

UNIVERSIDAD DE OVIEDO

ESCUELA POLITÉCNICA DE MIERES

MÁSTER EN TELEDETECCIÓN Y SISTEMAS DE INFORMACIÓN GEOGRÁFICA

**DEPARTAMENTO DE ORGANISMOS Y SISTEMAS
ÁREA DE INGENIERÍA AGROFORESTAL**

TRABAJO FIN DE MÁSTER

MODELOS DE PREDICCIÓN EN PLANTACIONES BIOENERGÉTICAS MEDIANTE TÉCNICAS LiDAR.

AUTOR: COVADONGA DIEGO ALONSO

TUTOR: PEDRO ÁLVAREZ ÁLVAREZ

COTUTOR: MARIA CASTAÑO DÍAZ

JUNIO, 2016

Agradecimientos

En primer lugar, me gustaría darle las gracias a Pedro Álvarez por la confianza depositada en mí para la realización de este trabajo. No sólo ha contribuido en mi formación con la dirección del trabajo fin de máster, si no que ha sido un excelente profesor durante mis últimos años de carrera. Le animo a continuar formando a futuros alumnos de la misma forma que lo hizo hasta ahora. Ojalá pueda seguir trabajando y aprendiendo con él en el futuro.

En segundo y último lugar, pero no por ello menos importante, a María Castaño ya que sin ella este trabajo no hubiese salido adelante. A pesar de los infortunios, trabajar contigo siempre es enriquecedor en todos los sentidos. Poco queda decir que no sepas, solo desearte suerte en lo que te queda y que nuestros caminos forestales se vuelvan a cruzar algún día, ya que los personales dudo que se separen nunca.

Resumen

El avance tecnológico asociado al sector forestal se encuentra en un momento de innovación en las técnicas relativas a la cuantificación de las masas vegetales, la extracción y el procesado. Por ello la aplicación de nuevas técnicas y metodologías serán una gran oportunidad para estimar rendimientos, capacidades productivas, y sobre todo una mejora en la gestión, al disponer de más y más fiable información.

Durante los últimos años los inventarios forestales suponían uno de los costes altos dentro de un determinado presupuesto que se reducía por considerarse suficiente información de una masa forestal la medición de unas pocas parcelas de muestreo. Con la incorporación de la tecnología LiDAR (*Light Detection and Ranging*) se reduce el coste en determinadas masas forestales regulares obteniendo información del conjunto arbolado y su estructura.

Los cultivos energéticos son masas forestales que debido a su corto turno de corta y alta densidad requieren de grandes esfuerzos de inventario. Cuando ya se demostró en estudios anteriores su productividad como alternativa energética renovable, ahora se busca mejorar su gestión mediante el diseño de modelos ajustados en base a mediciones LiDAR. La obtención de información biométrica de la plantación o del cultivo con alta precisión y sin recurrir a un elevado presupuesto hará del LiDAR una herramienta de gestión forestal óptima.

En este TFM se ha expuesto el proceso de procesado de datos LiDAR (no explicado en detalle en la mayoría de los documentos publicados) y el ajuste de modelos predictivos para diversas variables biométricas. Para ello se han empleado los archivos gratuitos del IGN con una densidad de puntos de 0,5 puntos/m². Finalmente se han ajustado diversos modelos, mediante técnicas paramétricas (regresión simple y múltiple) y técnicas no paramétricas (CART) con una buena capacidad predictiva y fiabilidad.

Palabras clave: LiDAR, inventarios forestales, cultivos energéticos, biomasa, modelos ajustados.

Abstract

The technological advances associated with the forestry sector are in a time of innovation in techniques relating to the quantification of plant mass, extraction and processing. Therefore, the application of new techniques and methodologies will be a great opportunity to estimate efficiency, production capacity, and especially an improved management to have more and more reliable information.

In recent years, forest inventories mean for one of the high costs within a particular budget that was reduced because it is considered enough information the measurement of a few sample plots. With the incorporation of LiDAR (Light Detection and Ranging) technology, in regular forest the cost is reduced and information about woodland and its structure is obtained.

Energy crops are forests that because of its short period of cut and high density require large inventory efforts. When already demonstrated in previous studies the productivity as a renewable energy alternative, now is necessary to improve its management by designing models adjusted based on LiDAR measurements. Obtaining biometric information of planting with high precision and without resorting to a high budget, LiDAR will be an optimal tool for forest management.

In this TFM it is discussed the process of LiDAR data processing (not explained in detail in most papers published) and adjusting predictive models for various biometric variables. To do this, we have used the free IGN files with a dot density of 0.5 points / m². Finally, different models have been adjusted by parametric techniques (simple and multiple regressions) and nonparametric techniques (CART) with good predictive power and reliability.

Keywords: LiDAR, forest inventories, energy crops, biomass, adjusted models.

ÍNDICE

RESUMEN	3
ABSTRACT	4
1. INTRODUCCIÓN Y ANTECEDENTES.	8
2. MATERIAL Y MÉTODOS.	12
2.1 Extracción de datos de campo.	12
2.1.1 Descripción del área de estudio	12
2.1.2 Inventario Dasométrico.	14
2.2 Extracción de datos LIDAR	15
2.2.1 Obtención de datos LiDAR	15
2.2.2 Tratamiento de datos LiDAR.	17
2.2.2.1 Preparación de los archivos LAS.	17
2.2.2.2 Generación del Modelo Digital del Terreno.	18
2.2.2.3 Extracción del Modelo Digital de Superficie	19
2.2.2.4 Extracción de la nube de puntos por parcela.	20
2.2.2.5 Cálculo de los estadísticos de los retornos de vegetación.	21
2.3 Análisis estadístico	22
2.3.1 Ajuste de modelos paramétricos	23
2.3.1.1 Correlaciones de Pearson	23
2.3.1.2 Regresión Simple	24
2.3.1.3 Regresión Múltiple	25
2.3.2 Ajuste de modelos no paramétricos	25
2.3.2.1 CART (Classification and Regression Trees)	25
2.3.3 Diagnóstico de los modelos	27
▪ Análisis de la bondad de ajuste	27
▪ Gráfico de residuos	28
▪ Test de colinealidad	28
3. RESULTADOS.	29
3.1 Resultados de las correlaciones de Pearson.	30
3.2 Resultados de modelos de regresión simple.	31
3.3 Resultados de modelos de regresión múltiple.	34
3.4 Resultados de diagnóstico de los modelos	37
3.5 Resultados del árbol de clasificación y regresión	39

4. DISCUSIÓN	42
5. CONCLUSIONES	44
6. BIBLIOGRAFÍA	45
Páginas web de consulta	49
7. ANEXO I: FICHA TÉCNICA DEL ENSAYO	50
8. ANEXO II: VARIABLES OBTENIDAS DEL COMANDO CLOUDMETRICS	51
9. ANEXO III: MAPA DE SITUACIÓN.	52
10. ANEXO IV: GRÁFICOS DE HOMOGENEIDAD DE VARIANZAS.	53

ÍNDICE DE TABLAS

Tabla 1: Variables del inventario de campo.....	14
Tabla 2: Especificaciones del vuelo LiDAR. Fuente: PNOA.....	16
Tabla 3: Coeficientes de correlación de Pearson para las variables dasométricas.....	30
Tabla 4: Coeficientes de correlación de Pearson para las variables de estimación de biomasa.	31
Tabla 5: Modelos de regresión simple para el conjunto del ensayo.....	32
Tabla 6: Modelos de regresión simple especificados por especie de plantación para las variables dasométricas de inventario.	33
Tabla 7: Modelos de regresión simple especificados por especie de plantación para las variables de estimación de biomasa.	34
Tabla 8: Regresión paso a paso general en SPSS.....	35
Tabla 9: Regresión paso a paso por especie en SPSS para variables de inventario de campo.	36
Tabla 10: Regresión paso a paso por especie en SPSS para variables de estimación de biomasas.	367

1. Introducción y antecedentes.

Actualmente el sector forestal se encuentra en un proceso de renovación tecnológica que lo haga competitivo en el mercado. Dentro de los diferentes avances que surgen; desde la extracción hasta la gestión, la reducción de costes en el proceso de inventario es objeto de mejora.

Una de las tecnologías que irrumpe en los últimos años es el LiDAR (*Light Detection and Ranging*) que mejora los inventarios forestales, reduciendo su coste y el tiempo empleado. Mediante el procesado de los datos tomados por un sensor aerotransportado se obtienen métricas sobre las masas forestales.

El acrónimo LiDAR viene de la traducción del inglés de '*Detección y Localización de la Luz*' (Wher y Lhor, 1999), esta tecnología es un sistema activo de detección remota basado en un sensor láser. Su fundamento son un conjunto de técnicas que permiten determinar la distancia desde un emisor a un objeto o superficie utilizando un haz láser (Sánchez, 2010). Esta distancia se establece midiendo el tiempo de retraso entre la emisión y su detección a través de la señal reflejada (Hernández Clemente *et al.*, 2016).

Los inventarios tradicionales de estimación de producción sobre masas forestales calculaban las existencias en función del presupuesto para el establecimiento de parcelas de medición (González-Ferreiro *et al.*, 2012), mientras que un vuelo LiDAR evalúa el conjunto de la superficie forestal.

La posibilidad de obtener datos acerca de la estructura forestal y sus variables dasométricas a partir de sensores aerotransportados es un gran avance en la cuantificación de los recursos que se obtienen de forma directa de los bosques (García-Ferrer, 2013).

Esta tecnología la integración de la telemetría láser, el sistema de posicionamiento global (GPS) y la medición de movimiento inercial (IMU). El haz láser que emplea, emite miles de pulsos de energía por segundo hacia la tierra creando una franja densa de puntos 3D (Manue, 2007).

Esa nube de puntos 3D se basa en la medición precisa del tiempo de retorno de las porciones de pulso al sensor, y así, calcular la distancia que separa a éste de la superficie terrestre y de los objetos que existen sobre ella (Magdaleno *et al.*, 2006). Dado que la posición y la orientación del sensor son conocidas para cada pulso emitido, cada señal de retorno tiene unas coordenadas tridimensionales únicas.

Existen dos formas de toma de datos en los sistemas LiDAR; de onda completa (*full waveform*) y retornos discretos (Valbuena, 2013). En el área forestal se suelen utilizar los segundos; estos almacenan las coordenadas tridimensionales en el primer y último retorno. Los sistemas *full-waveform*, en cambio, almacenan la onda completa (Crespo *et al.*, 2015).

El funcionamiento de los retornos del haz láser es simple, Gonçalves (2007) lo diferencia en tres tipos:

- Sobre una superficie sólida (suelo, edificaciones, etc.) el rayo se refleja y retorna al avión o helicóptero.
- Sobre el agua no devuelve ningún retorno debido a la alta absorción de esta.
- Sobre la cubierta vegetal, el rayo choca con las partes más altas del árbol (las copas) donde se refleja y retorna al avión; pero como no es una superficie sólida como tal una parte de ese rayo no es reflejada y va al suelo (o a otros elementos de la estructura del árbol, como ramas bajas) y vuelve al avión.

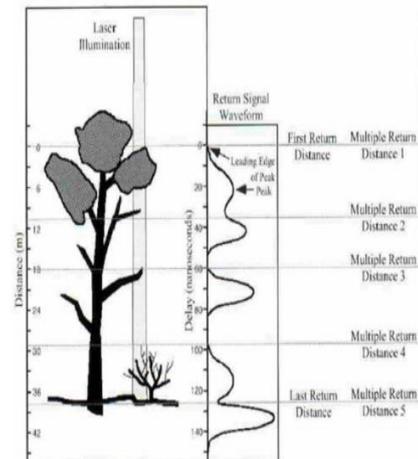


Figura 1: Izquierda; esquema de funcionamiento LiDAR de múltiple retorno. Fuente: Lefsky et al., 2002.

A pesar de que los sistemas LiDAR son capaces de ajustar en precisión de forma bastante elevada tienen diversas fuentes de error debidas entre otros factores a la alineación entre el láser, avión y sistema de navegación; errores en el sistema GPS/INS (Huising y Gomes Pereira 1998; Maas, 2001). En general la precisión de los datos, procedentes del PNOA a una densidad de 0.5 puntos/m², adquiridos por un usuario obtienen un RMSE_z menor o igual a 0.20 metros en altimetría y 0.50 metros en los MDT obtenidos con una malla de 5x5 metros (www.csic.es). Además, hay que tener en cuenta que cuando se trata de superficies forestales la orografía del terreno producirá solapes y/o áreas no cubiertas (García, 2014).

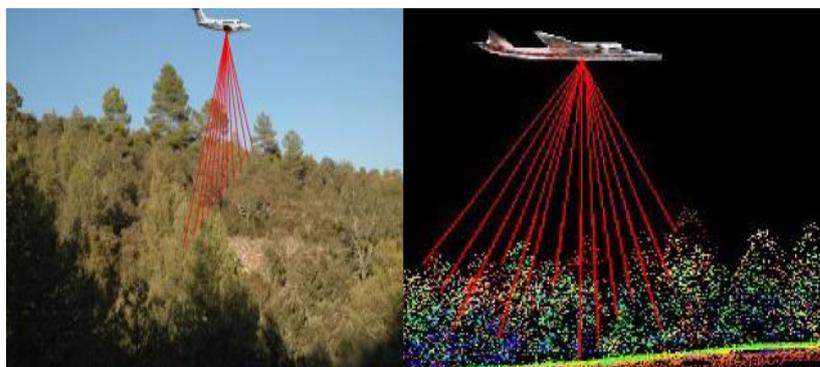


Figura 2: Generación de nube de puntos 3D sobre masas forestales. Fuente: García, 2014.

La distribución de datos LiDAR se hizo en un principio a través del formato ASCII (ficheros de texto para las coordenadas x, y, z de cada punto), algo que resultó tedioso y pronto quedó obsoleto (Hernández Clemente et al., 2016). A partir del año 2000 se utiliza el formato *LAS (Laser File Format Exchange Activities) desarrollado por la Sociedad Americana para la

Fotogrametría y la Teledetección (ASPRS 'American Society for Photogrammetry and Remote Sensing').

La información básica que contienen estos ficheros es: coordenadas x e y, datos de elevación (respecto a la superficie) e intensidad (cada material refleja de manera diferencial el rayo de luz láser), número de pulsos emitidos y reflejados y ángulo nadir. A partir de esta información se pueden procesar los puntos con distintas finalidades.

La gran ventaja de este sistema radica en su capacidad de penetrar en la cubierta vegetal y capturar información de la estructura vegetal y el suelo (Morán *et al.*, 2013) siendo capaces de generar una nube de puntos de esa estructura en tres dimensiones (Fabra, 2012). Es esta gran funcionalidad, la que permite ajustar modelos que ajustan con una alta predicción las variables dasométricas propias de un inventario forestal.

LiDAR puede ser una tecnología revolucionaria en el sector forestal, aunque nunca llegue a reemplazar la salida y estudio en campo (Andersen, 2005). La información que se obtiene de su procesamiento pueden ser tan común en los próximos años como lo es hoy en día el uso de fotografías aéreas y mapas topográficos.

Se plantea como caso de estudio la implantación de este sistema como método de inventario en plantaciones con fines bioenergéticos situadas en el municipio de Langreo, Principado de Asturias. De esta forma se puede comparar previos inventarios de campo con la extracción de variables LiDAR y analizar su relación mediante modelos de regresión ajustados.

Para ello se continúa los ensayos llevados a cabo por el Grupo HUNOSA y el grupo de investigación Sistemas Forestales Atlánticos (GIS-Forest) de la Universidad de Oviedo en los que se analiza la productividad de cultivos energéticos en terrenos baldíos de la minería.

