

TRABAJO FIN DE MÁSTER

ANÁLISIS DE METODOLOGÍAS PARA EL CÁLCULO DEL ÍNDICE DE ÁREA FOLIAR (LAI) POST-INCENDIO EN ENTORNOS FORESTALES

Autor: Jorge García Hernández

Director: Fernando Pérez Cabello

**Máster Universitario en
Tecnologías de la información geográfica para la ordenación del territorio: sistemas de información geográfica y teledetección**

Octubre de 2023



**Universidad
Zaragoza**

**Departamento de Geografía
y Ordenación del Territorio**



Resumen

Las diversas fuentes de datos LiDAR están adquiriendo protagonismo en la caracterización de masas forestales, las cualidades de penetración y rebote del haz laser permiten la estimación precisa de variables que tradicionalmente se han considerado estructurales de las masas forestales. En el ámbito de los incendios forestales, el índice de área foliar (LAI) es una propiedad estructural del dosel que está adquiriendo importancia como indicador de recuperación de masas forestales afectadas por incendios. Propiedad física de difícil estimación, en la actualidad existen sensores de campo y productos satelitales que cuantifican el valor de LAI. El objeto de este estudio es valorar la capacidad de distintos sensores LiDAR para la estimación de LAI, analizando las distintas características que pudieran convertirse en aspectos clave, entre otras consideraciones como la resolución temporal de los datos o el estadio vegetativo de las masas forestales.

Palabras Clave: LAI, Incendios forestales, teledetección, LiDAR, satélite

Abstract

Diverse sources of LiDAR data are gaining prominence in forest characterization, the penetration and rebound qualities of the laser beam allow the precise estimation of variables that traditionally have been considered forest structural. In the field of forest fires, the leaf area index (LAI) is a structural property of the canopy that is gaining importance as a recovery indicator of fire affected forests. Although being a difficult to estimate physical property, currently there are field sensors and satellite products that quantify LAI value. The purpose of this study is to assess the capacity of different LiDAR sensors for estimating LAI, analysing the distinctive characteristics that could become key aspects, among other considerations, such as the temporal resolution of the data or the vegetative stage of forests.

Key Words: LAI, Forest fires, remote sensing, LiDAR, satellite

INDICE

1.	INTRODUCCIÓN	1
2.	OBJETIVOS	2
3.	AREA DE ESTUDIO	2
4.	MATERIALES	3
4.1.	Sensor LI-COR LAI 2200C.....	3
4.2.	DJI Matrice 300 RTK	4
4.3.	Software Agisoft Metashape.....	4
4.4.	Imágenes MSI Sentinel 2	5
4.5.	Software Snap	6
4.6.	Entornos de desarrollo Python y R.....	6
4.7.	Software FUSION.....	7
4.8.	Paquete de herramientas Lastools.....	7
4.9.	Sistemas de información geográfica	7
5.	METODOLOGIA	8
5.1.	Datos de campo	8
5.1.1.	Captura de datos de LAI	8
5.1.2.	Captura de datos LiDAR	9
5.2.	Definición de regiones de interés (ROIs).....	9
5.3.	Descarga de datos LiDAR PNOA de segunda cobertura.....	10
5.4.	Extracción de métricas LiDAR.....	10
5.4.1.	Clasificación de puntos de suelo	11
5.4.2.	Creación de modelos digitales de elevaciones	14
5.4.3.	Normalización de nubes de puntos	16
5.4.4.	Recorte de nubes de puntos	16
5.4.5.	Reducción de densidad de nubes de puntos	17
5.4.6.	Cálculo de métricas LiDAR.....	17
5.5.	Cálculo de LAI a partir de imágenes Sentinel 2	18
5.5.1.	Selección y descarga de imágenes	18
5.5.2.	Tratamiento de imágenes	19
5.5.3.	Extracción de valores de LAI	20
5.6.	Unión de tablas con variables dependientes y estadísticos a analizar	20
5.7.	Análisis estadístico	21
5.7.1.	Filtrado de datos	21
5.7.2.	Detección de valores atípicos y estadísticas descriptivas de variables dependientes	22
5.7.3.	Análisis de distribución normal y prueba de Shapiro-Wilk	23
5.7.4.	Análisis de correlaciones y colinealidad de variables independientes	23

5.7.5.	Análisis de regresión LASSO	25
5.7.6.	Prueba de normalidad y homocedasticidad de los residuos	25
6.	RESULTADOS	26
6.1.	Resumen estadístico de variables dependientes.....	26
6.2.	Filtrado de datos UAV Matrice.....	26
6.3.	Filtrado de datos UAV Matrice densidad reducida	27
6.4.	Filtrado de datos LiDAR PNOA segunda cobertura	28
6.5.	Detección de valores atípicos y análisis de estadísticas descriptivas	29
6.6.	Análisis de distribución normal y prueba de Shapiro-Wilk	30
6.7.	Descarte de variables independientes por correlación y colinealidad	34
6.7.1.	Descarte de variables independientes de datos LiDAR Matrice de alta densidad	34
6.7.2.	Descarte de variables independientes de datos LiDAR Matrice de baja densidad	34
6.7.3.	Descarte de variables independientes de datos LiDAR PNOA segunda cobertura	35
6.8.	Modelos de regresión LASSO y pruebas de normalidad y homocedasticidad de residuos	35
6.8.1.	Modelos LASSO de datos LiDAR Matrice de alta densidad.....	36
6.8.2.	Modelos LASSO de datos LiDAR Matrice de baja resolución espacial	36
6.8.3.	Modelos LASSO de datos LiDAR PNOA de segunda cobertura	36
7.	DISCUSIÓN	37
8.	AGRADECIMIENTOS	38
9.	BIBLIOGRAFIA	38
10.	ANEXO I: SCRIPT EN R UNION DE TABLAS EXCEL	40
11.	ANEXO II: SCRIPT EN R DE RECORTE DE NUBES DE PUNTOS CON PARCELAS	41
12.	ANEXO 3: SCRIPT EN PYTHON DE NORMALIZACION DE NUBES DE PUNTOS	42

1. INTRODUCCIÓN

Los incendios forestales han constituido históricamente en la España mediterránea, uno de los principales factores en la evolución de los ecosistemas forestales, modelando con su alta ocurrencia las comunidades vegetales presentes hacia una adaptación a dicho factor ambiental en mayor o menor medida.

Las consecuencias inmediatas de los incendios incluyen, además de la reducción de la cobertura vegetal, la perturbación de los suelos en cuanto a reducción de materia orgánica (Giovannini et al, 1998), desestabilización de agregados (Giovannini y Iuchesi, 1983), el incremento de sustancias hidrofóbicas (Giovannini y Iuchesi, 1984) y la generación de costras superficiales (Cerdeira et al., 1995)

En los ecosistemas mediterráneos, debido a la adaptabilidad antes señalada, tras un incendio se desencadenan intensos mecanismos de rebrote/germinación de especies y semillas preexistentes que conllevan un rápido recubrimiento del suelo (Pérez-Cabello et al., 2011). Diversos estudios en ecosistemas mediterráneos muestran que los efectos erosivos post-incendio se reducen notablemente durante los primeros años tras el incendio, en relación con el incremento de la cobertura vegetal (Marques & Mora, 1992)(Cerdeira, 1998).

Sin embargo, en un escenario post-fuego, diversos factores pueden dilatar la respuesta del ecosistema, por ejemplo, la severidad del fuego puede condicionar la regeneración vegetal y la dinámica hidrogeomorfológica, por lo que una correcta evaluación de cada escenario permite focalizar áreas de actuación y la adopción de estrategias y medidas de recuperación específicas (Montorio Llovería et al., n.d.).