Los datos de partida del estudio será un inventario de campo en otoño del año 2012 coincidente con el último vuelo LiDAR realizado por el PNOA (Plan Nacional de Ortofotografía Aérea) sobre el Principado de Asturias. El proyecto comenzó en 2008 planteando la posibilidad de buscar una rentabilidad a antiguas escombreras en desuso a través de plantaciones de sauce de crecimiento rápido.

La búsqueda de nuevas materias primas como fuente de energía se enmarca en un proyecto global de reducción de gases contaminantes a la atmósfera. Una alternativa como la biomasa que mantiene un balance neutro de emisiones de CO₂ necesita de estudios que maximicen su rentabilidad a la hora de ser competitiva en el mercado energético; y que ayude a cumplir de esta forma los objetivos establecidos en el Protocolo de Kyoto en 1997.

Es por ello, que en base a dicho inventario de campo realizado en 54 parcelas de plantación se estudia la viabilidad y la precisión de la tecnología LiDAR como método de análisis de la estructura vegetal en detrimento de los viejos inventarios que incrementan los costes de producción.

El establecimiento de plantaciones de biomasa para producción energética, no es sólo una fuente de energía renovable atractiva (González-Ferreiro *et al.*, 2013) sino que tiene muchas ventajas por las que merece la pena su desarrollo. La recuperación de actividades en el medio rural, el balance de CO₂ neutro o la recuperación de áreas degradadas son algunas de las causas por las que los cultivos con fines bioenergéticos son una opción de futuro.

Uno de los futuros campos de trabajo del gestor forestal es la selvicultura energética, siendo ésta una rama de conocimiento aplicada al tratamiento de plantaciones con fines bioenergéticos. Para ello, es imprescindible el conocimiento de la estructura del bosque (organización espacial vertical y horizontal) y sus componentes (Baca Venegas, 2000).

Entre las múltiples ventajas de la aplicación de la tecnología LiDAR a la selvicultura energética destacan:

- Información tridimensional geográficamente extensa, continua y de alta precisión. Lo que supone contar con un volumen de información enorme de la estructura del bosque.
- Modelos Digitales del Terreno (MDT/MDE), de Superficie (MDS) y de vegetación (MDV) de alta precisión.
- Bajo coste por unidad de superficie, el LiDAR permite disminuir considerablemente los costes de realización de cartografía e inventarios forestales.

Las aplicaciones de los sensores LiDAR al campo forestal son numerosas, destacan las siguientes; atributos forestales (Lim *et al.*, 2003; Wang y Glenn, 2008), modelos de combustible y simulación del comportamiento del fuego (González *et al.*, 2012, Mutlu *et al.*, 2008, Riaño *et al.*, 2007), establecimiento de modelos de copa individuales (Hyypä y Inkien, 1999), índices de calidad de ribera (Magdaleno *et al.*, 2006), estimación de biomasa y sumideros de carbono (Zolkos *et al.*, 2013, García *et al.*, 2010), conservación y análisis de hábitats (Mundt *et al.*, 2006) y numerosos estudios sobre inventarios forestales (Hyypä *et al.*, 2008, Naesset, 2002, 2004, Maltamo *et al.*, 2005) y alturas de vegetación (Hollaus *et al.*, 2006; Packalén *et al.*, 2008).

Estos estudios se basan en la alta correlación existente entre la estructura de la vegetación y las variables biofísicas de gestión forestal, de esta forma se pueden ajustar modelos para el cálculo de variables como área basimétrica, volumen de madera o biomasa, convirtiéndose así en una herramienta de gestión de alta funcionalidad.

Además, no se trata solamente de la obtención de parámetros de medición sino de modelos (Modelo Digital de Elevaciones, Modelo Digital de Superficies o Modelo Digital de Vegetación) que permiten el estudio del medio físico, la ordenación de un determinado territorio o evaluación de otros parámetros ambientales.

La existencia de información disponible y gratuita por parte del Plan Nacional de Ortofotografía Aérea en densidades de retorno de 0.5 pulsos/m² ofrece una oportunidad de descenso de costes en los inventarios forestales (González-Ferreiro *et al.*, 2012). Esta disponibilidad de información espacialmente continua sobre la estructura y composición de las

masas forestales, unido al abaratamiento de esta tecnología está comenzando a dejar obsoletos los antiguos inventarios de campo (Valbuena, 2014).

Atendiendo a las necesidades de estudio del presente trabajo, **se establecen como principales objetivos;** (i) *determinación de modelos ajustados que permitan estimar distintas variables dasométricas a través de la extracción de datos LiDAR perteneciente al vuelo del PNOA de 2012 y coincidente con la extracción de datos de campo en la misma época,* (ii) *estudiar la predicción de dichos modelos como herramienta que facilita la gestión de plantaciones forestales con fines energéticos,* (iii) *comprobación de las hipótesis halladas en estudios anteriores sobre productividad.*

2. Material y Métodos.

2.1 Extracción de datos de campo.

2.1.1 Descripción del área de estudio

La zona de estudio pertenece a un área restaurada de una antigua escombrera de carbón de la empresa Grupo HUNOSA en el monte de Mozquita situado en el concejo de Langreo, Principado de Asturias (Ver Mapa 1. Situación, Anexo III). En esta área se establece un ensayo de plantación de especies del género *Salix* con fines energéticos en 54 parcelas ubicadas geográficamente a 43° 16' 13,56" Norte y 5° 42' 03,69" Oeste (UTM Huso 30; x = 280.822, y = 4.794.372) aproximadamente.



Figura 3: Vista general del estado actual del ensayo de plantación.
Fuente: propia

Se establece este género de plantación en el marco de los cultivos SCR (*Short Rotation Coppice*) como cultivos de elevada capacidad de rebrote que es el objetivo buscado en este tipo de plantaciones con fines energéticos.

A su vez cabe destacar las tres especies del género *Salix* elegidas:

- Bjor (*Salix schwerinii* x *Salix viminalis*)
- Inger (*Salix trianta* x *Salix viminalis*)
- Olof (*Salix viminalis* x (*Salix schwerinii* x *Salix viminalis*))



Figura 4: Foto del clon Olof en el área de estudio.
Fuente: propia.

Se estableció un diseño de parcelas de plantación *Split-plot* o de parcelas subdivididas, cada una de las cuales ocupan una superficie de 400 m² (20 x 20 m), ajustándose al tamaño mínimo recomendado para obtención de datos experimentales fiables en los estudios realizados, ya que las parcelas de menores dimensiones sobreestiman la producción (Vande Walle *et al.*, 2007).

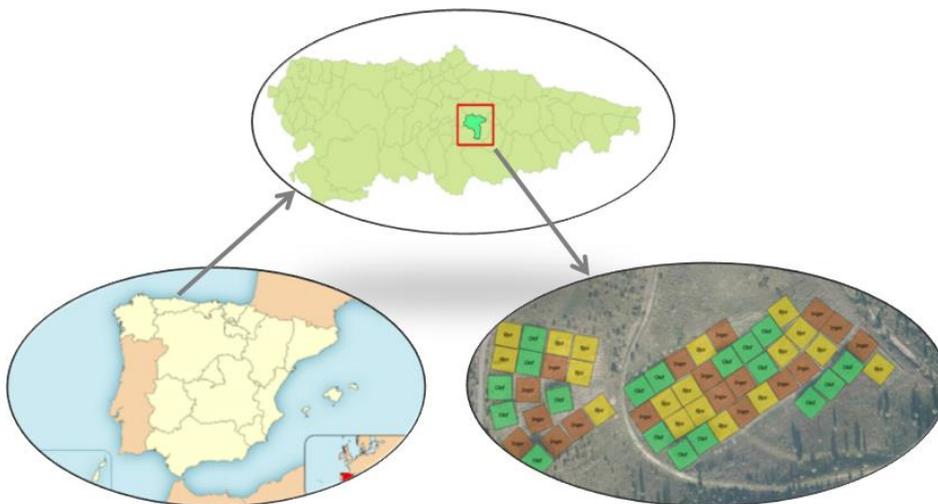


Figura 5: Ubicación de las parcelas de ensayo.
Fuente: propia.

2.1.2 Inventario Dasométrico.

Para el ajuste de los modelos de regresión, los datos de campo y los datos LiDAR deben coincidir exactamente en espacio y tiempo (Donés, 2009). La nube de punto LiDAR se encuentra georreferenciada cuando se descarga del IGN (Instituto Geográfico Nacional) así que los datos de campo deben estarlo también. El sistema de referencia utilizado es el ERTS89 UTM Huso 30.

A través de los puntos de referencia de cada parcela (4 vértices de la parcela) obtenidos mediante un GPS Trimble con precisión submétrica y con la ortofoto de máxima actualidad del PNOA como base se digitalizaron las parcelas de estudio a través del software QGIS 2.14.

El inventario que se diseñó en 2012 consistió en la contabilización en cada cepa¹ del número de brotes emitidos por estaquilla, el diámetro basal a 20 cm y la altura total de cada brote. Para tomar los datos dentro de cada parcela se midió a partir de la 5ª fila en horizontal y la 8ª cepa en vertical, evitando así el efecto borde.

Dicho inventario fue realizado en otoño de 2012 coincidente con el vuelo LiDAR realizado sobre el Principado de Asturias por el PNOA, con un desfase de tiempo menor a 2 meses entre ambas mediciones. La importancia de que los datos de campo y los datos LiDAR sean coincidentes en el tiempo radica en que si no se cumple esa condición, los resultados obtenidos en el ajuste de los modelos paramétricos y no paramétricos no obtienen la precisión ni la fiabilidad buscada en el estudio.

En cuanto al instrumenta utilizado en el inventario de campo, para la medición de alturas se empleó una pértiga telescópica desplegable, y el hipsómetro Vertex IV, éste calcula las alturas trigonométricamente con la distancia y los ángulos. En las parcelas con predominio de brotes inferiores a 2 m se utilizó una cinta métrica convencional. Los diámetros se midieron con un calibre digital.

Tabla 1: Variables del inventario de campo

<i>Variables dasométricas de inventario</i>	D₂₀ : diámetro del brote a los 20 cm de altura
	D₁₃₀ : diámetro del brote a los 130 cm de altura
	H : altura del brote en metros
<i>Estimación de biomasa</i>	PV : biomasa en peso verde en t/ha
	PS : biomasa en peso seco en t/ha

¹ Cepa: conjunto de la base del tronco (tocón), cuello de la raíz y zona próxima de las raíces gruesas.

En cuanto a las variables de estimación de biomasa, se calculan una vez finalizado el turno² de la plantación, y en el momento de corta se calculan los pesos secos y verde en t/ha. El peso verde se calcula en campo por cepa y posterior extrapolación para el conjunto de la plantación mientras que el peso seco se calcula en laboratorio.



Figura 6: Material utilizado en el inventario de campo. Cinta métrica, calibre digital y Vertex IV.
Fuente: propia.

2.2 Extracción de datos LIDAR

2.2.1 Obtención de datos LiDAR

A través del centro de descargas del IGN se descargan los ficheros digitales con información altimétrica de la nube de puntos LiDAR, distribuidos en ficheros de 2x2 km de extensión. El formato de descarga es un archivo *.LAZ (formato de compresión de ficheros *.LAS), en la información auxiliar el IGN ofrece una herramienta de descompresión y visualización de ficheros *.LAZ y *.LAS (*LASTask Tools*).

Las nubes de puntos han sido capturadas mediante vuelos con sensor LiDAR con una densidad de 0,5 puntos/m², y posteriormente clasificación de manera automática y coloreada mediante RGB obtenido a partir de ortofotos del Plan Nacional de Ortofotografía Aérea (PNOA) con tamaño de pixel de 25 o 50 cm.

² Turno de corta: Número planificado de años que transcurre entre la formación o regeneración de una masa y el momento de su corta final

Tabla 2: **Especificaciones del vuelo LiDAR. Fuente: PNOA**

Parámetro	Especificación
Tamaño de pixel	0.25 metros
Densidad de puntos	0.5 puntos/m ²
Sistema geodésico de referencia	ETRS89
Proyección cartográfica	UTM
Frecuencia de pulso	45 kHz
Áreas cubiertas	Hojas 1:5.000 completas
Sensores	Leica ALS50-60, LMS-Q680 de Riegl

Una vez descomprimidos los archivos LiDAR (a través de un software de descompresión de datos *.LAZ) se comienza con el tratamiento de los mismos a través de dos softwares específicos:

- *Fusion USDA versión 3.40*: es una herramienta de tratamiento de datos LiDAR desarrollado por la Dirección General de Investigación del Servicio Forestal de Estados Unidos (McGaughey, 2010). Está orientado al análisis forestal con datos LiDAR.
- *FugroViewer*: software para la visualización de datos geoespaciales. Resulta de utilizada para realizar un análisis exploratorio de la nube de puntos.

La nube de puntos LiDAR descomprimida se numera por hojas; si se desconoce la hoja que ocupa la nube de puntos, el IGN pone a disposición de los usuarios la base de datos en formato *shapefile* de la hoja perteneciente a cada municipio para visualizarla en un GIS o un Excel con las hojas por municipio.

Una vez conseguida la nube de puntos referida al área de estudio, se procede al tratamiento para obtener una nube de puntos por parcela. Se llevan a cabo una serie de comandos ejecutados con el software Fusion; que trabaja a través de la consola de MS-DOS de Windows, para obtener los datos LiDAR referentes a cada una de las 54 parcelas de estudio.

El software fusión está implementado mediante lenguaje de programación C++, trabaja ejecutando una serie de comandos a través de un intérprete; la consola MS-DOS en Windows (cmd.exe). Cuando se abre la consola se conecta al software y se trabaja ejecutando los scripts desde ella.

Estos scripts o comandos se pueden implementar de dos formas diferentes; **a)** mediante la generación de textos de programación en una hoja de cálculo Excel que permite aplicar el mismo comando para un número grande de puntos de forma rápida o **b)** a través de archivos de extensión .bat* (batch) editado en Notepad++ que permite ejecutar el comando directamente sin la necesidad de vincular la consola al software.

A continuación, se describen los comandos empleados en el tratamiento de los archivos *LAS, especificados para la ruta de almacenamiento empleada en el estudio. Tan sólo habría

que copiarlo en la consola para que se ejecutasen (con especial atención a las rutas de entrada y salida) en caso de aplicación a otros casos prácticos.

2.2.2 Tratamiento de datos LiDAR.

2.2.2.1 Preparación de los archivos LAS.

Antes de comenzar a tratar la información contenida en la nube de puntos como tal, es necesario ejecutar un paso previo de clasificación de dicha nube. Mediante este proceso se diferencia la tierra desnuda de la vegetación u otros objetos existentes sobre la superficie.

El comando *GroundFilter* está diseñado para filtrar esa nube de puntos LiDAR, siendo la salida un archivo que contiene sólo los puntos clasificados almacenados en formato *LAS. Este comando produce conjuntos de puntos que dan lugar a modelos de superficie adecuados para calcular la altura de la vegetación. El algoritmo se basa en una predicción lineal (Kraus y Pfeifer, 1998) con un ajuste individual para cada medida.

En primer lugar, esta función asigna un valor al suelo desnudo, que no incluirá puntos de vegetación. Los puntos relativos a la vegetación se encontrarán por encima de esta primera clasificación. Posteriormente, se pueden asignar pesos a los puntos LiDAR basados en la distancia y dirección que existe entre cada punto y la primera clasificación calculada. De esta forma se consiguen distintos estratos de clasificación.

Comando:

```
groundfilter /wparam:2 C:\datoslidar\las\puntosuelo.las 1 C:\datoslidar\las\280_4796.las
```

- Comando = *GroundFilter*
- Conmutadores = /wparam:2
- Output = C:\datoslidar\las\puntosuelo.las
- Tamaño de celda = 1
- Input= C:\datoslidar\las\280_4796.las (hoja LiDAR del PNOA)

Para el algoritmo diseñado por Kraus el switches 'wparam' tiene un valor de 2 (w en la ecuación) (McGaughey, 2010).



Figura 7: Nube de puntos filtrada.
Fuente: propia.

2.2.2.2 Generación del Modelo Digital del Terreno.

Una vez estén los puntos clasificados se procede a construir un Modelo Digital del Terreno (MDT). Doyle, 1978, definió como MDT al conjunto ordenado de números que representan la distribución espacial y las características del terreno. Mediante el algoritmo que procesa el comando *GridSurfaceCreate* se calcula la elevación de cada celda de la cuadrícula utilizando la elevación media de todos los puntos dentro de la hoja LiDAR.

Existen ciertos casos en los que la obtención de un MDT de alta precisión se ve obstruido por la existencia de un sotobosque espeso que impida que los retornos que provienen del suelo lleguen con información válida (Lefsky *et al.*, 2002).

Este no es el caso, ya que quizás otras masas forestales si contienen sotobosque en estratos bajos pero las plantaciones energéticas al ser cultivos de corto turno y con alta competencia sobre herbáceas que precisan de tratamientos selvícolas de mejora, como desbroces, que mantienen el suelo desprovisto de vegetación. Además, el diseño de la plantación en líneas permite al sensor aerotransportado obtener retornos del suelo sin dificultades.

Para la obtención de un MDT de alta calidad es necesario eliminar todos los puntos correspondientes a la vegetación, pero sin eliminar los correspondientes al suelo, esto es especialmente difícil en zonas con pocos retornos de suelo (Kraus y Pfeifer, 1998). Así funciona el comando empleado:

```
gridsurfacecreate C:\datoslidar\dtm\suelodef.dtm 1 m m 1 0 0 0  
C:\datoslidar\las\puntosuelo.las
```

- Comando: *GridSurfaceCreate*
- Output: C:\datoslidar\las\suelodef.dtm
- Conmutador: surfacefile: 1 m m 1 0 0 0 (Codificación de números relativa a: tamaño de celda, unidades para X e Y en metros, unidades para Z en metros, sistema de coordenadas UTM, zona del sistema de coordenadas, datum horizontal y vertical)
- Input: C:\datoslidar\lda\puntosuelo.las

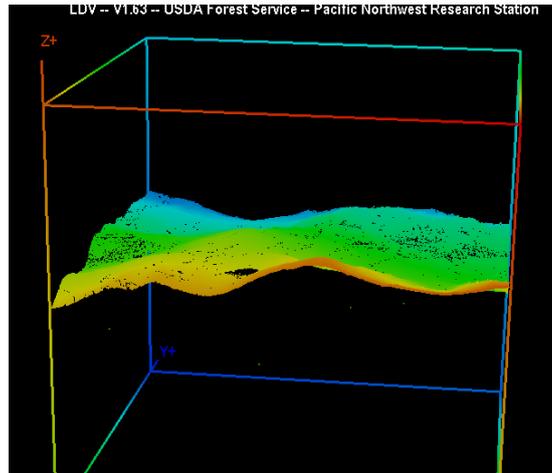


Figura 8: MDT en vista 3D en Fusion (USDA).
Fuente: propia

2.2.2.3 Extracción del Modelo Digital de Superficie

CanopyModel crea un modelo de superficie (MDS) de cobertura mediante la utilización de una nube de puntos LiDAR. Por defecto, el algoritmo asigna la elevación más alta del retorno más alto al centro de la celda. Esta función ejecuta un suavizado de la superficie generada usando la mediana o un filtro de media (McGaughey, 2010).

Resta las elevaciones del terreno a partir del retorno de las elevaciones de los puntos para producir un modelo de altura del dosel. La salida es un archivo de extensión '*.dtm*'.

```
canopymodel C:\datoslidar\dtm\copas.dtm 1 m m 1 0 0 0 C:\datoslidar\las\280_4796.las
```

- Comando: *CanopyModel*
- Output: C:\datoslidar\dtm\copas.dtm
- Conmutador: surfacefile: 1 m m 1 0 0 0 (ídem que para la creación del MDT)
- Input: C:\datoslidar\las\280_4796.las

2.2.2.4 Extracción de la nube de puntos por parcela.

Por un lado, se necesita cortar el área de estudio de Mozquita relativo a la zona dónde están ubicadas las parcelas de estudio del Grupo HUNOSA. Para ello se utiliza el comando *ClipData*; el resultado es un archivo de extensión *.LAS.

La necesidad de este paso radica en que los switches introducidos en el comando empleado permiten la **normalización de la nube de puntos**; esto es que cada punto asigne dentro de la cabecera de información su cota sobre el suelo. Si no se realiza este paso la cota que tendría cada retorno sería sobre el nivel del mar.

Se crea un área circular de 500 m de radio para cortar con *ClipData* (sobre un punto central en QGIS se crea un área de influencia que permita obtener las coordenadas).

```
ClipData /shape:1 /dtm:C:\datoslidar\dtm\suelodef.dtm /height  
C:\datoslidar\las\280_4796.las C:\datoslidar\las\mozquita.las 280911.5 4794259.6
```

- Comando: *ClipData*
- Output: C:\datoslidar\las\mozquita.las
- Conmutador: shape: 1 (da como resultado un área circular); (coordenadas esquina superior derecha – esquina inferior izquierda; minX minY maxX maxY), dtm (modelo digital del terreno que normalice los datos) y height (obtener alturas de puntos sobre el nivel del suelo).
- Input: C:\datoslidar\las\280_4796.las



Figura 9: Área de estudio cortada en vista 3D en Fugro Viewer.
Fuente: propia.

Después de recortar el área de estudio se procede a cortar las parcelas de estudio mediante el comando *PolyClipData*. El comportamiento predeterminado es producir un único archivo de salida que contiene todos los puntos que están dentro de todos los polígonos en el archivo de formas. Los archivos individuales producidos contienen los puntos dentro de cada polígono en el archivo de formas, y los puntos de recorte en un solo polígono especificado mediante un campo de la base de datos del archivo de formas.

Se elige nombrarlas por el campo PLOT, que es el código identificativo de la parcela.

```
PolyClipData /multifile /shape:1,* C:\datos lidar\shp\moz_bd_lidar.shp  
C:\datoslidar\las\parcelas.las C:\datoslidar\las\mozquita.las
```

- Comando: *PolyClipData*
- Output: C:\datoslidar\las\parcelas.las
- Conmutador: shape:1,* (indica que tome de nombre el campo uno de la base de datos asociada al archivo de formas).
- Input: C:\datoslidar\las\mozquita.las C:\datos lidar\shp\moz_bd_lidar.shp

Durante el trabajo de gabinete se encontró un problema en el recorte de datos. Cortar cada parcela con el comando *ClipData* implica el cálculo de la esquina superior derecha y la esquina inferior izquierda de cada parcela lo cual resulta lento cuando se tiene un elevado número de zonas a muestrear.

Además, se debe decir que se elige que el corte de cada parcela sea rectangular, asumiendo el efecto borde, ya que se desea tener la máxima información posible sobre el estudio debido a la baja densidad de puntos del vuelo LiDAR. Al utilizar el archivo shape generado en QGIS para la delimitación de las parcelas, la nube de puntos recortada tiene esa misma forma.



Figura 10: Área de estudio cortada en vista 3D en Fugro Viewer.
Fuente: propia.

2.2.2.5 Cálculo de los estadísticos de los retornos de vegetación.

CloudMetrics calcula una variedad de parámetros estadísticos que describen un conjunto de datos LiDAR. Las métricas se calculan utilizando las elevaciones puntuales y valores de intensidad. La salida es un archivo de 'csv' (texto delimitado por comas) que se visualiza en Excel.

El software Fusion (USDA) calcula las métricas LiDAR con las variables correspondientes a la altura de la nube de puntos, la intensidad del haz láser de retorno (que será diferente en función de la cubierta vegetal) y el conteo de retornos.

Las elevaciones permitirán conocer parámetros como la altura media de la parcela o la altura dominante (relacionada con la elevación del percentil 95), los retornos permiten conocer el volumen de cobertura arbórea de una zona (relacionados con el cálculo de fracción de cabida cubierta) y las intensidades permiten diferenciar la cobertura de la vegetación.

Para conocer las variables que se obtienen consultar Anexo II, dónde se especifica el conjunto de variables derivadas de la aplicación del comando siguiente (referidas a los retornos de la nube de puntos, la intensidad del haz láser y las elevaciones).

```
CloudMetrics /above:1 /minht:0.5 C:\datoslidar\las\*.las C:\datoslidar\csv\est_parcelas.csv
```

- Comando: *CloudMetrics*
- Output: C:\datoslidar\csv\est_parcelas.csv
- Conmutador: /above:1 /minht:0.5 ('above' se usa para obtener unas métricas particulares y referidas a la FCC³; 'minht' se usa para delimitar que los estadísticos no se refieran a alturas de brote menos a 0.5 metros).
- Input: C:\datoslidar\las*.las (incluye los archivos recortados *.LAS de las 54 parcelas)

2.3 Análisis estadístico

El siguiente apartado recoge una breve descripción de los distintos análisis estadísticos llevados a cabo para la obtención de modelos ajustados que describan el comportamiento de las variables obtenidas mediante dos softwares; SPSS 17.0 (Statistical Product and Service Solutions) y Statgraphics Centurion XVI.

³ FCC o fracción de cabida cubierta se define como el grado de recubrimiento del suelo por la proyección vertical de las copas de la masa vegetal.

Se procede al estudio de modelos de regresión paramétricos y no paramétricos. La estadística paramétrica es una rama de la estadística que comprende los procedimientos estadísticos y de decisión que están basados en las distribuciones de los datos reales (Contreras, 2000). Estas son determinadas usando un número finito de parámetros.

Los modelos no paramétricos, en cambio, no asumen información de la distribución. Esta última también se denomina de libre distribución. Estos métodos funcionan bien cuando no se supone la normalidad de la muestra siendo en ocasiones mejores que los anteriores para detectar diferencias entre las poblaciones (Del Carmen, 2014).

2.3.1 Ajuste de modelos paramétricos

2.3.1.1 Correlaciones de Pearson

En una primera fase de exploración de los datos muestrales y previo a la obtención de relaciones estadísticas entre las variables de estudio, es necesario saber el grado de asociación o correlaciones entre dichas variables.

Dentro de los diferentes métodos de análisis de correlaciones se determina que Pearson es el más adecuado al caso de estudio que otros habituales, como Spearman, ya que se trata de un coeficiente aplicable cuando la métrica de las variables correlaciones es de intervalo, y se supone que éstas se distribuyen en la población de forma normal (Rodríguez, 2016). Spearman es más práctica en variables de tipo ordinal con violación del supuesto de normalidad.

El coeficiente de Pearson mide el grado de asociación lineal entre dos variables cualesquiera (Díaz y Fernández, 2002). El resultado de estas correlaciones puede ser positivo o negativo dependiendo de si esa relación es directa o inversa. Se trata de un indicador de la relación lineal existente entre dos variables cuando estas se distribuyen en una población de forma normal (Rodríguez, 2016).

Para el presente estudio se utiliza el software SPSS 17.0 que permite obtener que variables de estadísticas LiDAR (variable independiente) mantienen correlaciones más elevadas con las variables de inventario de campo (variable dependiente). Se obtienen correlaciones para un nivel de significación del 99%.

El coeficiente de correlación de Pearson es un índice de fácil interpretación. Sus valores absolutos oscilan entre 0 y 1. Esto es, si se tienen dos variables X e Y, y se definen el coeficiente de correlación de Pearson entre estas dos variables como r_{xy} entonces: tan fuerte es una relación de +1 como de -1. Viene definido por la siguiente expresión:

$$r_{xy} = \frac{\sum Z_X Z_Y}{N} \quad \text{[Ecuación 1]}$$

Esta correlación puede ser lineal y directa, positiva (al aumentar una variable, aumenta la otra), negativa (inversa a la positiva) y nula cuando no existe ninguna correlación entre la nube de puntos (Rodríguez, 2016).

2.3.1.2 Regresión Simple

Los modelos de regresión simple tratan de explicar el comportamiento de una variable cuantitativa Y como función de otra variable cuantitativa X observable. Siendo en este ejemplo Y la variable respuesta (dependiente) y X la variable regresora o predictora (independiente) (Pardo y Ruiz, 2002).

Existen diversos modelos de regresión simple, cada uno se adecúa más o menos al caso de estudio siendo el más común el lineal (Martín, 2011).

$$Y_i = \beta_0 + \beta_{1x_i} + U_i \quad [\text{Ecuación 2}]$$

- Siendo β_0 y β_1 la ordenada en el origen y la pendiente de la recta de regresión. U_i es el error experimental.
- β_0 representa el valor medio de la respuesta Y cuando la variable explicativa X vale 0.
- β_1 representa la variación que experimenta en media la respuesta Y cuando la variable explicativa X aumenta en una unidad.

Existen diversos modelos que, no siendo lineales, se explican a partir de la teoría supuesta y que se pueden ajustar mejor a las variables que el modelo lineal, pero todos ellos son casos de regresión simple (Rodríguez, 2016). El uso de este tipo de modelos en los que solo entra una variable respuesta se justifica por su carácter práctico y operativo.

▪ *Error Medio Absoluto*

Definido por Stauffer y Seaman en 1990, se define como el promedio de los errores obtenidos en el valor absoluto, entendiéndose a éste como el residuo en valor absoluto o el error promedio que se puede esperar en una predicción.

$$MAE = \sum_{i=1}^N \frac{|\phi_i - \phi_{iobs}|}{N} \quad [\text{Ecuación 3}]$$

La ecuación describe el valor absoluto de la media aritmética ($\phi_i - \phi_{iobs}$) de los errores individuales de predicción. Es un estimador apropiado en datos lineales y simétricos. A menor MAE mejor análisis y predicción (Moreno-Gil, 1998).

2.3.1.3 Regresión Múltiple

El software SPSS empleado para la obtención de las correlaciones de Pearson emplea el método denominado “*stepwise*” o “regresión paso a paso” para la obtención del conjunto de variables regresoras que deben estar en el modelo de regresión múltiple (Pardo y Ruiz, 2002).

Este procedimiento empieza introduciendo todas las posibles variables de la ecuación, a partir de las cuales el programa escoge en primer lugar la variable explicativa con una mayor correlación con la variable respuesta.

Se calculan, posteriormente, las correlaciones parciales entre el resto de las variables independientes y la variable respuesta; la segunda variable seleccionada es la que tiene la mayor correlación parcial con la variable respuesta, se procede de este modo sucesivamente hasta completar el número de variables que se ha definido, escogiendo las variables con una mayor correlación y a la vez eliminando las introducidas, si ya no son estadísticamente significativas al introducir nuevas variables (Rodríguez, 2016).

La ventaja que proporciona este método de regresión es que evita la multicolinealidad al no mantener fijas las variables que ya entraron en la primera etapa al modelo (Rodríguez, 2016). Como inconveniente destaca la alta restricción en la entrada de variables al modelo que caracteriza este método (Barón y Téllez, 2000).

2.3.2 Ajuste de modelos no paramétricos

2.3.2.1 CART (Classification and Regression Trees)

El primer planteamiento del análisis estadístico por clasificación y regresión de árbol (CART) lo planteó Leo Breiman en 1984 (Breiman *et al.*, 1984) generando un gran interés en la comunidad científica por su fácil implementación y clara interpretación.

Se trata de un método que utiliza datos históricos para construir árboles de clasificación y de regresión, los cuales son usados para clasificar o predecir nuevos datos. Estos árboles CART pueden manipular fácilmente variables numéricas y/o categóricas. Este modelo no paramétrico es de gran utilidad en las actividades de toma de decisiones, busca determinar de

forma automática cuales son las variables explicativas más significativas respecto a una variable respuesta (Mitjans, 1996).