La elaboración de cartografías a partir de sensores remotos puede apoyar sustancialmente la evaluación de ecosistemas afectados por incendios, para lo cual los índices de vegetación son ampliamente utilizados (Cabello et al., 1998).

El índice de área foliar (LAI), es una propiedad estructural del dosel que se define como la superficie foliar fotosintéticamente activa por unidad de superficie (Chen and Black, 1991), según Jiménez Ruano et al., (2016), se ha comprobado que los niveles de LAI se reducen en torno a un 75% tras un incendio como consecuencia de la combustión de la vegetación, y que existe una relación directa entre el tiempo transcurrido tras un incendio y el incremento de los valores de LAI, según se desarrollan estructuras de vegetación más complejas, lo que indica un importante potencial de LAI para evaluar la evolución de comunidades vegetales post-incendio.

La obtención de LAI, que tradicionalmente se ha realizado de forma empírica, está ejecutándose a partir de imágenes satelitales multiespectrales, inicialmente de baja resolución (MODIS) y actualmente a partir de imágenes de alta resolución (Landsat 8, Sentinel 2), además, se han desarrollado sensores de campo para la estimación de LAI como el sensor LI-COR LAI 2200, y las nuevas tecnologías drone, junto con sensores LiDAR más accesibles han generado una extensa bibliografía respecto de la estimación de LAI a partir de nubes de puntos LiDAR (Wang y Fang, 2020)

Estas tecnologías han alcanzado distintos grados de efectividad en el cálculo de LAI en masas forestales, sin embargo, el seguimiento de la evolución post-incendio requiere de su aplicación sobre comunidades vegetales poco desarrolladas, lo que implica, previsiblemente, una alta sensibilidad de los sensores y métodos empleados que nos permita observar y valorar la evolución temporal de estas.

Por otra parte, la profusión y democratización de la tecnología LiDAR en España, con la existencia de un programa de captura estatal (PNOA) que genera coberturas LiDAR de libre acceso para toda la superficie del estado de manera periódica, los cuales son ampliamente utilizados en labores de investigación, gestión e inventario forestal. Supone una oportunidad como herramienta para realizar seguimientos de LAI a partir de datos de mayor resolución respecto a imágenes satelitales, pese a su baja resolución espacial en comparación con vuelos LiDAR dedicados en plataformas tripuladas o vehículos aéreos no tripulados (UAV).

En el presente estudio, se pretende contrastar la viabilidad de diversos sensores para la estimación de valores de LAI sobre ecosistemas forestales mediterráneos que han sufrido incendios forestales a lo

largo de las últimas décadas, presentando diversos grados de recuperación post-incendio, y establecidas parcelas de control sobre superficies no quemadas del entorno en todos los casos.

En el caso de los sensores LiDAR, una vez establecida la viabilidad de los datos LiDAR de alta densidad de puntos para realizar estimaciones de LAI, obtenidos a partir de vuelos con un vehículo aéreo no tripulado sobre las parcelas de campo, se procederá a la reducción aleatoria de dicha densidad de puntos hasta alcanzar valores de puntos por unidad de superficie semejantes a los proporcionados por los vuelos LiDAR del programa PNOA sobre las parcelas de campo, con el objetivo de valorar la capacidad de nubes de puntos de dicha densidad en cuanto a la estimación de valores de LAI.

2. OBJETIVOS

En el presente estudio, se pretende: (1) contrastar la viabilidad de diversos sensores, Sentinel 2 y LiDAR sobre plataforma tripulada o vehículo aéreo no tripulado, para la estimación de valores de LAI sobre ecosistemas forestales mediterráneos/submediterráneos que han sufrido incendios forestales a lo largo de las últimas décadas, y que en la actualidad presentan diversos grados de recuperación post-incendio.

En el caso de los sensores LiDAR, (2) realizar una valoración de cuáles son los niveles de densidad de las nubes de puntos óptimos para la estimación de LAI. Una vez establecida la viabilidad de los datos LiDAR de alta densidad de puntos para realizar estimaciones de LAI, obtenidos a partir de vuelos con un vehículo aéreo no tripulado sobre parcelas de campo, se procederá a la reducción aleatoria de dicha densidad de puntos hasta alcanzar valores de puntos por unidad de superficie semejantes a los proporcionados por los vuelos LiDAR del programa PNOA sobre parcelas de campo.

(3) analizar el grado de asociación entre los datos LiDAR de la 2ª cobertura del programa PNOA, con fecha de captura en 2020 para la zona de estudio, análisis que nos permitirá valorar su viabilidad para la estimación de valores de LAI pese al desfase temporal de varios años respecto de los datos de campo y las imágenes Sentinel 2.

Finalmente, dado que la captura de valores de LAI con los distintos sensores y la de métricas LiDAR con el vehículo aéreo no tripulado se han realizado de forma simultánea, pero con un desfase de hasta 3 meses entre parcelas que las sitúa en distinto estadio vegetativo en el momento de captura, (4) valorar la influencia del estadio vegetativo diferenciado en la variabilidad de LAI y las métricas LiDAR

También se intentarán establecer correlaciones entre los datos LiDAR de segunda cobertura del programa PNOA, con fecha de captura en 2020 para la zona de estudio, lo que nos permitirá valorar su viabilidad para la estimación de valores de LAI pese al desfase temporal de varios años respecto de los datos de campo y las imágenes Sentinel 2.

3. AREA DE ESTUDIO

El objeto de estudio son superficies forestales quemadas en la comunidad autónoma de Aragón desde el año 1980, seleccionándose 76 transectos de estudio que incluyen coberturas vegetales que fueron afectadas por el fuego y coberturas de control cercanas que no fueron afectadas por los incendios.

Estos transectos están distribuidos a lo largo de la geografía aragonesa en distintos municipios de las provincias de Huesca y Zaragoza, según el siguiente listado:

- Ainzon: 2 transectos en 1 ubicación
- Ambel: 4 incendios en 1 ubicación
- Caldearenas: 13 transectos en 2 ubicaciones
- Loarre: 8 transectos en 4 ubicaciones
- Longas: 2 transectos en 1 ubicación
- Nueno: 12 transectos en 3 ubicaciones

- Orés: 4 transectos en 2 ubicaciones
- Peñas de Riglos: 9 transectos en 3 ubicaciones
- Sabiñánigo: 5 transectos en 2 ubicaciones
- Tabuenca: 7 transectos en 3 ubicaciones
- Talamantes: 2 incendios en 1 ubicación
- Zuera: 8 transectos en 3 ubicaciones

Las especies principales de vegetación presentes en los distintos enclaves incluyen mayoritariamente distintas especies del género *Pinus*, tales como *Pinus sylvestris*, *Pinus nigra* y *Pinus halepensis*, pero también masas de quejigo y encina (*Quercus faginea* y *Quercus ilex* respectivamente), incluyendo masas de repoblación y naturales, pudiendo localizar, como formas fundamentales de masa, monte alto, monte medio y monte bajo.

Entre las especies principales de matorral cabe destacar coscojas (*Quercus coccifera*), aulagas (*Genista scorpius*), enebro rojo y común (*Juniperus oxycedrus* y *Juniperus communis*), o distintas especies de rosales silvestres y zarzas (*Rosa sp.* y *Rubus sp.*)

El rango altitudinal se sitúa entre los 537 metros sobre el nivel del mar de los transectos en las ubicaciones más bajas de los Montes de Zuera a los 1579 metros de la Sierra del Águila en el entorno del Pico del Águila, con un valor medio de altitudes de 921 metros y una mediana ubicada en los 890 metros de altura.