Es un método de segmentación binaria donde el árbol es construido dividiendo repetidamente los datos. En cada división los datos son partidos en dos grupos mutuamente excluyentes (Serna, 2009). El nodo inicial es llamado nodo raíz y se divide en dos nodos, luego el procedimiento de partición es aplicado a cada nodo descendiente por separado. Las divisiones se seleccionan de modo que “la impureza” de los nodos sea menor que la del nodo raíz y éstas están definidas por un valor de una variable explicativa (Deconinck *et al.*, 2006), buscando la mayor homogeneidad posible en las divisiones.

Se basa en el algoritmo conocido como particionamiento recursivo, que determina la medida de impureza, es un proceso paso a paso para construir el árbol de decisión y es la clave del método no paramétrico CART (Izenman, 2008).

Sea Y una variable respuesta y sean p variables predictoras x_1, x_2, \dots, x_p , donde las x 's son tomadas fijas y Y es una variables aleatoria. El problema estadístico es establecer una relación entre Y y las x 's de tal forma que sea posible predecir Y basado en los valores de x 's. Matemáticamente, se quiere estimar la probabilidad condicional de la variables aleatorias Y (Sepúlveda y Correa, 2013),

$$P[Y = y | x_1, x_2, \dots, x_p]$$

cuando la variable Y es discreta, o un funcional de su probabilidad tal como la esperanza condicional

$$E[Y | x_1, x_2, \dots, x_p]$$

cuando la variable Y es continua.

Otra de las características a destacar del modelo es “la poda”, se trata del sobreajuste general que se hace sobre éste, cortando sucesivamente ramas o nodos terminales hasta encontrar el tamaño deseado del árbol (Serna, 2009).

Estos árboles permiten explorar resultados y visualmente determinar cómo trabaja el modelo. Los resultados visuales ayudan a encontrar subgrupos específicos y relaciones que tal vez no se encuentren utilizando estadísticos tradicionales. Dado que los árboles de clasificación dividen los datos en ramas y nodos, se podrá ver fácilmente cómo se dividen y dónde terminan los grupos (Serna, 2009).

2.3.3 Diagnóstico de los modelos

En el estudio de cualquier modelo estadístico, se debe contrastar que se verifican las hipótesis básicas o estructurales del modelo cuando este sea de tipo paramétrico, establecidas a priori sobre los errores del modelo:

- *Bondad de ajuste*
- *Normalidad*
- *Homocedasticidad del error*
- *Independencia de las observaciones*

Dado que los errores del modelo son desconocidos, y que no se lleva a cabo el proceso de validación, las hipótesis anteriores deben estudiarse a partir de los residuos. Para evaluar el ajuste de los modelos desarrollados se utilizan dos conceptos; la bondad de ajuste y el ajuste a la recta de regresión de la nube de puntos (análisis gráficos). Además, en los modelos de regresión múltiple se estudia la correlación entre las variables mediante el análisis de multicolinealidad.

- ***Análisis de la bondad de ajuste***

Una de las medidas más comunes de evaluación del ajuste es el coeficiente de determinación ajustado R^2_{adj} ; representa el grado de predicción de una variable con la información existente de otra y la relación que se establece entre ambas (Ecuación 4). Este coeficiente es un criterio de valoración de la capacidad de explicación de los modelos de regresión y representa el porcentaje de la varianza explicado por la variable independiente (Pardo y Ruiz, 2002).

El coeficiente de determinación no informa de si se ha seleccionado el mejor conjunto de variables independientes ni si hay sesgo debido a la omisión de alguna variable. Para realizar esa comprobación, se realiza el análisis de regresión multivariante (*stepwise*). Este parámetro no sólo mide la capacidad explicativa de un modelo, sino que además permite elegir entre varios modelos cual es el más adecuado. Así los modelos tienen la misma variable dependiente y el mismo número de variables explicativas, será más adecuado el que tenga un coeficiente de determinación mayor (Rodríguez, 2016).

$$R^2_{adj} = 100 \left(\left(1 - \frac{n-1}{n-p-1} \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right) \right) \% \text{ [Ecuación 4]}$$

Siendo R^2 el coeficiente de determinación, donde y_i es el valor observado de Y , \widehat{y}_i es el valor predicho con el modelo ajustado, y p es el número de variables independientes incluidas en el modelo. Tiene en cuenta el número de variables regresoras y no tiene por qué crecer al introducir nuevas variables.

▪ **Gráfico de residuos**

El modelo teórico de la regresión lineal múltiple se define mediante la siguiente expresión matricial:

$$Y = X \beta + \varepsilon \rightarrow \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{11} & \dots & X_{1k} \\ \vdots & \dots & \vdots \\ X_{1n} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad [\text{Ecuación 5}]$$

- La hipótesis que planta es que ε_j son variables aleatorias con media 0 e independientes de las X_j .
- Se cumpla el principio de homocedasticidad en el que ε_j tiene varianzas iguales (σ^2)
- No autocorrelación, ε_j son incorreladas entre sí

La comprobación de que se cumplen estas condiciones se realiza mediante el análisis de los gráficos de residuos. El histograma de los residuos permite comprobar gráficamente la hipótesis de normalidad; aspecto que deberá tenerse en cuenta para la interpretación de los resultados de la inferencia estadística. Por otro lado, el diagrama de dispersión de puntos (*scatterplot*) analiza los residuos estandarizados en función de la variable dependiente para determinar la existencia de correlación entre ellos. Finalmente, el gráfico normal 'P-P Plot' de residuos estandarizados analiza el ajuste de los mismos a la recta de regresión (Pardo y Ruiz, 2002).

▪ **Test de colinealidad**

Una de las hipótesis del modelo de regresión múltiple establece la no existencia de una relación lineal entre los regresores, es decir, que no existe multicolinealidad en el modelo (no existe correlación entre predictores) (Rodríguez, 2016). Por ello, una de las medidas para detectar la multicolinealidad en el conjunto estadístico es el VIF (*Variance Inflation Factor*). Este estadístico mide en qué medida la varianza de los coeficientes de regresión estimados ha sido

inflada, en comparación con un contexto en el que las variables predictoras no están linealmente relacionadas (Ramírez *et al.*, 2005).

La multicolinealidad es problemática debido a que puede aumentar la varianza de los coeficientes de la regresión, lo que haría que fuesen inestables y difíciles de interpretar.

$$VIF = \frac{1}{1-R^2} \text{ [Ecuación 6]}$$

- VIF = 1 No correlacionados
- $1 < VIF < 5$ Moderadamente correlacionados
- VIF > 5 a 10 Altamente correlacionados

La existencia de valores de VIF mayores a 10 determinan la existencia de una multicolinealidad grave (Mandeville, 2008).

3. Resultados.

El centro de descargas del CNIG (Centro Nacional de Información Geográfica) tiene a su disposición nubes de puntos en formato *LAZ de España. A partir de esta adquisición gratuita se obtuvo información LiDAR (elevaciones, intensidades y retornos de los pulsos láser) referente a las 54 parcelas experimentales. El procesado de dichos datos se realizó mediante el software Fusion, al tratarse de un software de código libre, considerado como uno de los más completos para aplicaciones forestales.

Previo a este tratamiento, se contaba con una base de datos procedente del inventario de campo realizado en 2012 sobre 54 parcelas de plantación con medidas dasométricas de masa (altura (cm), diámetro normal (cm) y diámetro basal (cm)) y medidas de estimación de biomasa (peso verde (t/ha) y peso seco (t/ha)). Los resultados obtenidos son un producto del tratamiento estadístico de una base de datos que aúna el inventario de campo con variables por parcela de ensayo, y las métricas LiDAR obtenidas a través del procesado de la nube de puntos mediante el comando *CloudMetrics*, también por parcela.

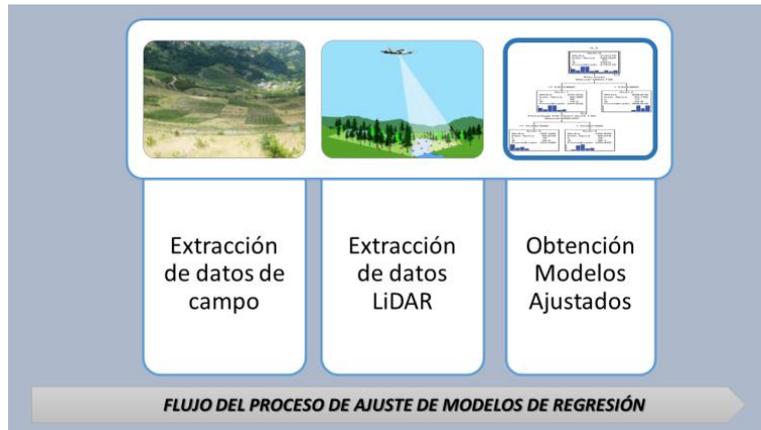


Figura 11: Diagrama de trabajo desarrollado en el estudio.
Fuente: propia.

3.1 Resultados de las correlaciones de Pearson.

El análisis de las correlaciones de Pearson arrojó resultados altamente significativos para las variables dasométricas inventariadas en campo (Tabla 3) y las variables de estimación de biomasa (Tabla 4) a un nivel de significación del 95% de probabilidad ($p \leq 0,05$). Los valores obtenidos muestran una correlación directa y positiva entre las variables de masa y las variables LiDAR, no existiendo correlaciones negativas entre las principales variables LiDAR correlacionadas.

En cuanto a la variable LiDAR que destaca sobre las demás, por su coeficiente de correlación de Pearson, es la elevación media, con un valor 0.89 tanto para la altura medida en campo como para el peso verde estimado. En relación a la variable con coeficiente de Pearson menor es el porcentaje de los primeros retornos sobre 1 con un valor de 0.69 para al diámetro basal medido por cepa.

Tabla 3: Coeficientes de correlación de Pearson para las variables dasométricas.

Variable de campo	Variable LiDAR	Correlación Pearson
H (altura cm)	<i>Elev maximum</i>	0,81
	<i>Elev mean</i>	0,89
	<i>Percentage first returns above 1.00</i>	0,74
D₂₀ (diámetro basal cm)	<i>Elev mean</i>	0,79
	<i>Elev máximo</i>	0,73
	<i>Percentage first returns above 1.00</i>	0,69
D₁₃₀ (diámetro normal cm)	<i>Elev maximum</i>	0,77
	<i>Elev mean</i>	0,82
	<i>Elev mode</i>	0,72

Tabla 4: Coeficientes de correlación de Pearson para las variables de estimación de biomasa.

Variable de campo	Variable LiDAR	Correlación Pearson
PV (peso verde kg/ha)	<i>Elev maximum</i>	0,81
	<i>Elev mean</i>	0,89
	<i>Percentage first returns above 1.00</i>	0,74
PS (peso seco kg/ha)	<i>Elev mean</i>	0,79
	<i>Elev maximum</i>	0,73
	<i>Percentage first returns above 1.00</i>	0,69

3.2 Resultados de modelos de regresión simple.

Los modelos ajustados, tanto para el conjunto de la plantación, como especificado por los clones empleados en el ensayo, se observan en la tabla 5 y en la tabla 6. El análisis se centra en la regresión general del conjunto de parcelas en un primer lugar, y seguidamente realizando una diferenciación por las especies ensayadas.

Para el conjunto de la plantación se observa que la variable LiDAR de elevación media es aquella que explica las variables de inventario en campo (altura (cm), diámetro normal (cm) y diámetro basal (cm)), mientras que las de estimación de biomasa (peso verde (t/ha) y peso seco (t/ha)) son explicadas por el porcentaje de todos los retornos sobre 1 (Tabla 5). Se debe añadir, que, aunque sea la altura la variable de mejor coeficiente de determinación ajustado, el error medio absoluto es elevado y por tanto devalúa su ajuste.

Tabla 5: Modelos de regresión simple para el conjunto del ensayo.

Variable Dependiente	Tipo de modelo	Variable Independiente	Parámetro del modelo	MAE	Coefficiente de correlación	R ² ajustado (%)
H (m)	<i>Double squared</i>	β_0	0,873492	3,93	0,94	83,43
		<i>Elev mean</i>	1,55261			
D₂₀ (cm)	<i>Squared Y</i>	β_0	-1,30034	1,57	0,83	68,70
		<i>Elev mean</i>	2,46665			
D₁₃₀ (cm)	<i>Squared Y</i>	β_0	-1,50919	0,83	0,88	77,46
		<i>Elev mean</i>	1,62471			
PV (t/ha)	<i>Logaritmico Y square root X</i>	β_0	0,0347361	0,52	0,83	68,60
		<i>Percentage of all returns above 1.0</i>	0,513635			
PS (t/ha)	<i>Double squared root</i>	β_0	0,22962	0,55	0,85	70,92
		<i>Percentage of first returns above 1.0</i>	0,544741			

En cuanto a los resultados obtenidos por especie (Tabla 6), para las variables de inventario (altura (cm), diámetro normal (cm) y diámetro basal (cm)) se obtienen ajustes con los percentiles de las elevaciones salvo para el clon *Inger* que lo hace con el porcentaje de todos los retornos sobre 1. En el caso de las variables de estimación de biomasa (peso verde (t/ha) y peso seco (t/ha)) sucede un caso diferente; para el clon *Bjor* e *Inger* entran en el modelo de regresión los porcentajes de retornos mientras que el clon *Olof* sigue utilizando las variables de percentiles de elevaciones.

Señalar para finalizar, que los errores (MAE) disminuye cuando se tratan los datos por especie de ensayo, en comparación con el análisis general del área de estudio. Esto puede deberse a que no hay el mismo número de pulsos/m² para cada parcela, ya que parcelas del clon *Olof* que tiene mayor producción tienen una densidad de puntos mayor que el clon *Bjor* que es el de menor crecimiento.

Tabla 6: Modelos de regresión simple especificados por especie de plantación para las variables dasométricas de inventario.

Variable Dependiente	Clon	Tipo de modelo	Variable Independiente	Estimación del parámetro	MAE	Coefficiente de correlación	R ² ajustado (%)	
H (m)	Bjor	Squared	β_0	0,820082				
		root Y – squared X	Elev P50	0,16365	0,17	0,80	62,19	
	Inger	Double reciprocal	β_0	0,398016				
			Percentage of all returns above 1.0	0,248287	0,10	0,90	81,05	
	Olof	Square X	β_0	2,18137				
			Elev P50	0,151876	0,56	0,91	82,47	
D₂₀ (cm)	Bjor	Squared	β_0	0,774154				
		X	Elev P25	0,435969	0,40	0,75	53,99	
	Inger	Double reciprocal	β_0	-	0,0164221	0,11	0,89	78,34
			Elev P75	1,23579				
	Olof	Square root Y – square X	β_0	1,21762				
			Elev P50	0,0231265	0,09	0,89	78,51	
D₁₃₀ (cm)	Bjor	Squared	β_0	0,246599				
		X	Elev P40	0,280162	0,24	0,81	63,94	
	Inger	S-curve Model	β_0	1,0035				
			Elev P60	1,51134	0,24	-0,80	62,61	
	Olof	Square root Y – square X	β_0	1,04337				
			Elev P50	0,0215818	0,08	0,91	82,15	

Tabla 7: Modelos de regresión simple especificados por especie de plantación para las variables de estimación de biomasa.

Variable Dependiente	Clon	Tipo de modelo	Variable Independiente	Estimación del parámetro	MAE	Coefficiente de correlación	R ² ajustado (%)	
PV (t/ha)	Bjor	Double squared root	β_0	0,450479	0,58	0,86	66,07	
			Percentage of all returns above 1.0	0,73409				
	Inger	Multiplicative (ln)	Percentage of first returns above mode	β_0	-0,176831 0,883665* ln(7,278481)	0,42	0,86	71,98
PS (t/ha)	Olof	Square root Y – square X	β_0	1,47552	0,68	0,85	70,26	
			Elev P95	0,0982484				
	Bjor	Double squared root	Percentage of all returns above 1.0	β_0	0,23304 0,56694	0,44	0,85	70,23
PS (t/ha)	Inger	Multiplicative (ln)	Percentage of first returns above mode	β_0	-0,909199 0,920332* ln(7,278481)	0,43	0,86	71,87
			Olof	Square root Y – square X	Elev P75	β_0	1,31891 0,09557125	0,54

3.3 Resultados de modelos de regresión múltiple.

La determinación de modelos de regresión múltiple se basa en el procedimiento de regresión paso a paso o 'stepwise' mediante el software de análisis estadístico SPSS.

En la primera fase, referente a los modelos de regresión paso a paso, se selecciona el número de variables que entran en el modelo y cumplen las condiciones de no multicolinealidad entre las variables regresoras (observando que el VIF no alcanza el umbral de multicolinealidad en ningún caso) y que el error

sea asumible para el modelo. En una segunda fase de análisis de resultados se observa que en la tabla 8 dan valores superiores a 0.75 en R^2 .

Destaca un patrón constante, y es en las variables de campo (Tabla 8) (altura (cm), diámetro normal (cm), diámetro basal (cm)) entran en los modelos de regresión múltiple los percentiles de las elevaciones y el conteo de los primeros retornos sobre 0.5. Mientras que en el caso de las variables de estimación de biomasa (Tabla 8) (peso verde (t/ha) y peso seco (t/ha)) son el porcentaje de todos los retornos sobre la media y los percentiles de las elevaciones las variables elegidas para el modelo.

En el caso de la regresión múltiple por especie no destaca la repetición de algún patrón en lo referente a variables LiDAR. Si se puede destacar que en cuanto a las variables de inventario de campo (Tabla 9) (altura (cm), diámetro normal (cm), diámetro basal (cm)) las especies de clon *Inger* y *Olof* elaboran modelos con pocas variables, salvo el clon *Olof* para el diámetro normal que introduce cinco variables. También hay que decir que la altura sigue siendo la variable que mayor error obtiene.

En cuanto a las variables de estimación de biomasa (Tabla 10) (peso verde (t/ha) y peso seco (t/ha)) el *Bjor* es el clon con mejores resultados (R^2 de 0.9 aprox.) pero, por ejemplo, en la estimación del peso seco entran en el modelo cinco variables mientras que para el clon *Olof* solo entraría una para un R^2 de 0.66.

Tabla 8: Regresión paso a paso general en SPSS.

Variable Dependiente	R^2	Variables Independientes	B	Std. Error	VIF
H (m)	0,90	(Constant)	349,62	77,12	
		Elev P50	81,02	7,62	1,88
		Total return count	-0,85	0,19	1,74
		Return 1 count above 0.50	0,58	0,17	1,21
D₂₀ (cm)	0,75	(Constant)	2,26	0,56	
		Elev P50	0,33	0,05	1,88
		Total return count	-0,01	0,01	1,74
		Return 1 count above 0.50	0,01	0,01	1,21
D₁₃₀ (cm)	0,81	(Constant)	1,32	0,41	
		Elev P30	0,38	0,04	1,95
		Return 1 count above 0.50	0,01	0,01	1,19
		Total return count	-0,01	0,01	1,84
PV (t/ha)	0,75	(Constant)	-4,89	1,95	
		Elev P60	4,68	0,85	1,76
		Percentage all returns above mean	0,59	0,12	1,76
PS (t/ha)	0,77	(Constant)	-2,70	0,99	
		Elev P50	2,75	0,48	1,79
		Percentage all returns above mean	0,35	0,06	1,79

Tabla 9: Regresión paso a paso por especie en SPSS para variables de inventario de campo.

Variable Dependiente	R ²	Variables Independientes	B	Std. Error
H (m)	0.91	(Constant)	111,598	17,62
		Percentage all returns above 1.00	32,433	3,48
		Return 1 count above 0.50	-5,932	0,92
		Elev skewness	-115,916	20,71
		Elev variance	71,163	22,17
	0.68	(Constant)	135,973	20,95
		Percentage first returns above 1.00	4,112	0,67
	0.87	(Constant)	-916,104	384,74
		Elev P50	123,375	12,27
		Int P05	89,991	34,98
D₂₀ (cm)	0.93	(Constant)	0,878	0,19
		Total first returns	-0,009	0,01
		Elev P60	1,252	0,13
		Canopy relief ratio	2,495	0,64
	0.63	Int L4	-0,327	0,11
		(Constant)	0,247	0,29
	0.92	Elev P60	0,856	0,15
		(Constant)	-4,255	1,44
		Elev P50	0,553	0,05
	D₁₃₀ (cm)	0.89	Int P10	0,603
(Constant)			0,521	0,16
Total first returns			-0,006	0,01
Elev P50			1,007	0,12
0.77		Canopy relief ratio	1,191	0,54
		(Constant)	2,052	0,56
		Elev L1	0,449	0,13
0.95		Total first returns	-0,006	0,01
		Int skewness	0,215	0,09
		(Constant)	-33,601	12,78
	Elev P50	1,418	0,32	
	Int P10	0,579	0,10	
	Int P30	-0,258	0,06	
	Elev P40	-1,027	0,35	
	Int P01	3,077	1,31	

Tabla 10: Regresión paso por paso en SPSS para variables de estimación de biomasa

VARIABLE DEPENDIENTE	R ² ajustado (%)	Variables Independientes	B	Std. Error
PV (t/ha)	0.92	(Constant)	2,777	0,819
		Percentage all returns above 1.00	2,326	0,366
		All returns above 1.00	-0,522	0,111
		Elev skewness	-3,992	0,920
		Elev L4	9,128	3,725
PV (t/ha)	0.77	(Constant)	-1,355	2,170
		(All returns above mode) / (Total first returns) * 100	0,489	0,066
		Int kurtosis	1,092	0,374
PV (t/ha)	0.84	(Constant)	-67,239	12,760
		Elev P75	100,216	25,310
		Int P90	1,827	0,496
		Elev P70	-89,366	25,076
PV (t/ha)	0.90	(Constant)	0,156	0,511
		Percentage all returns above 1.00	1,245	0,208
		All returns above 1.00	-0,288	0,061
		Elev skewness	-2,632	0,566
		Elev kurtosis	0,768	0,267
PV (t/ha)	0.79	(Constant)	-1,184	1,190
		(All returns above mode) / (Total first returns) * 100	0,283	0,036
		Int kurtosis	0,633	0,205
PV (t/ha)	0.66	(Constant)	-8,093	3,975
		Elev P75	4,917	0,835

3.4 Resultados de diagnóstico de los modelos

En la evaluación de la diagnosis de los modelos de regresión, se evaluó la homogeneidad de varianzas mediante los gráficos obtenidos en SPSS de ZRESID Y ZPRED (valores pronosticados tipificados y residuos tipificados, Anexo IV).

Dicho estudio se llevó a cabo para el clon *Olof*, ya que al ser el clon de mayor productividad es el que interesa estudiar y ajustar lo más preciso que se pueda sus modelos de regresión. No obstante, en un primer análisis previo a la obtención de modelos se observó que dichas premisas se cumplían para el conjunto de los datos muestrales (Figura 12).

Los resultados dan valores de ausencia de datos atípicos y dispersos en la muestra, así como una linealidad dentro del rango muestral. El histograma muestra normalidad en los datos a través de su forma de campana para todos los conjuntos de datos estudiados. Además, no se encontraron valores dispersión o fuera del rango de la muestra en el análisis del diagrama de dispersión, y los valores se ajustan a la recta. Los mismos resultados se cumple en el clon *Olof* (Anexo IV) en el que se analizan el histograma de residuos y el gráfico de dispersión.

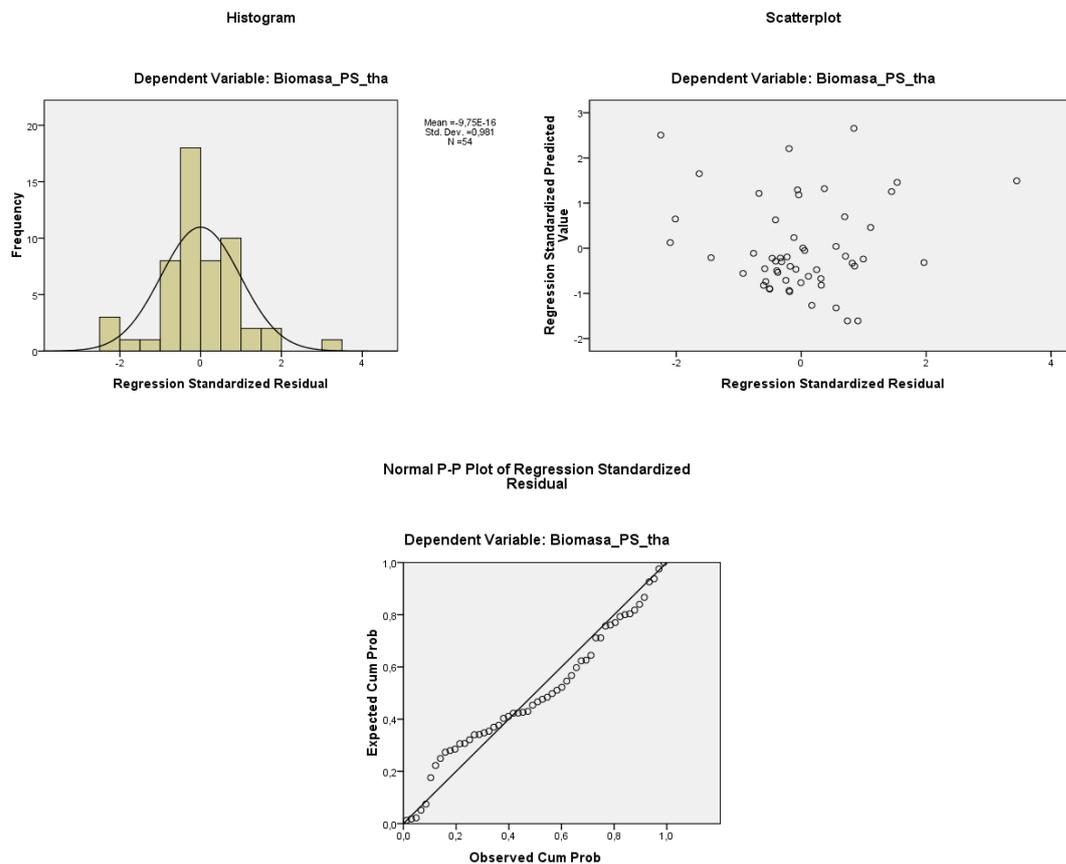


Figura 12: Histograma, gráfico de dispersión y ajuste de valores a la recta para la variable de estimación de biomasa peso seco (t/ha) dónde se observa que los datos cumplen las hipótesis establecidas.

3.5 Resultados del árbol de clasificación y regresión

Para la interpretación de los árboles de clasificación y regresión se eligen los más representativos del proceso, en este caso se analiza el peso seco para el clon *Olof* (Figura 9) y para el conjunto del ensayo (Figura 12) por ser la variable de cuantificación de biomasa más utilizada dentro de la comunidad científica. También se analizan la altura (Figura 10) y el diámetro normal (Figura 11) para el conjunto del ensayo.

En primer lugar y referido al peso seco de la especie *Olof* el árbol presenta dos niveles ramificados en cuatro nodos terminales. El primer nodo describe la variable dependiente, en este caso el peso seco del clon *Olof*. Esa variable se ramifica en dos nodos en función de la elevación media que es la variable principal predictora, siendo esta capaz de predecir un valor de 14.062 t/ha de biomasa seca media. *Improvement* es el valor de mejora de esa variable sobre el modelo.

Los nodos 1 y 2 se dividen a su vez en dos nuevos nodos, el nodo 1 se refiere a los datos menores a 3.64 t/ha de biomasa seca y el nodo 2 a datos mayores. Al encontrarse el grueso del rango muestral explicado por el nodo 2, se continúa el análisis observado que es la elevación mínima la variable principal predictora. En este caso, y a modo de resumen sólo entrarían tres variables al modelo.

En el caso de la altura del ensayo realizado continúa siendo la elevación media la variable predictora del nodo principal y para el caso del nodo 1 (ya que el 2 es un nodo terminal) el porcentaje de los primeros retornos sobre 1 sería la segunda variable del modelo.

Para el diámetro normal del ensayo de plantación sucede el mismo patrón que en el caso anterior pero en este modelo la segunda variable predictora es el conteo de todos los retornos sobre 0.5.

Se concluye, por tanto, que los métodos de análisis paramétrico de tipo CART asignan a la variable LiDAR elevación media el peso más fuerte como variable predictora por defecto para el ajuste de variables de inventario de campo y de estimación de biomasa.

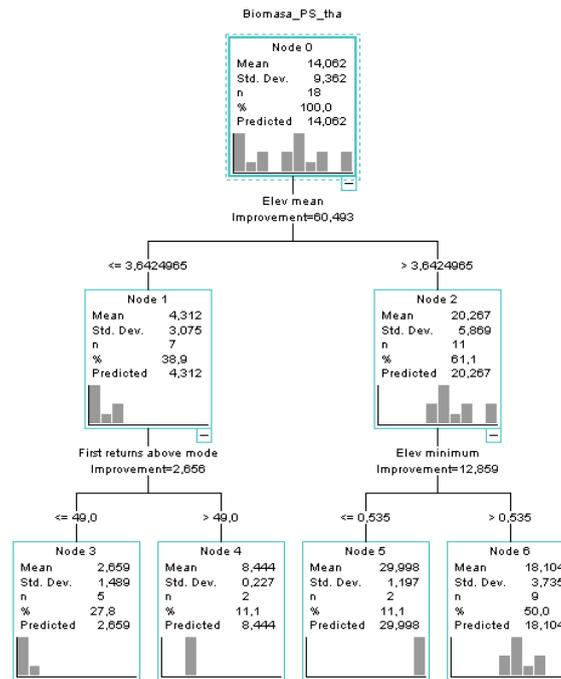


Figura 13: CART de peso seco (t/ha) para Olof

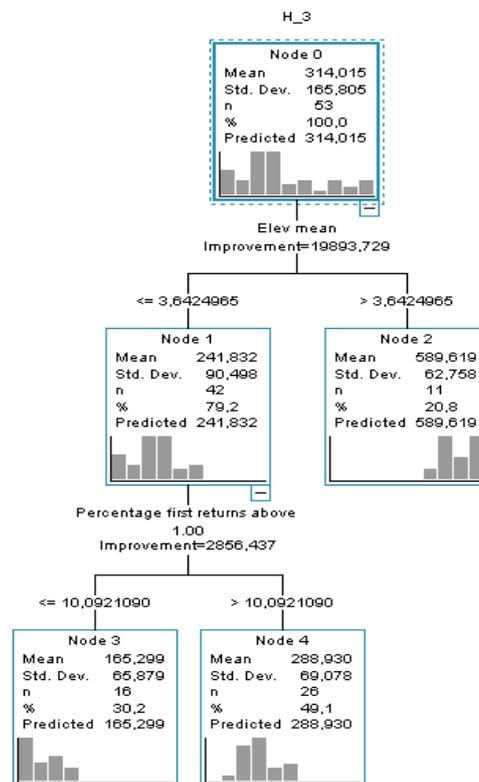


Figura 14: CART de altura (cm) para el ensayo

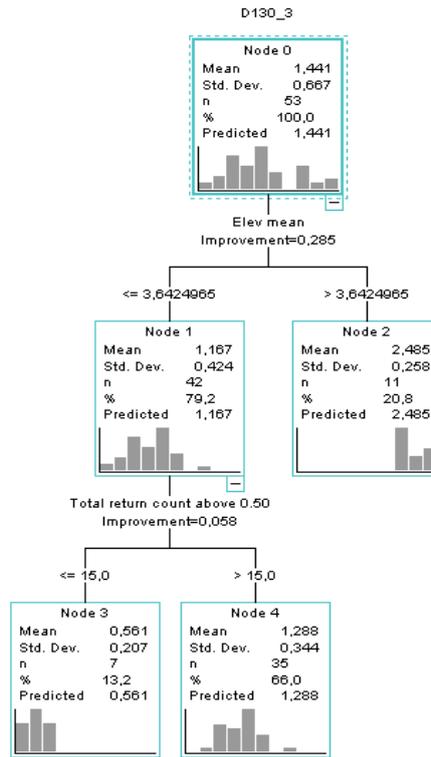


Figura 11: CART diámetro normal (cm) para el ensayo

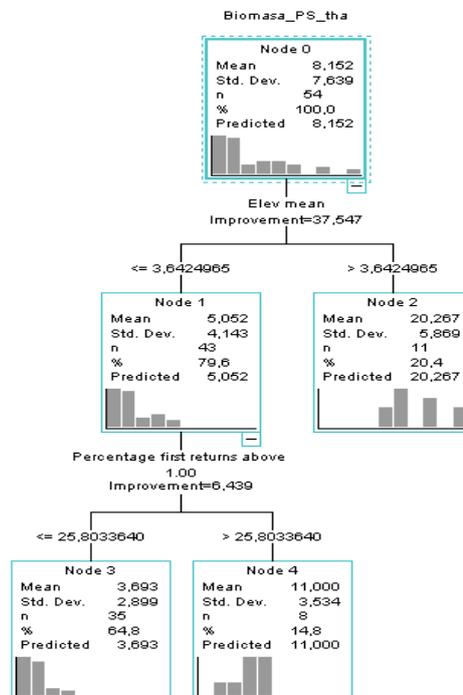


Figura 12: CART de peso seco (t/ha) para el ensayo

4. Discusión

En cuanto los resultados obtenidos referentes a la capacidad predictiva de los modelos se consideran altamente significativos en relación con otros autores.