En cuanto a las pendientes, encontramos un rango que se sitúa entre una pendiente media mínima del 3 por ciento y una pendiente media máxima del 65.6 por ciento, alcanzando un valor medio de un 35.05 por ciento y una mediana del 36 por ciento. Si bien la orientación media de los transectos es sureste, y la mediana es orientación sur, podemos encontrar diversidad de valores en este aspecto, abarcando todo el abanico de orientaciones posible.

4. MATERIALES

4.1. Sensor LI-COR LAI 2200C

El LAI-2200C calcula el Índice de Área Foliar (*Leaf Area Index*, LAI), además de otros parámetros relacionados con la estructura foliar del dosel. El principio básico del sensor es que cuando un haz de luz atraviesa un dosel vegetal, hay una gran probabilidad de que sea interceptado por el follaje en función de la longitud de la trayectoria en el dosel, la densidad foliar (área de follaje sobre volumen del dosel) y la propia orientación del follaje. Por tanto, el instrumento se basa en un sensor óptico de “ojo de pez”, para realizar mediciones encima y debajo del dosel y poder calcular la interceptación de luz (<490 nm) según una semiesfera conformada por 5 anillos, y con un campo de visión de 360 grados (puede reducirse mediante mascarar para evitar la influencia del operador u otros elementos), utilizando un modelo de transferencia radiativa.

Es un sensor de campo con el que se pueden realizar muestreos puntuales o recorridos a lo largo de una parcela (como en nuestro caso) con capturas continuas que permitirán luego agrupar valores de LAI según el interés de la investigación.



Table 4-2. Distance Factors.

		Distance Factors				
		Sensor Field of View:				
Ring	Angle (°)	360°	270°	180°	90°	45°
1	7	10	20	30	50	100
2	22	4	5	8	20	30
3	38	3	3	5	10	20
4	52	2	3	4	8	15
5	68	2	2	3	7	14

Figura 1. Sensor LAI 2200c y tabla de ángulos cenitales y campos de visión (Fuente: LI-COR Biosciences)

4.2. DJI Matrice 300 RTK

Este UAV es un multicoptero de cuatro hélices con un peso máximo al despegue de 9 kg, y una tara de 3.6 o 6.3 kilogramos en función de la configuración de baterías, lo que le permite portar una carga de pago de hasta 2.7 kg, según los datos del fabricante. La autonomía máxima está declarada en 55 minutos, siempre con la configuración de baterías y carga de pago más favorable.

El equipo consta de un sensor de posicionamiento satelital capaz de recibir datos de las constelaciones GPS, GLONASS, BeiDou y Galileo. Además, es capaz de comunicarse con redes de antenas GNSS que ofrecen posicionamiento mejorado, siendo capaz de trabajar en modo RTK (Real Time Kinematic), trabajar en modo base + rover con una antena GNSS compatible fija sobre el terreno, así como de generar datos Rinex para la mejora de la precisión en postproceso utilizando los datos aportados por redes GNSS estáticas.

El sistema de posicionamiento satelital en combinación con un sistema de medición inercial de alta precisión permite el posicionamiento y orientación preciso de los datos captados por un sensor equipado y compatible.

El equipo Zenmuse L1 de DJI, instalado en el UAV durante los vuelos, consta de un sensor LiDAR marca Livox y modelo AVIA de triple eco, una longitud de onda de 905nm, un alcance de 450 metros en condiciones ideales, un peso de 498 gramos y una precisión de medida de 2 cm, acompañado por una cámara óptica y un sistema de medición inercial mejorado de DJI, que alcanza la precisión necesaria para el posicionamiento con garantías de la nube de retornos LiDAR.



Figura 2. Sensor LIVOX AVIA (izq.) y UAV Matrice 300 rtk con Zenmuse L1 (der.) (Fuentes: Livox y DJI)

4.3. Software Agisoft Metashape

El software Metashape de Agisoft está dedicado al procesamiento de imágenes de vuelos fotogramétricos mediante el método de pares estereoscópicos, con la posibilidad de geo-posicionar las imágenes, respecto de un sistema de referencia dado, en base a puntos de control del terreno visibles en las imágenes con coordenadas conocidas y/o la importación de imágenes de vuelo con coordenadas de captura y ángulos de orientación de las imágenes de mayor o menor precisión.

Esta capacidad le permite generar nubes de puntos 3d geo-posicionados con una alta densidad, permitiendo mediante procesos de clasificación e interpolación la obtención de modelos digitales de elevaciones y la orto-rectificación de las imágenes de vuelo para generar mosaicos de ortofotografías de las superficies voladas.

Es destacable la capacidad del software para realizar clasificaciones de nubes de millones de puntos de suelo en base a criterios morfológicos, concretamente, la relación de alturas y ángulos entre puntos contiguos dentro de una ventana móvil de tamaño configurable.

Las últimas versiones del software permiten la importación de nubes de puntos en distintos formatos, lo que hace posible la importación de nubes de puntos LiDAR para su clasificación, razón por la que ha sido utilizado para la importación de las nubes de puntos generadas con el sensor LiDAR con el objetivo de clasificar puntos del terreno y generar un modelo digital de elevaciones que permitiera normalizar las alturas de las nubes de puntos como paso previo a la obtención de métricas sobre dichas nubes.

4.4. Imágenes MSI Sentinel 2

Los Sentinel son una flota de satélites diseñada específicamente para proporcionar los datos requeridos por el programa Copernicus de la Comisión Europea, desarrollados en colaboración con la Agencia Espacial Europea (ESA) con el objetivo de establecer una capacidad europea para la observación de la Tierra.

Concretamente, la misión Sentinel 2 consta de dos satélites gemelos, Sentinel 2a y Sentinel 2b, operando a una altitud media de 786 kilómetros en órbita helio-sincrónica, que asegura un ángulo de luz solar consistente sobre la superficie de la tierra observada, ambos satélites ocupan una misma órbita, separados por 180 grados, con un ciclo de 10 días, lo que supone un periodo de revisita de 5 días en el ecuador, considerando ambos satélites. Las imágenes se capturan entorno a las 10:30 am, hora local.

Ambos satélites están equipados con el sensor Multiespectral Instrument (MSI), un radiómetro de empuje (push-broom) que recolecta los datos en el sentido de avance de la órbita del satélite. El sensor captura la información en trece bandas espectrales, distribuidas con el objetivo de estudiar principalmente sistemas agrícolas, forestales y marinos, desde el infrarrojo visible (VNIR) al infrarrojo de onda corta (SWIR), con resoluciones espaciales desde los 10 metros hasta los 60 metros.

Band Number	S2A		S2B		Spatial resolution (m)
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	
1	442.7	20	442.3	20	60
2	492.7	65	492.3	65	10
3	559.8	35	558.9	35	10
4	664.6	30	664.9	31	10
5	704.1	14	703.8	15	20
6	740.5	14	739.1	13	20
7	782.8	19	779.7	19	20
8	832.8	105	832.9	104	10
8a	864.7	21	864.0	21	20
9	945.1	19	943.2	20	60
10	1373.5	29	1376.9	29	60
11	1613.7	90	1610.4	94	20
12	2202.4	174	2185.7	184	20

Tabla 1. Bandas espectrales de los satélites Sentinel 2a y 2b (Fuente: ESA)

4.5. Software Snap

Snap es un software desarrollado por la Agencia Espacial Europea que incluye las distintas cajas de herramientas para el trabajo con imágenes satelitales de las misiones Sentinel, aunque dichas herramientas pueden ser utilizadas para realizar análisis de imágenes multispectrales o Radar de otras fuentes.