Se deduce en un primer lugar, partiendo de la literatura estudiada, que la disposición de las líneas plantación en el diseño experimental llevado a cabo en Mozquita - Langreo resulta favorable para la obtención de un modelo digital de terreno de alta precisión. En aquellas zonas en las que la cobertura del suelo tenga un alto grado de espesura los retornos procedentes del suelo no generarán un MDT preciso (Lefsky *et al.*, 2002).

Cuando se obtiene un MDT de alta fiabilidad, las métricas obtenidas a través del procesado de la nube de puntos se asemejan más a la realidad, aunque dicha nube de puntos sea de baja densidad como en el caso de estudio, ya que las alturas obtenidas en el procesado están referidas a ese modelo generado.

Además, la orografía favorable característica del ensayo, con pendientes menores al 30% descarta errores por desplazamientos en los retornos debidos a terrenos demasiado escarpados (García, 2014).

En cuanto a la funcionalidad de la nube de puntos procedente del PNOA de 0,5 pulsos/m², y comparando con otros estudios se puede determinar que tiene una alta funcionalidad.

Estudios de ajuste de modelos de regresión para *Pinus radiata* en Galicia (Ferreiro *et al.*, 2012) consiguen ajustes para la altura media de 0.759 de coeficiente de determinación para densidades de 8 pulsos/m², y para plantaciones de *Eucalyptus globulus* en predicciones de biomasa de copa entre 0,5 y 0,7 para densidades de puntos de 0,5 y 4 pulsos/m² (Ferreiro *et al.*, 2013). Con métodos como el análisis de componentes principales y densidades de más de 4 pulsos/m² no superan en R² de 0.7 en estudios de masas de Douglas en EE.UU (Li *et al.*, 2008).

En masas de *Pinus* y *Quercus* en Texas (EE.UU) estimaciones de biomasa aérea mediante modelos de regresión lineal simple obtienen valores de R² de 0.9 para densidades de puntos de 2.6 pulsos/m² (Zhao *et al.*, 2008) mientras que en el presente ensayo se consiguen para el clon *Olof* (*Salix viminalis* x (*Salix schwerinii* x *Salix viminalis*)) coeficientes de 0.82 con un nube de puntos menor para el diámetro normal.

En el caso de ajustes de regresión múltiple (Condés *et al.*, 2013) para masas de *Pinus sylvestris* en la provincia de Soria a una densidad de 2 pulsos/m² obtiene valores de R² de 0.89 aproximadamente para distintas variables de masa estudiadas, mientras que en el ensayo realizado se pueden citar valores de 0.84 para el cálculo del peso verde (t/ha) para el clon *Olof*.

Con una densidad de puntos baja de 0.5 pulsos/m² se consiguen modelos mejor o igual de representativos que autores que trabajan con vuelos a densidades mayores, además se trata de información gratuita y con total disponibilidad.

Para finalizar, los modelos de ajuste paramétricos introducidos en el ensayo, permiten estudiar y comprender el comportamiento explicativo de las variables LiDAR con buenos ajustes, un método hasta ahora no utilizado en este tipo de estudios pero que resulta de gran utilidad.

5. Conclusiones

El estudio permitió evaluar la capacidad de la información proveniente del procesado de nubes de puntos LiDAR del IGN para la estimación de las principales variables de interés en la gestión forestal, en particular en lo correspondiente a los inventarios de campo en terrenos dedicados a cultivos energéticos. En vista de los resultados obtenidos, de este trabajo derivan las siguientes conclusiones:

- I. El estudio concreta los resultados de productividad obtenidos mediante el análisis de datos de campo para los tres clones de sauce del estudio. Se determina que el clon *Olof* (*Salix viminalis* x (*Salix schwerinii* x *Salix viminalis*)) es aquel que mayores crecimientos alcanza, siendo a su vez la especie que mejor se ajusta a los modelos de regresión obtenidos.
- II. Los modelos de regresión simple resultan prácticos y operativos obteniendo valores de R^2_{adj} altos para las variables de inventario y de estimación de biomasa. El ajuste de modelos paramétricos, específicamente el CART, mejora la flexibilidad del modelo y ayuda en la interpretación de los resultados.
- III. La disponibilidad de una baja densidad de puntos LiDAR no supone un inconveniente para el diseño de inventarios forestales en plantaciones con fines bioenergéticos, ni tampoco para la estimación de sus variables dasométricas. Esto se ve favorecido por la disposición de las líneas de plantación, considerando finalmente este método como una herramienta de gestión que facilita el trabajo del gestor forestal.
- IV. Dichas líneas de plantación, junto a los tratamientos selvícolas asociados que mantienen la cubierta del suelo prácticamente desnuda, ayudan a que se generen modelos digitales del terreno precisos que hagan que la densidad de puntos necesaria para obtener buenos resultados no deba ser tan elevada. Este hecho es una gran ventaja, ya que a pesar de que el vuelo en sensor aerotransportado requiere un alto coste, no es necesario que afine en precisión para obtener unos resultados de alta fiabilidad.

6. Bibliografía

BARÓN LÓPEZ, F. J.; TÉLLEZ MONTIEL, F. Apuntes de bioestadística. Apuntes de Bioestadística, 2000.

BREIMAN, Leo, et al. Classification and regression trees. CRC press, 1984.

CONDÉS, S., et al. Influencia del inventario de campo en el error de muestreo obtenido en un inventario con tecnología Lidar. En Congresos-CARGA FINAL. 2013.

CONTRERAS, Sixto Jesús Álvarez. Estadística aplicada: teoría y problemas. 2000.

CRESPO, Miguel Fabra. Aplicaciones de la tecnología LIDAR al sector forestal y comparación de costes frente a metodologías tradicionales. Revista Montes, 2012, no 110, p. 33-37.

CRESPO-PEREMARCH, Pablo, et al. Análisis temporal de la estructura forestal mediante métricas derivadas de LiDAR full-waveform. En XVI Congreso de la Asociación Española de Teledetección. 2015.

DECONINCK, Eric, et al. Classification tree models for the prediction of blood-brain barrier passage of drugs. Journal of chemical information and modeling, 2006, vol. 46, no 3, p. 1410-1419.

DEL CARMEN GARCÍA, María, et al. Criterios de información y predictivos para la selección de un modelo lineal mixto. SaberEs, 2014, no 6.

DÍAZ, Pértegas; FERNÁNDEZ, Pita. Determinación del tamaño muestral para calcular la significación del coeficiente de correlación lineal. Unidad Epidemiol. Clínica y Bioestad, 2002, p. 1-6.

DONÉS, J.; CABRERA, M. Proyecto de 7ª Revisión de la Ordenación del monte número 2 del Catálogo de los de Utilidad Pública de la Provincia de Segovia "Pinar" de Valsaín. Unpublished management plan, Segovia, 2009.

GARCIA-GUTIERREZ, Jorge, et al. Evolutionary feature selection to estimate forest stand variables using LiDAR. International Journal of Applied Earth Observation and Geoinformation, 2014, vol. 26, p. 119-131.

GARCÍA-FERRER, Alfonso. Potencialidad y nuevas orientaciones en el uso de la teledetección en agricultura y silvicultura. Ambients: La revista del Ministerio de Medio Ambiente, 2013, no 105, p. 6-15.

GARCÍA, Mariano, et al. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. Remote Sensing of Environment, 2010, vol. 114, no 4, p. 816-830.

GARCÍA, M. 2014. Documento técnico: obtención de variables forestales a partir de datos LiDAR. Ministerio de Agricultura, Alimentación y Medio Ambiente. Disponible en:

http://www.magrama.gob.es/es/red-parques-nacionales/plan-seguimiento-evaluacion/documento-tecnico-obtencion-variables-lidar_tcm7-337536.pdf

GONZÁLEZ-FERREIRO, Eduardo; DIÉGUEZ-ARANDA, Ulises; MIRANDA, David. Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. *Forestry*, 2012, vol. 85, no 2, p. 281-292.

GONZÁLEZ-FERREIRO, Eduardo; DIÉGUEZ-ARANDA, Ulises; MIRANDA, David. Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. *Forestry*, 2012, vol. 85, no 2, p. 281-292.

GONZÁLEZ-OLABARRIA, José-Ramón, et al. Mapping fire risk in the Model Forest of Urbión (Spain) based on airborne LiDAR measurements. *Forest Ecology and Management*, 2012, vol. 282, p. 149-156.

HERNÁNDEZ CLEMENTE Rocío, VARO MARTÍNEZ M^a Ángeles. Fundamentos de la tecnología LiDAR. LiDAR y Estructura de Bosque. Métodos de Masa. ETS Ingenieros Agrónomos y de Montes (UCO-ETSIAM), 2016. Plataforma de formación on-line.

HÖFLE, Bernhard; PFEIFER, Norbert. Correction of laser scanning intensity data: Data and model-driven approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2007, vol. 62, no 6, p. 415-433.

HOLLAUS, Markus, et al. Airborne laser scanning of forest stem volume in a mountainous environment. *Sensors*, 2007, vol. 7, no 8, p. 1559-1577.

HUISING, E. J.; PEREIRA, LM Gomes. Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of photogrammetry and remote sensing*, 1998, vol. 53, no 5, p. 245-261.

HYYPÄ, Juha; INKINEN, Mikko. Detecting and estimating attributes for single trees using laser scanner. *The photogrammetric journal of Finland*, 1999, vol. 16, no 2, p. 27-42.

HYYPÄ, Juha, et al. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *Geoscience and Remote Sensing, IEEE Transactions on*, 2001, vol. 39, no 5, p. 969-975.

IZENMAN, A. *Modern multivariate statistical techniques*. New York: Springer, 2008.

KRAUS, Karl; PFEIFER, Norbert. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and remote Sensing*, 1998, vol. 53, no 4, p. 193-203.

LEFSKY, Michael A., et al. Lidar Remote Sensing for Ecosystem Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, 2002, vol. 52, no 1, p. 19-30.

LI, Yuzhen; ANDERSEN, Hans-Erik; MCGAUGHEY, Robert. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *Western Journal of Applied Forestry*, 2008, vol. 23, no 4, p. 223-231.

LIM, Kevin, et al. LiDAR remote sensing of forest structure. *Progress in physical geography*, 2003, vol. 27, no 1, p. 88-106.

MCGAUGHEY, Robert. Fusion/LDV: Software for LiDAR Data Analysis and Visualization. USDA Forest Service. Pacific Northwest Research Station. FUSION Manual. March 2010 – FUSION Version 2.80.

MAGDALENO MAS, Fernando; MARTINEZ ROMERO, Roberto. Aplicaciones de la teledetección láser (LIDAR) en la caracterización y gestión del medio fluvial. *Ingeniería Civil*, 2006, no 142, p. 1-15.

MALTAMO, Mustonen, et al. The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. *Canadian Journal of Forest Research*, 2004, vol. 34, no 9, p. 1791-1801.

MANDEVILLE, P. B. ¿Por qué se deben centrar las covariables en regresión lineal. *Ciencia UANL*, 2008, vol. 11, no 3, p. 300-305.

MARTÍN FERNANDEZ, S., et al. Guía completa de Statgraphics. Desde MS-DOS a Statgraphics Plus.(Complete Guide of STATGRAPHICS. From MS-DOS to Statgraphics Plus). Díaz de Santos, Madrid, 2001.

MAUNE, David Francis (ed.). Digital elevation model technologies and applications: the DEM users manual. Asprs Publications, 2007.

MITJANS, Eduard Nafria. Segmentación automática para árboles de decisión: estudio de la estabilidad y la complejidad del algoritmo. 1996. Tesis Doctoral. Universitat Politècnica de Catalunya.

MORENO-GIL, Jaime Serret. Procedimientos estadísticos con Statgraphics. ESIC Editorial, 1998.

MUNDT, Jacob T.; STREUTKER, David R.; GLENN, Nancy F. Mapping sagebrush distribution using fusion of hyperspectral and lidar classifications. *Photogrammetric Engineering & Remote Sensing*, 2006, vol. 72, no 1, p. 47-54.

MUTLU, Muge, et al. Mapping surface fuel models using lidar and multispectral data fusion for fire behavior. *Remote Sensing of Environment*, 2008, vol. 112, no 1, p. 274-285.

NÆSSET, Erik. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 2002, vol. 80, no 1, p. 88-99.

NÆSSET, Erik. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 2004, vol. 19, no 2, p. 164-179.

PACKALÉN, Petteri, et al. Comparison of individual tree detection and canopy height distribution approaches: a case study in Finland. *Proceedings of SilviLaser*, 2008, vol. 2008, p. 8th.

PARDO, Antonio; RUIZ, Miguel Ángel. *SPSS 11 Guía para el análisis de datos*. Madrid: McGraw-Hill, 2002.

RODRÍGUEZ PUERTA Paco. Ajuste y diagnóstico de modelos de regresión aplicados al inventario. *LiDAR y Estructura de Bosque. Métodos de Masa*. ETS Ingenieros Agrónomos y de Montes (UCO-ETSIAM), 2016. Plataforma de formación on-line.

RODRÍGUEZ PUERTA Paco. Conceptos básicos de estadística. *LiDAR y Estructura de Bosque. Métodos de Masa*. ETS Ingenieros Agrónomos y de Montes (UCO-ETSIAM), 2016. Plataforma de formación on-line.

RABADÁN, Manuel Ángel Valbuena. Determinación de variables forestales de masa y de árboles individuales mediante delineación de copas a partir de datos LIDAR Aerotransportado. Aplicación a las masas de *Pinus sylvestris* L. en Álava. 2014. Tesis Doctoral. Universidad de La Rioja.

RAMIREZ, Guillermo, et al. Detección gráfica de la multicolinealidad mediante el h-plot de la inversa de la matriz de correlaciones, 2005.

RIANO, David, et al. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment*, 2003, vol. 86, no 2, p. 177-186.

SÁNCHEZ, Javier. Guía para la aplicación del sistema nacional de cartografía de zonas inundables. *Jornada Retos y Oportunidades sobre evaluación y gestión de los riesgos de inundación*. Dirección General del Agua. Ministerio de Medio Ambiente y Medio Rural y Marino, 2010. Capítulo 2. 2.2.1 Introducción a la tecnología Lidar.

SERNA PINEDA, Sandra Carolina, et al. Comparación de árboles de regresión y clasificación y regresión logística, 2009. Tesis Doctoral. Universidad Nacional de Colombia.

SEPÚLVEDA, Juan Felipe Díaz; CORREA, Juan Carlos. Comparación entre árboles de regresión CART y regresión lineal. *Comunicaciones en Estadística*, 2013, vol. 6, no 2, p. 175-195.

WALLE, Inge Vande, et al. Short-rotation forestry of birch, maple, poplar and willow in Flanders (Belgium) I—Biomass production after 4 years of tree growth. *Biomass and bioenergy*, 2007, vol. 31, no 5, p. 267-275.

WANG, Cheng; GLENN, Nancy F. A linear regression method for tree canopy height estimation using airborne lidar data. *Canadian Journal of Remote Sensing*, 2008, vol. 34, no sup2, p. S217-S227.

WEHR, Aloysius; LOHR, Uwe. Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 1999, vol. 54, no 2, p. 68-82.

ZHAO, Kaiguang; POPESCU, Sorin; NELSON, Ross. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, 2009, vol. 113, no 1, p. 182-196.

ZOLKOS, S. G.; GOETZ, S. J.; DUBAYAH, R. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, 2013, vol. 128, p. 289-298.

Páginas web de consulta

- Web de descarga: www.cnig.com

Link: <http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do#selectedSerie>

[Consulta: 02/2016] ; [Consulta 04/2013]

7. ANEXO I: Ficha Técnica del Ensayo

LOCALIZACIÓN GEOGRÁFICA

Comunidad Autónoma: Principado de Asturias

Municipio: Langreo

Área de estudio: Mozquita

Sistema de coordenadas: ETRS89 Huso 30

Latitud: 43° 16' 13,56 Norte

Longitud: 5° 42' 03,69" Oeste



ENSAYO DE PLANTACIÓN

Especies de plantación:

- ❖ BJOR (*Salix schwerinii* x *Salix viminalis*)
- ❖ INGER (*Salix trianta* x *Salix viminalis*)
- ❖ OLOF (*Salix viminalis* x (*Salix schwerinii* x *Salix viminalis*))

Tipo de plantación: método de plantación de doble fila sueco con una densidad de plantación de 10.000 y 15.000 pies/ha.



BJOR



INGER



OLOF

CARACTERES CULTURALES

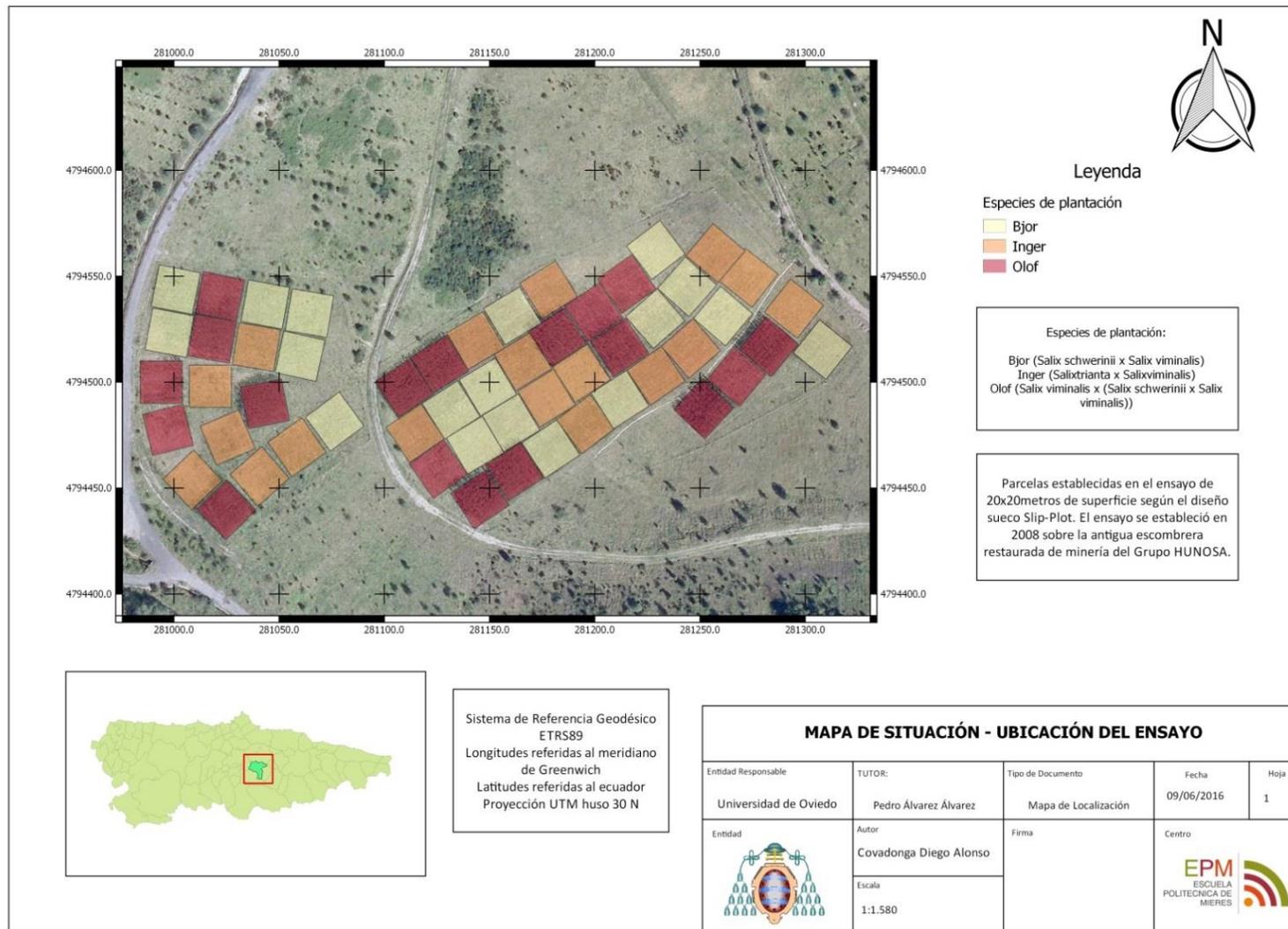
Dentro de la biomasa, se trabaja con cultivos energéticos de tipo lignocelulósico. Se denominan Short Rotation Coppice (SCR) por su alta capacidad de rebrote, rápidos crecimientos y con turnos de corta de pocos años.

8. ANEXO II: Variables obtenidas del comando CloudMetrics

Estadísticos LiDAR relacionados con la elevación, la intensidad y el número de retornos ⁴		
Total return count	Elev P05	Int L skewness
Total return count above 0.50	Elev P10	Int L kurtosis
Return 1 count above 0.50	Elev P20	Int P01
Return 2 count above 0.50	Elev P25	Int P05
Return 3 count above 0.50	Elev P30	Int P10
Return 4 count above 0.50	Elev P40	Int P20
Return 5 count above 0.50	Elev P50	Int P25
Return 6 count above 0.50	Elev P60	Int P30
Return 7 count above 0.50	Elev P70	Int P40
Return 8 count above 0.50	Elev P75	Int P50
Return 9 count above 0.50	Elev P80	Int P60
Other return count above 0.50	Elev P90	Int P70
Elev minimum	Elev P95	Int P75
Elev maximum	Elev P99	Int P80
Elev mean	Canopy relief ratio	Int P90
Elev mode	Elev SQRT mean SQ	Int P95
Elev stddev	Elev CURT mean CUBE	Int P99
Elev variance	Int minimum	Percentage first returns above 1.00
Elev CV	Int maximum	Percentage all returns above 1.00
Elev IQ	Int mean	$(\text{All returns above 1.00}) / (\text{Total first returns}) * 100$
Elev skewness	Int mode	First returns above 1.00
Elev kurtosis	Int stddev	All returns above 1.00
Elev AAD	Int variance	Percentage first returns above mean
Elev MAD median	Int CV	Percentage all returns above mean
Elev MAD mode	Int IQ	Percentage all returns above mode
Elev L1	Int skewness	$(\text{All returns above mean}) / (\text{Total first returns}) * 100$
Elev L2	Int kurtosis	$(\text{All returns above mode}) / (\text{Total first returns}) * 100$
Elev L3	Int AAD	First returns above mean
Elev L4	Int L1	First returns above mode
Elev L CV	Int L2	All returns above mean
Elev L skewness	Int L3	All returns above mode
Elev L kurtosis	Int L4	Total first returns
Elev P01	Int L CV	Total all returns

⁴ Tabla de relación de variables obtenidas en el procesado de la nube de puntos LiDAR; estadísticos referidos a retornos (returns), elevaciones (elev) e intensidades (int).

9. ANEXO III: Mapa de situación.



10. ANEXO IV: Gráficos de homogeneidad de varianzas.

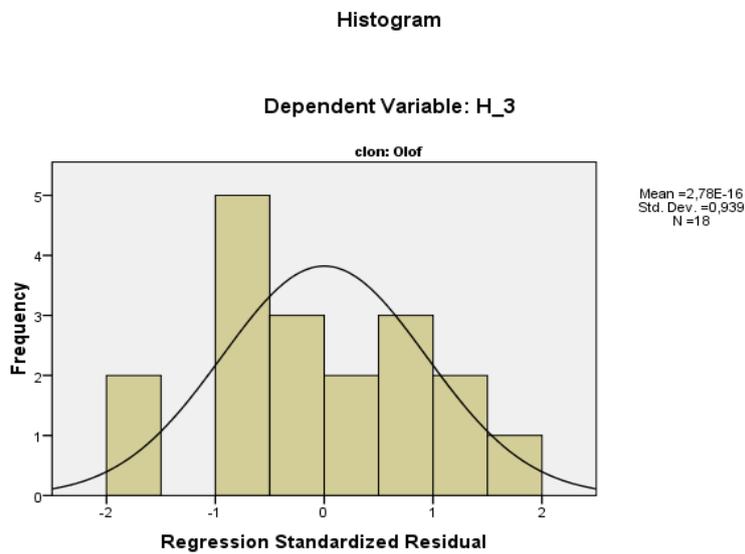


Gráfico1: Histograma de la variable altura (cm) del clon Olof

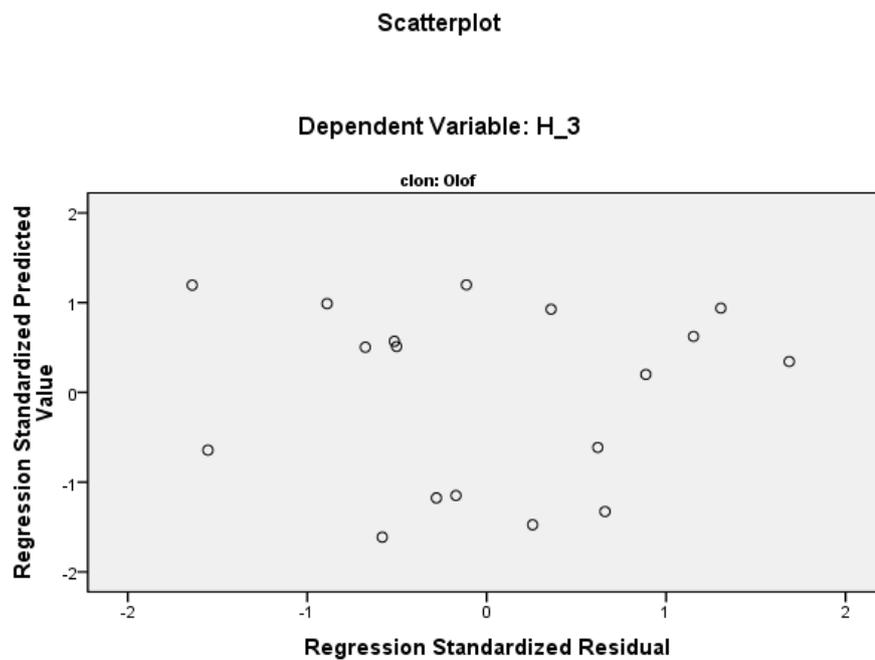


Gráfico2: Diagrama de dispersión de la variable altura (cm) del clon Olof

Histogram

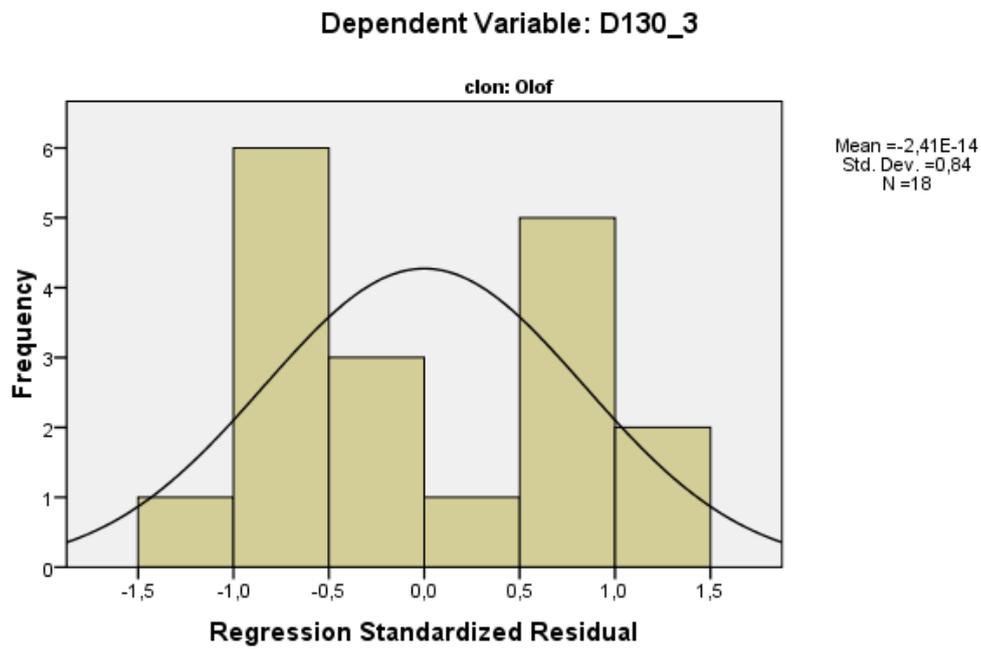


Gráfico 3: Histograma de residuos para la variable diámetro normal (cm)