Entre dichas herramientas se incluyen, entre otras, operaciones geométricas como mosaicados y remuestreo, conversiones de formato, generación de máscaras, extracción por superficie, extracción de bandas, análisis de componentes principales, distintos métodos de clasificación, visualización de datos espectrales, distintas herramientas de preprocesado de imágenes, corrección atmosférica y procesamiento de distintos productos temáticos para vegetación o entornos marinos, incluyendo el cálculo distintos índices de vegetación y biofísicos como pudiera ser el índice de área foliar o la fracción de cubierta, estos últimos dedicados concretamente a imágenes Sentinel 2 o Landsat 8.

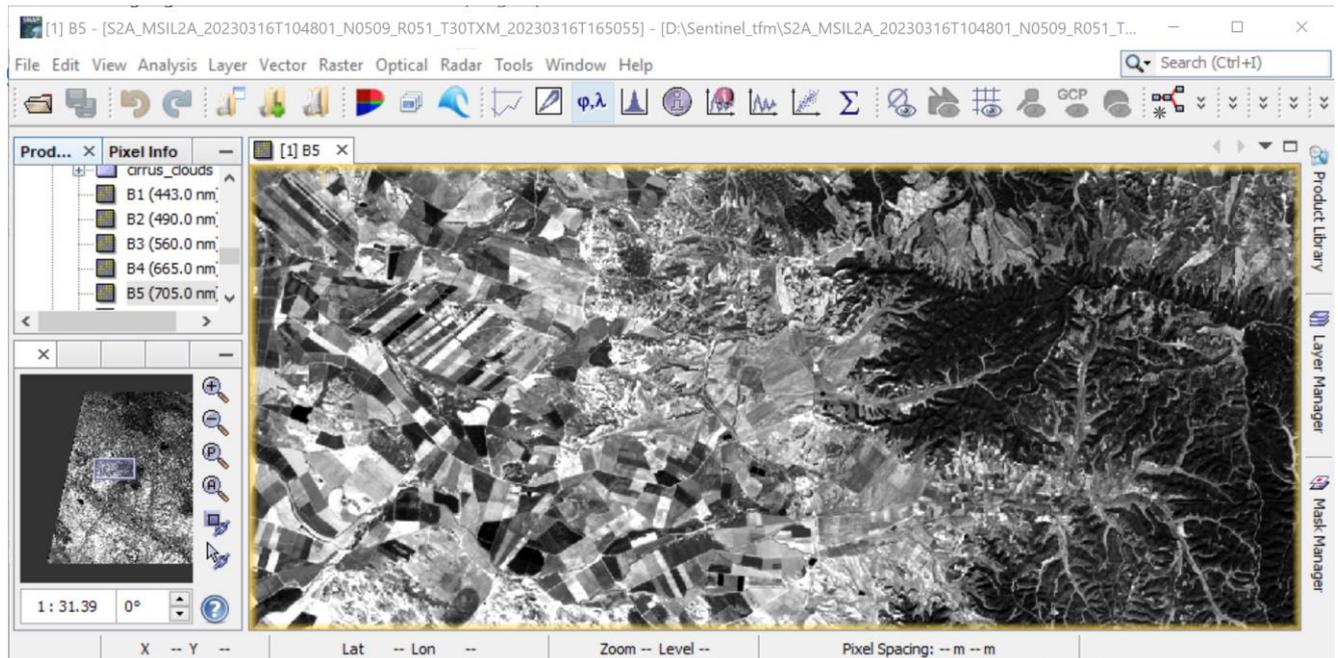


Figura 3. Vista de imagen Sentinel 2 en software Snap

4.6. Entornos de desarrollo Python y R

La alta resolución de los datos LiDAR, junto con una extensión de los trabajos que abarca buena parte de la comunidad autónoma de Aragón, ha requerido, con el objetivo de reducir la potencia de procesamiento necesaria y aumentar el rendimiento en tiempo, de la programación de procesos geoespaciales en los lenguajes de programación R y Python que permitieran compartimentar y lanzar en bucle distintos análisis geoespaciales necesarios en la edición de las nubes de puntos LiDAR y los modelos digitales de elevaciones, principalmente, así como la edición y unión de tablas de datos.

Además, el lenguaje de programación R ha sido la herramienta sobre la que se ha diseñado todo el análisis estadístico realizado.

Entre los paquetes utilizados se encuentran, en R: lidR, Raster, rgdal, sf, tidyverse, readxl, dplyr y openxlsx, ggplot2, e1071, glmnet, MASS, caret, corrplot y car. Las librerías Python utilizadas son: os, pylas, rasterio y numpy.

En cuanto a entornos de desarrollo, en ambos casos se ha seleccionado software de licencia libre, los scripts de R han sido desarrollados y ejecutados en el software RStudio, mientras que la programación y ejecución en Python se han ejecutado en el entorno de desarrollo Eclipse IDE.



Figura 4. Logos de los software libres RStudio y Eclipse IDE

4.7. Software FUSION

FUSION es una compilación de software desarrollado para visualizar y analizar nubes de puntos LiDAR generadas mediante vuelos tripulados. Es un software libre propiedad del servicio forestal de los Estados Unidos, concretamente del Departamento de Agricultura, con el objetivo de contribuir a las labores de inventario propias de la gestión forestal.

El software permite visualizar nubes de puntos a través de una representación 2D, sobre la que se pueden seleccionar pequeños conjuntos de datos LiDAR que puedan ser visualizadas en 3D o editados manualmente, sin embargo, su potencia reside en la capacidad de realizar análisis geoespaciales sobre la nube de puntos durante las que pueden ser empleadas y generadas capas vectoriales o raster, lo que permite operaciones de recorte, extracción, clasificación, normalización de elevaciones, y cálculo de estadísticos por parcelas que pueden ser exportados en formato tabla para su análisis.

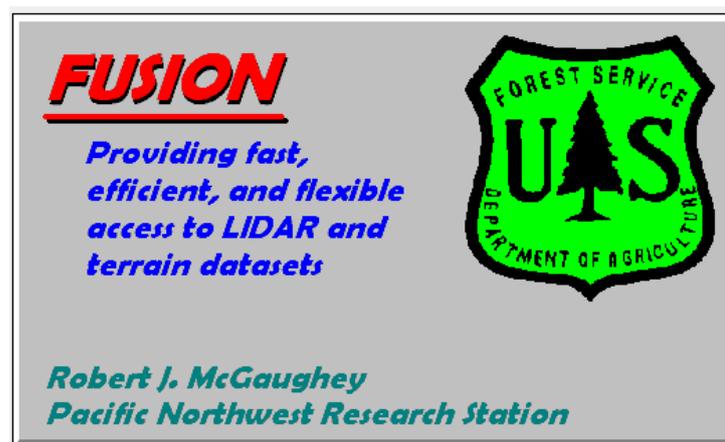


Figura 5. Logotipo de FUSION

4.8. Paquete de herramientas Lastools

Lastools es una colección de herramientas que permiten distintos procesamientos sobre nubes de puntos en formato las o en formato comprimido laz, dichas herramientas son programables en formato código o ejecutables sobre algunos sistemas de información geográfica, como ArcMap, donde se pueden ejecutar sobre varios archivos en un mismo lote(batch).

Entre la colección de herramientas, se ha utilizado Lasthin, herramienta que permite reducir la densidad de una nube de puntos, para reducir la densidad de las nubes de puntos obtenidas con el UAV con el sensor Zenmuse L1 con el objetivo de asemejar sus características a las de las nubes de puntos proporcionadas por el programa LiDAR PNOA.

4.9. Sistemas de información geográfica

Los sistemas de información geográfica (SIG) han sido utilizadas para la visualización de los distintos productos generados, y en algunos casos, para ejecutar procesos de conversión de formatos capas

vectoriales o raster, o más concretamente para extraer valores de capas raster a capas vectoriales, así como en la generación de representaciones cartográficas necesarias para ilustrar apartados de la presente memoria.

Los softwares utilizados han sido ArcMap de ESRI y el software libre QSIG.

5. METODOLOGIA

5.1. Datos de campo

Los trabajos de campo necesarios durante el presente estudio comprenden la toma de valores de índice de área foliar (LAI) con el sensor LI-COR LAI 2200c, y la toma de datos LiDAR de alta densidad, mediante vuelos efectuados con el UAV DJI Matrice 300 equipado con el sensor Zenmuse L1.

Dichos trabajos de campo fueron realizados por personal investigador del grupo Geoforest-IUCA de la Universidad de Zaragoza en el contexto del proyecto de investigación "PAF: Análisis dinámico de la resiliencia de los paisajes forestales afectados por el fuego mediante indicadores espectrales multi-sensor (PID2020-118886RB-I00)" del Ministerio de Ciencia e Innovación del Gobierno de España.

Durante los meses de marzo, mayo y junio de 2023, se tomaron valores de LAI y datos LiDAR de alta densidad en 76 transectos distribuidos sobre superficies afectadas por incendios forestales a lo largo de la geografía aragonesa, 35 transectos fueron visitados durante el mes de marzo, 12 durante el mes de mayo y 29 transectos fueron visitados durante el mes de junio, ejecutándose mediciones de LAI y toma de datos LiDAR durante el mismo día de trabajo o con la menor amplitud de fechas posible cuando los trabajos de vuelo y toma de datos LAI no pudieran ser ejecutados en el mismo día.

En cuanto a la distribución de los transectos, se han alternado en una misma zona transectos sobre superficies afectadas por el fuego, y transectos de control situados sobre superficies forestales próximas no afectadas por el incendio. Ejecutándose 41 parcelas sobre superficie quemada y 35 parcelas de control, cuya denominación incluye las letras "Q" o "C" respectivamente, a efectos de identificación.

FECHA	Transectos
03/03/2023	1C1, 1C2, 1Q1, 1Q2
10/03/2023	2C1, 2C2, 2C3, 2Q1, 2Q2, 2Q3, 3Q1, 3Q2, 3Q3
15/03/2023	4C1, 4C2, 4C3, 4Q1, 4Q2
16/03/2023	6C1, 6C2, 6C3, 6Q2, 6Q3, 7Q1
22/03/2023	7C2, 7Q2
29/03/2023	9C1, 9C2, 9Q1, 9Q2, 10C1, 10Q1, 10Q2, 11Q1, 11C1
17/05/2023	12C1, 12Q1, 12Q2, 13Q1, 13Q2, 13Q3, 12C2, 13C1
25/05/2023	14C1, 14C2, 14Q1, 14Q2
09/06/2023	15Q1, 15Q2, 15C1, 15C2, 16C1, 16Q2
14/06/2023	17Q1, 17C1, 17Q2, 18C1, 18C2, 18C3, 18Q2, 18Q3, 18Q4
16/06/2023	19C1, 19Q1, 20C2, 20Q2, 20Q3, 20C3
22/06/2023	21C1, 21C2, 21Q1, 21Q2, 22C1, 22C2, 22Q1, 22Q2

Tabla 2. Fechas de visita de las distintas parcelas inventariadas

5.1.1. Captura de datos de LAI

La toma de datos de LAI con el sensor LI-COR LAI 2200c se ha ejecutado sobre transectos rectilíneos de 10 metros, rectilíneos en la medida en que la vegetación de la parcela lo permitiera, efectuando mediciones parciales cada medio metro recorrido, y realizando el transecto primero en sentido hacia el sol y posteriormente dejando el sol a la espalda del operador, utilizando en esta segunda pasada la máscara adecuada para evitar la influencia del operador sobre la luz captada por el sensor. La ejecución de

las dos pasadas sobre el transecto de 10 metros con mediciones parciales realizadas cada medio metro, aproximadamente, nos permite obtener una media de 20 mediciones parciales por transecto.

Se han escogido únicamente las mediciones efectuadas en el sentido de la luz solar debido a cierta inconsistencia observada en aquellas mediciones en las que el sol quedaba a la espalda del operador, eligiendo entre estas el valor de la mediana como estadístico representativo de los transectos, con el objetivo de minimizar el impacto de valores extremos obtenidos en las mediciones parciales.

La medición del último anillo del sensor, definida por una banda angular con centro en los 62 grados respecto a la normal del sensor, ha sido descartada por su reducido peso específico en la medida del sensor respecto a las bandas de menor ángulo cenital.

En cuanto a la geolocalización de los transectos, los puntos de inicio y final de cada transecto han sido observados con la antena GNSS de precisión centimétrica LEICA GS15 en modo RTK, lo que permite disponer la localización de los transectos con una precisión suficiente para el contraste de los datos LAI y LiDAR. El sistema de referencia utilizado para el almacenamiento de los puntos es ETRS89 UTM Huso 30 Norte, con valores de alturas tomadas respecto del geoide de referencia en España, EGM08.

5.1.2. Captura de datos LiDAR

Los vuelos LiDAR se han efectuado sobre superficies de en torno a 15 hectáreas, abarcando cuando ha sido posible varios transectos de LAI. Para su ejecución se han planificado pasadas contiguas a 100 metros de altura respecto a un MDE cargado de un metro de resolución espacial, con un solape transversal del 80% y una velocidad media de vuelo de 7 m/s, obteniendo una densidad de puntos de 450 puntos por metro cuadrado, en función de las características del área.

El método de geoposicionamiento utilizado durante la captura de los datos ha sido el posicionamiento cinemático en tiempo real, RTK, utilizando las redes de datos móviles para recibir correcciones de la antena GNSS fija de la red ARAGEA más cercana en cada caso.

Posteriormente a la captura de los datos, personal especializado de la Universidad de Zaragoza ha realizado el ajuste necesario entre las distintas pasadas del vuelo mediante el software dedicado DJI a tal efecto. Durante el cual se corrigen las diferencias producidas por el sensor LiDAR en sus extremos transversales a la pasada, que generan algunas diferencias de cota entre puntos coincidentes en el espacio de pasadas contiguas.

5.2. Definición de regiones de interés (ROIs)

Las regiones de interés (ROIs) respecto de las que se extraerán datos LiDAR y valores de LAI obtenidos a partir de las imágenes satelitales, queda definidos por la superficie respecto a los transectos de LAI cuya vegetación ha tenido la capacidad de influir en la medición del sensor.

Dado que las bandas de distinto ángulo cenital del sensor LI-COR LAI 2200c utilizadas captan la incidencia de luz solar hasta un máximo de unos 52 grados respecto de la dirección normal al sensor, y considerando que la altura máxima del arbolado en el conjunto de transectos estudiados se encuentra por debajo de los 20 metros, el estudio de datos LiDAR y la extracción de valores de LAI de imágenes satelitales se ha definido para superficies comprendidas por un buffer de 20 metros respecto a la línea definida por el punto inicial y el punto final de cada transecto de medición de LAI observado con el sensor LI-COR LAI 2200c, que correspondería a la máxima superficie con elementos que afectan al sensor, en el caso de presentarse vegetación de la máxima altura en el entorno del transecto.