Scatterplot

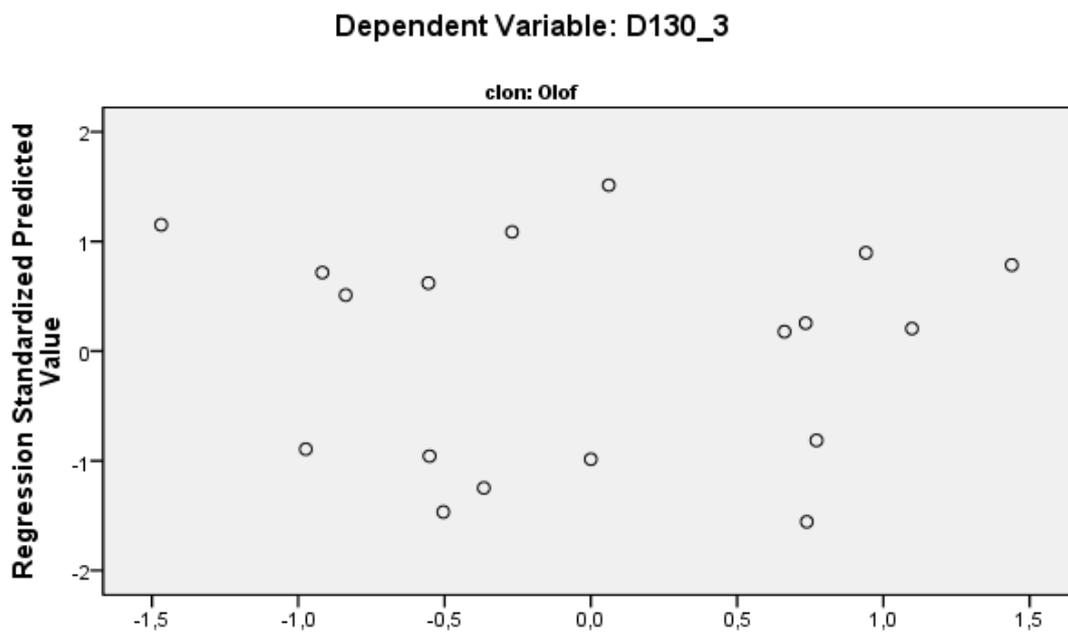


Gráfico 4: Gráfico de dispersión de residuos para el diámetro normal (cm)

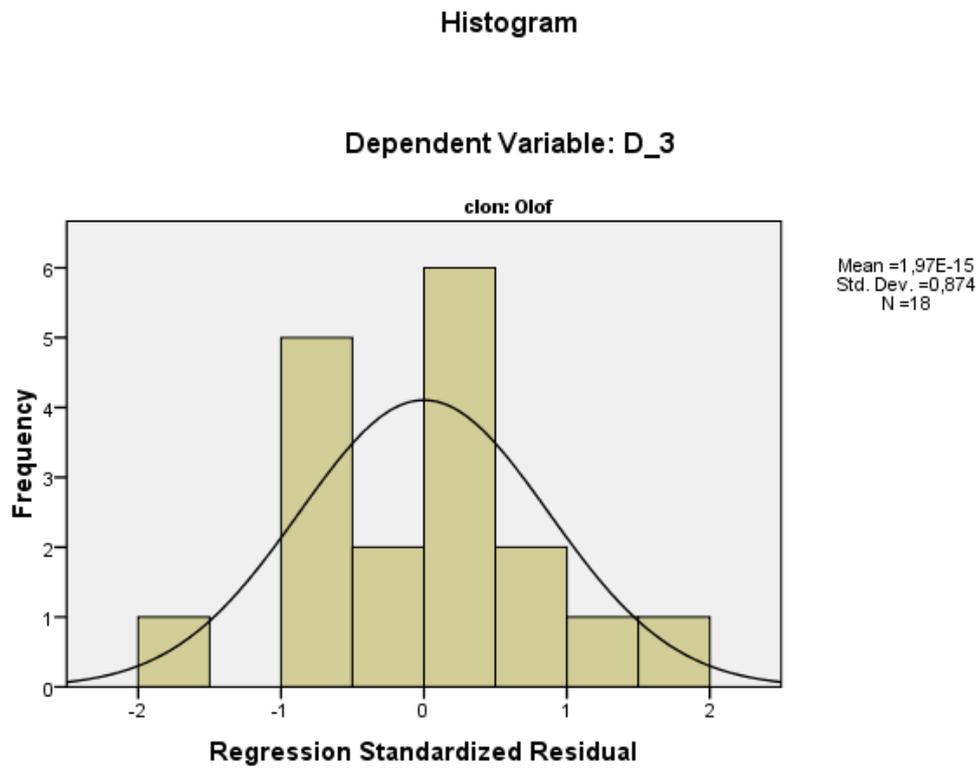


Gráfico 5: Histograma de residuos para la variable diámetro basal (cm)

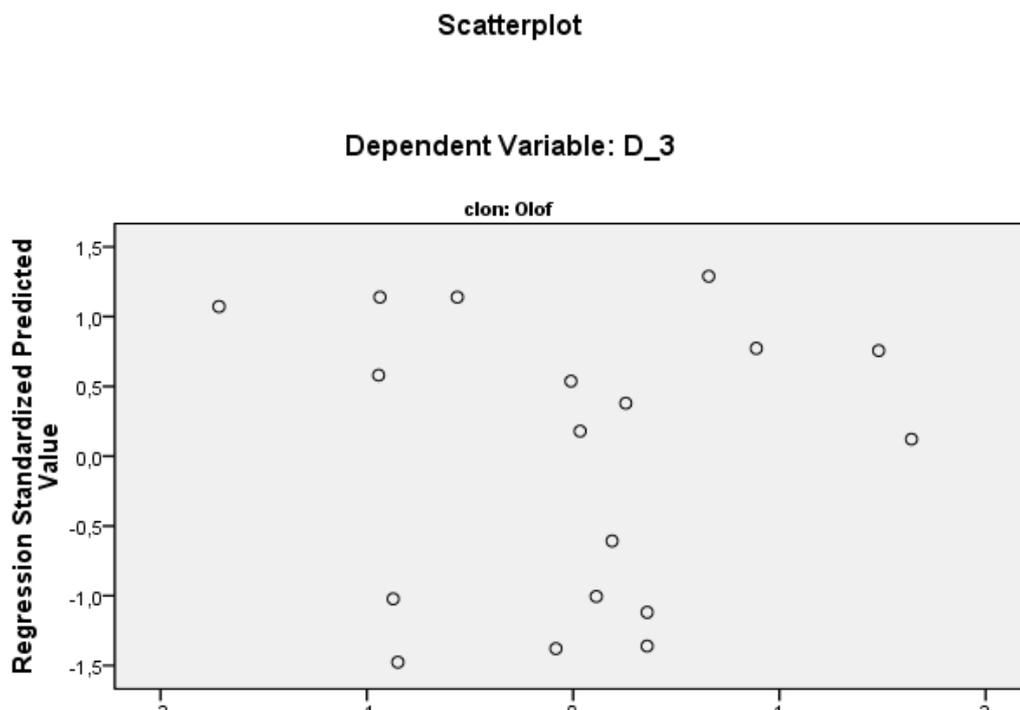


Gráfico 5: Gráfico de dispersión de residuos para el diámetro basal (cm)

Histogram

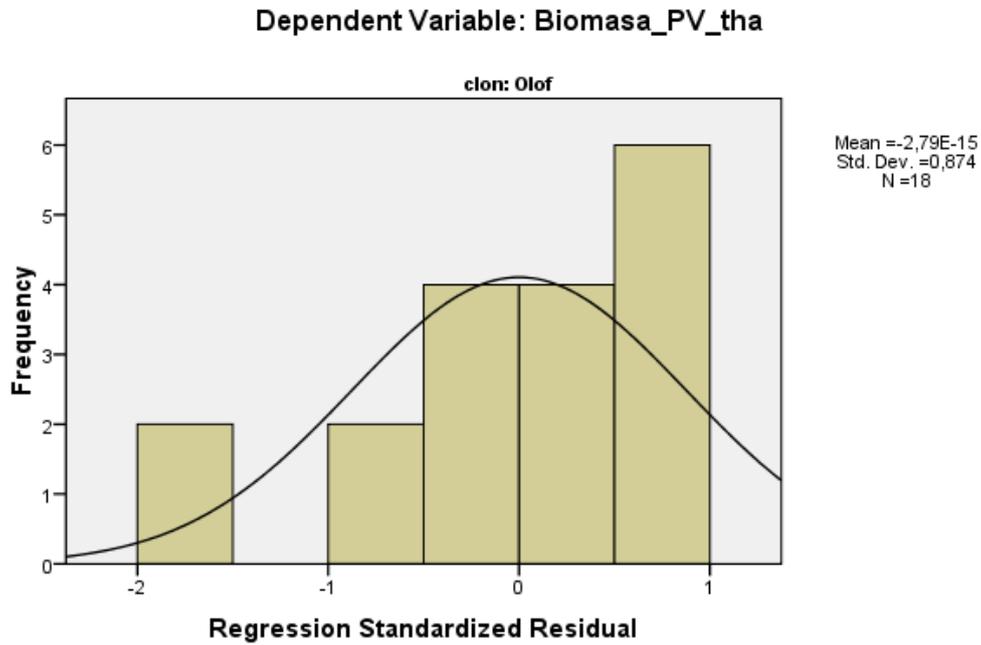


Gráfico 7: Histograma de residuos para la variable peso verde (t/ha)

Scatterplot

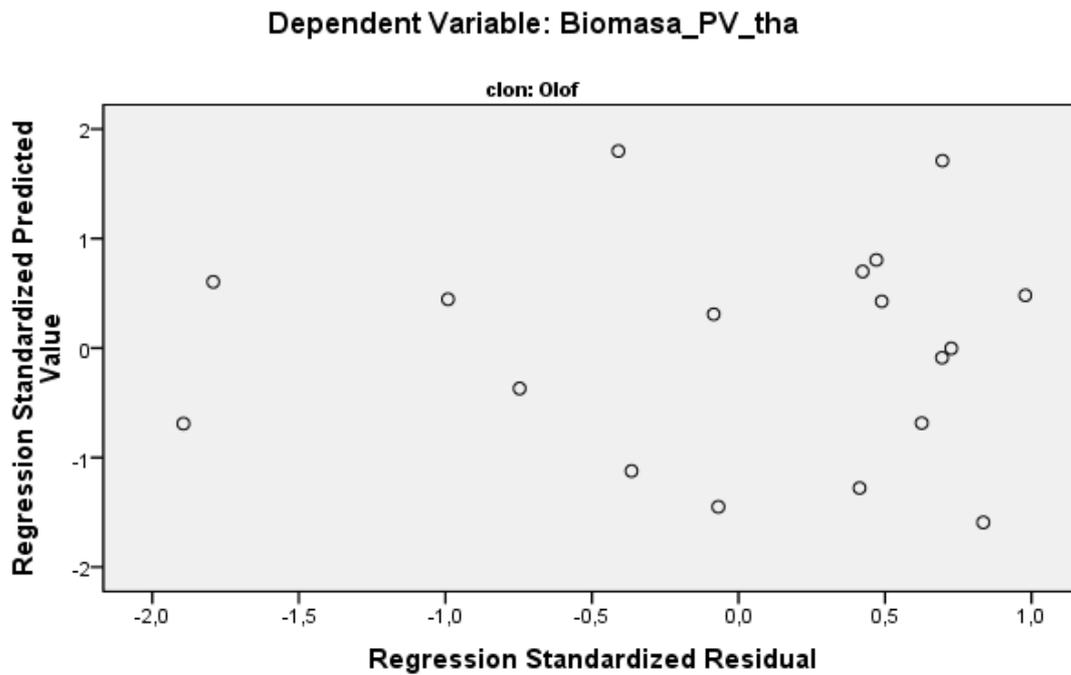


Gráfico 8: Gráfico de dispersión de residuos para la variable peso verde (t/ha)

Histogram

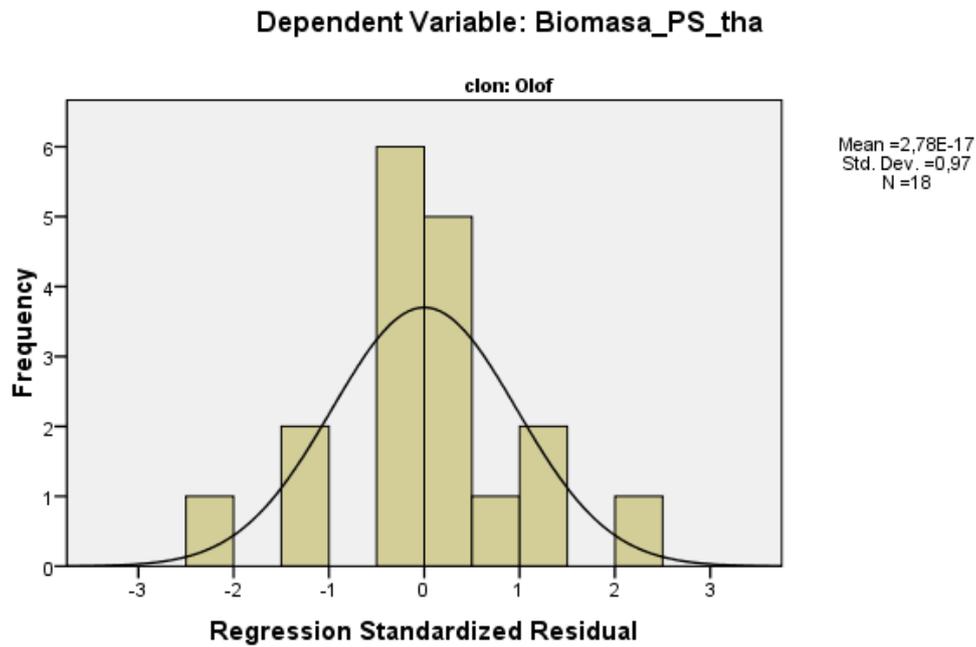


Gráfico 9: Histograma de residuos para la variable peso seco (t/ha)

Scatterplot

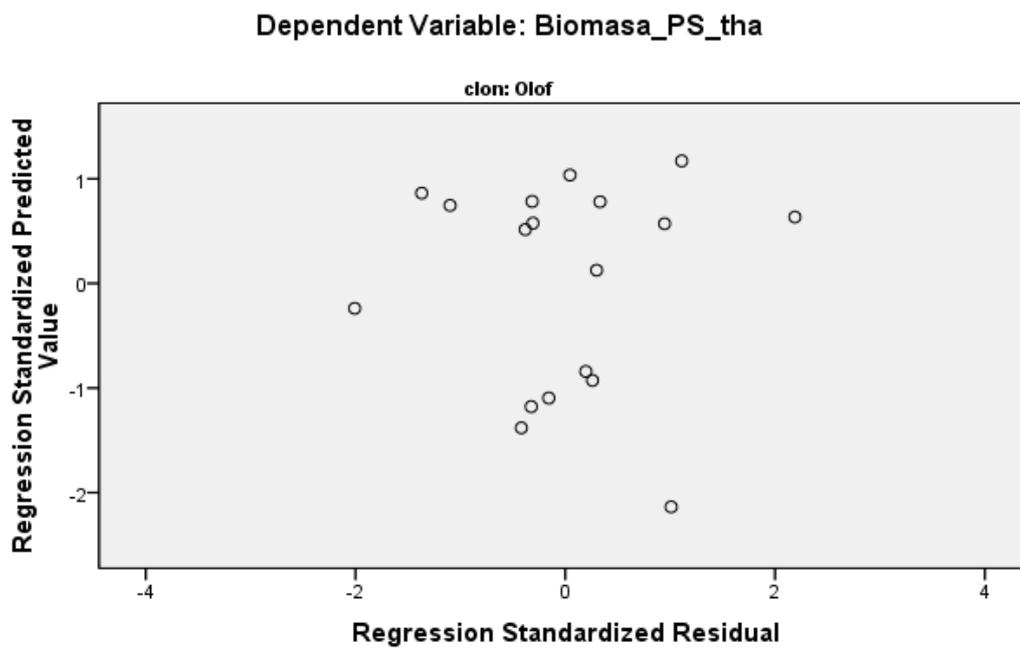


Gráfico 10: Gráfico de dispersión de residuos para la variable peso seco (t/ha)