Tomando como base la capa de puntos observados con antena GNSS sobre las localizaciones de los puntos de inicio y final de cada transecto capturado con el sensor LAI 2200c, utilizando el software ArcMap se han digitalizado las líneas que unen el punto inicial y final de cada transecto, ejecutando posteriormente un buffer de 20 metros respecto de las líneas obtenidas, quedando así definida una capa de polígonos que contiene las ROIs de cada uno de los transectos.

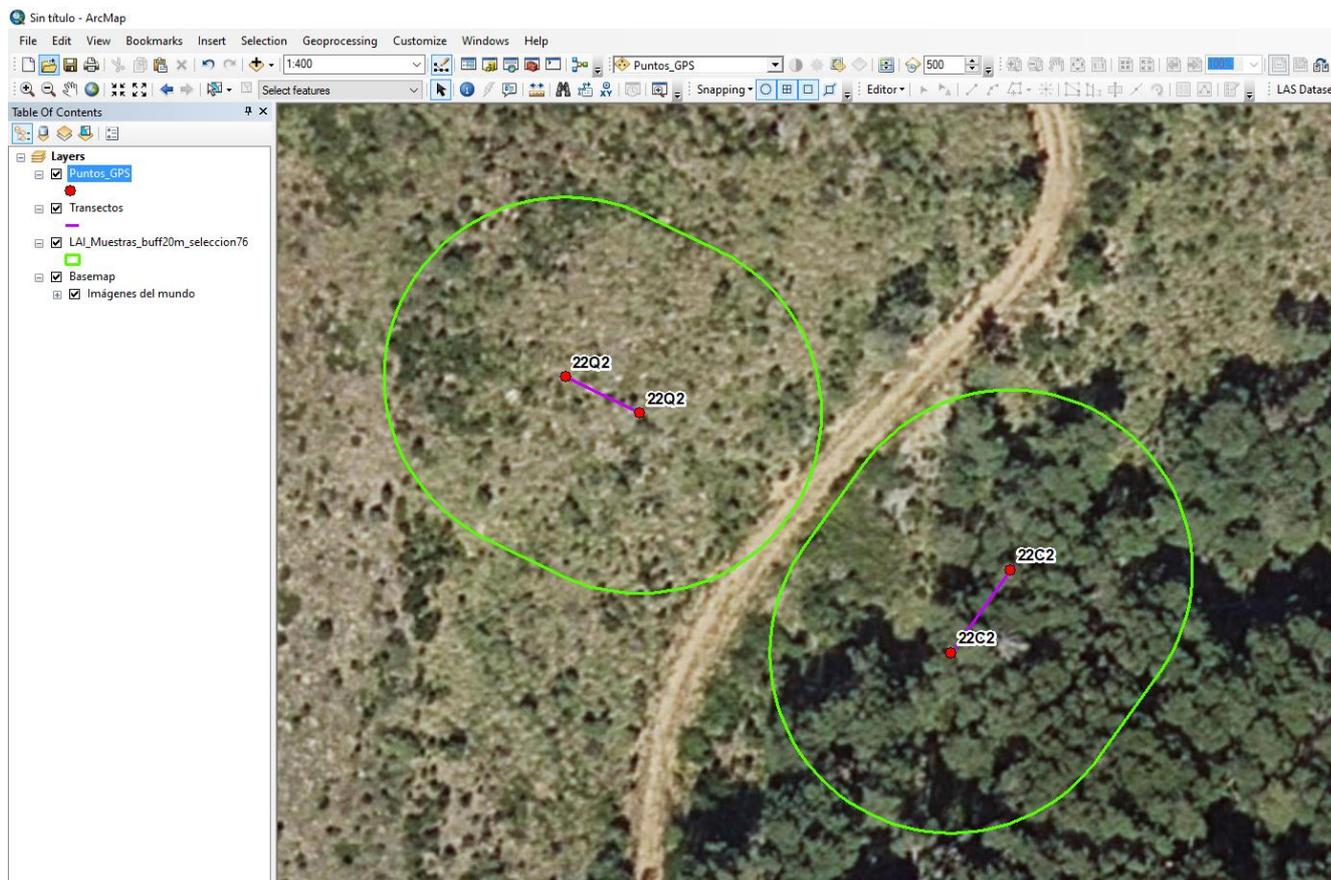


Figura 6. Definición de ROIs en ArcMap

5.3. Descarga de datos LiDAR PNOA de segunda cobertura

Con el objetivo de analizar la correlación de los valores de LAI con métricas obtenidas a partir de datos LiDAR PNOA de la segunda cobertura de vuelos estatales, se han descargado los datos a partir del centro de descargas del Centro Nacional de Información Geográfica (CNIG), que serán procesados de manera análoga al procesamiento realizado para la obtención de métricas LiDAR con los datos de los vuelos con el UAV.

Para la descarga de los datos mediante el centro de descargas de CNIG, se ha generado una capa vectorial de polígonos a partir de un buffer de 200 metros respecto de los transectos tomados con el sensor LI-COR LAI 2200c. Esta capa de polígonos, para su aceptación por el buscador del centro de descargas, debe proyectarse en sistema de referencia WGS84, y comprimirse, con todos sus archivos asociados en un archivo formato “Zip”, tras lo cual puede cargarse en el visor del centro de descargas, seleccionando como producto a descargar los datos LiDAR de segunda cobertura, con lo que se filtran los archivos de datos LiDAR que contienen puntos sobre la capa de polígonos cargada.

Dado que para todo el territorio existen archivos LiDAR descargables en formato color y en formato intensidad de retorno LiDAR, se aplica un filtro por nombre, solicitando al visor que mantenga únicamente aquellos archivos terminados en “COL.laz”, tras lo que se puede ejecutar la descarga masiva de todos los archivos a través de java.

Finalmente, mediante el uso de una herramienta programada en R para tal fin, se han cargado todos los archivos “laz” descargados en RStudio, y se ejecuta un recorte con la capa de polígonos generada con un buffer de 200 metros, para reducir el peso de las nubes de puntos, tras lo que se unifican y se exportan en formato “las” para su posterior procesamiento en el software Metashape.

5.4. Extracción de métricas LiDAR

La obtención de métricas LiDAR consiste en el cálculo de un gran número de estadísticos a partir de nubes de puntos LiDAR que se corresponden con cada uno de los transectos objeto de análisis, cálculo

que se ejecuta automáticamente sobre el software dedicado a la obtención de parámetros para la realización de inventarios forestales, FUSION.

Este procedimiento requiere de la normalización de las cotas de los puntos a partir de un modelo digital de elevaciones (MDE), para lo que es necesaria una clasificación de puntos de terreno que permita generar dicho MDE, además del recorte de las nubes de puntos para los polígonos individuales que conforman cada transecto analizado.

Si bien el software FUSION dispone de todas las herramientas para el cálculo de las métricas a partir de una nube de puntos LiDAR clasificada y con puntos con cotas elipsoidales u ortométricas, su uso se ha limitado al mismo cálculo de las métricas, delegando los procesos de clasificación, normalización y recorte de las nubes de puntos.

Los motivos de esta decisión son, por un lado, que si bien la nube de puntos utilizada ya está clasificada mediante el software de procesamiento de DJI y puede clasificarse con FUSION o diversos algoritmos de clasificación, el software Metashape de Agisoft permite un proceso de clasificación morfológica de puntos de terreno, que, ejecutado por pasos y adaptado por el operador a las condiciones de los terrenos analizados, junto con la posibilidad de visualización ágil de los resultados obtenidos para el ajuste de parámetros, permite clasificaciones de muy alta calidad en comparación con otras clasificaciones automatizadas. El máximo detalle de la clasificación es necesario en este caso para evitar la confusión de puntos de vegetación de baja cota como suelo, habitual en las clasificaciones automatizadas, cuando tenemos unos transectos que han sufrido incendios forestales y presentan en muchos casos vegetación de bajo porte, cuyas métricas perderíamos con una clasificación deficiente.

Por otra parte, la extensión geográfica de la superficie analizada, con transectos repartidos por buena parte de la geografía de Aragón, requiere una potencia de procesado en cuanto a las operaciones de recorte y normalización de la nube de puntos que no es asumible por el software FUSION, al menos en los equipos informáticos disponibles, por lo que dichas operaciones se han ejecutado a partir de distintas herramientas programadas expresamente para este estudio en lenguajes R y Python.

5.4.1. Clasificación de puntos de suelo

Para una clasificación eficiente en cuanto a potencia y tiempos de procesado, se ha procedido a distribuir las distintas nubes de puntos, uniéndolas con criterios geográficos para limitar la extensión de las superficies a procesar, dado que independientemente del número de puntos de las nubes LiDAR cargadas en Metashape, los tiempos de procesado dependen en gran medida de la extensión total del rectángulo sobre cuya superficie quedarían proyectadas según el eje Z todas las nubes de puntos cargadas en el software.

En el caso de los datos obtenidos a partir de los vuelos con el UAV, con este objetivo, a partir de las 30 nubes de puntos originales que abarcan la totalidad de los 76 transectos a analizar, se han generado 11 nubes de puntos, respecto de la extensión de todas las parcelas: 2 conglomerados al Noreste, 5 conglomerados al Noroeste, 1 en la zona central y 3 en el sur. Estas 11 nubes de puntos se han cargado independientemente en Metashape para su clasificación y generación de MDEs.

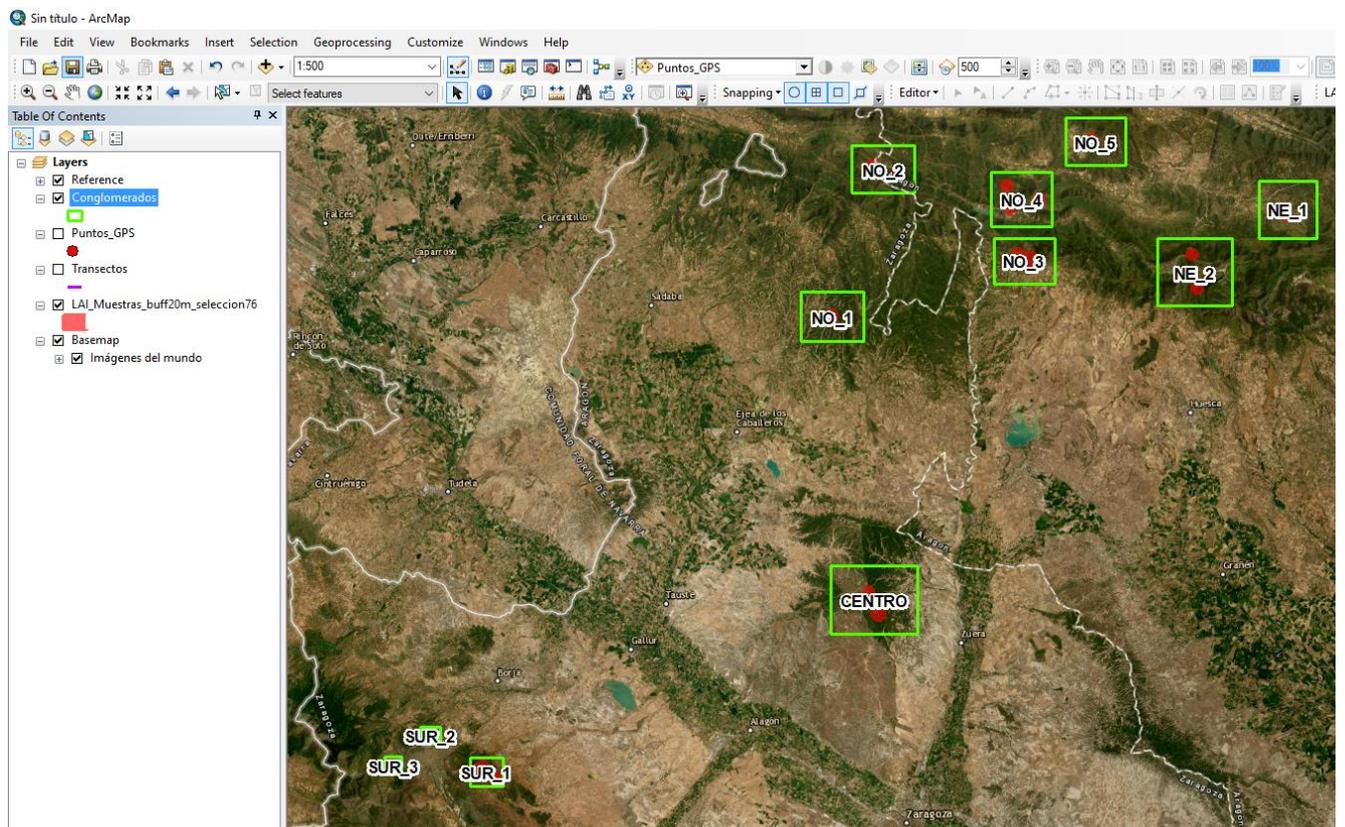


Figura 7. Distribución de conglomerados de nubes de puntos

En el caso de los datos del LiDAR del programa PNOA, se ha generado un único proyecto de Metashape para la clasificación, pues la notable diferencia en cuanto a puntos por metro cuadrado respecto a los vuelos de UAV ha permitido unos tiempos de procesamiento razonables para su ejecución.

Una vez cargadas las nubes de puntos, se han establecido tres pasos de clasificación, el objetivo del primer paso es separar los puntos pertenecientes a vegetación de altura media y alta y que pudiera tener una densidad tal que no permitiera al haz LiDAR alcanzar el terreno en grandes superficies, así como posibles construcciones que pudieran encontrarse en la zona, situación análoga morfológicamente a la anterior, para ello se establece un ángulo máximo entre puntos de 50° , una distancia máxima (altura) entre puntos de 90 centímetros y una ventana móvil de 30 metros.

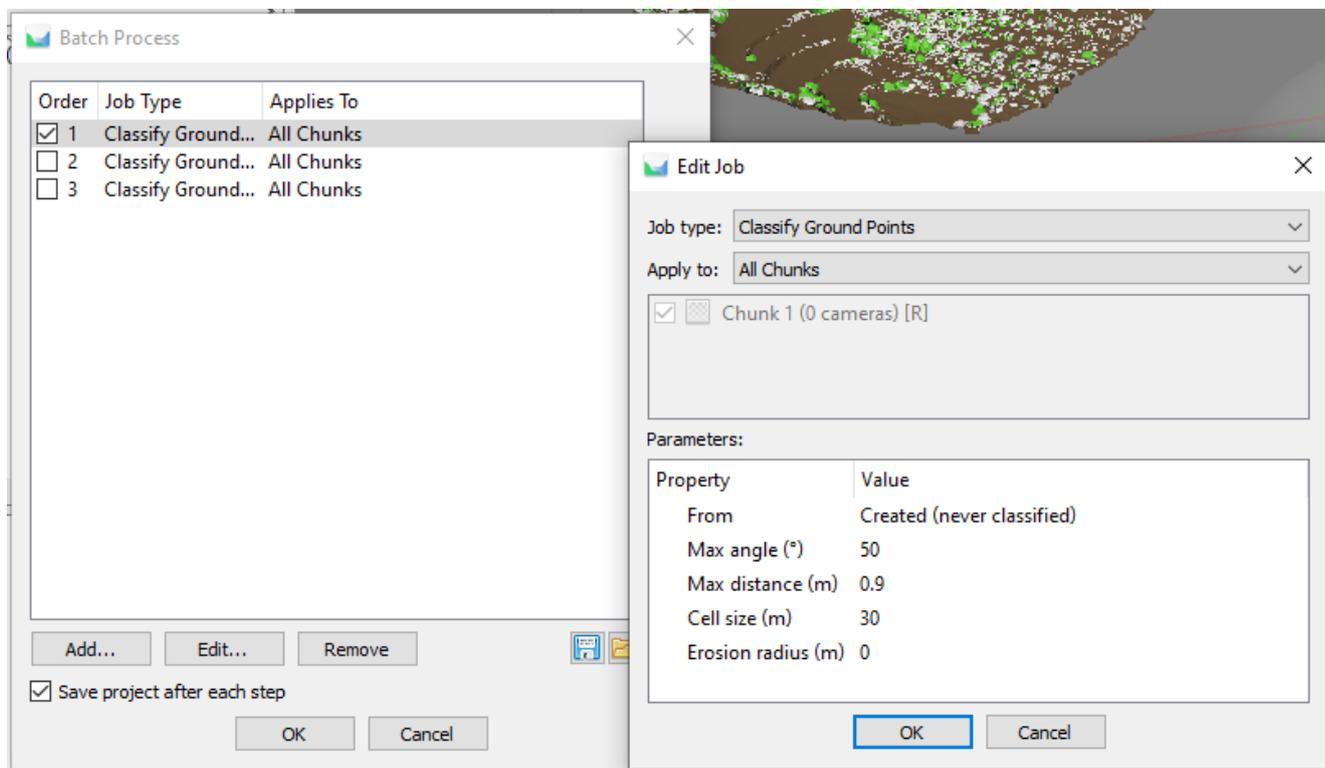


Figura 8. Vista de Metashape con la configuración del primer paso de clasificación

Para el segundo paso descartamos los puntos que no han sido clasificados como suelo en el primero, y aplicamos la segunda clasificación sobre los restantes. El objetivo de este paso es descartar puntos de vegetación media, pero conservando líneas de rotura del terreno (Cabezas de talud, bordes de caminos en pendiente, montículos, cimas, etc...) que pudieran ser susceptibles de ser clasificadas como no suelo, para ello, reducimos el ángulo entre puntos admisible y la altura máxima entre puntos a 25 grados y 40 centímetros respectivamente, no muy restrictivos, y reducimos la ventana móvil a 15 metros.

Llegados a este punto, los puntos de no suelo que aún no hemos conseguido apartar son, mayoritariamente, perímetros de arbustos y vegetación de bajo porte, para conseguir apartar estos puntos, se establecen unos criterios muy restrictivos, pero con una ventana móvil de pequeño tamaño que no permita perder información sobre grandes superficies entorno a líneas de rotura. El ángulo máximo entre puntos se mantiene en 20 grados dado que estamos en entornos de cierta pendiente, mientras la altura máxima entre puntos se reduce a 10 centímetros, y la ventana móvil, a 3 metros.

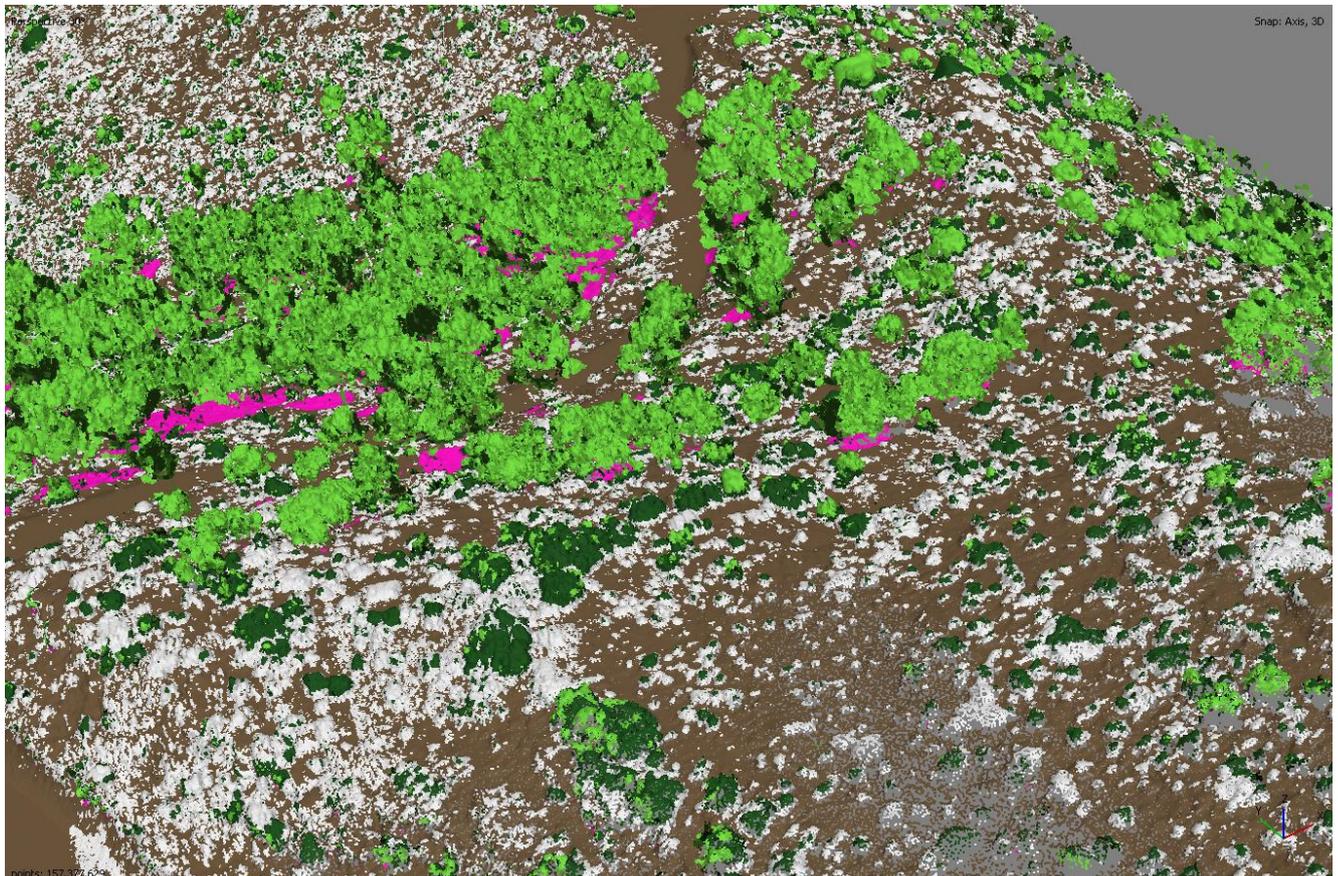


Figura 9. Clasificación de puntos, verde claro descartados en paso 1, verde oscuro en paso 2, blanco paso 3

5.4.2. Creación de modelos digitales de elevaciones

Clasificados los puntos de suelo, se genera una malla triangulada mediante algoritmos de interpolación de puntos, utilizando únicamente los puntos clasificados como suelo. La herramienta puede configurarse además para realizar una reducción de ruido en mayor o menor grado, que suaviza la malla resultante.

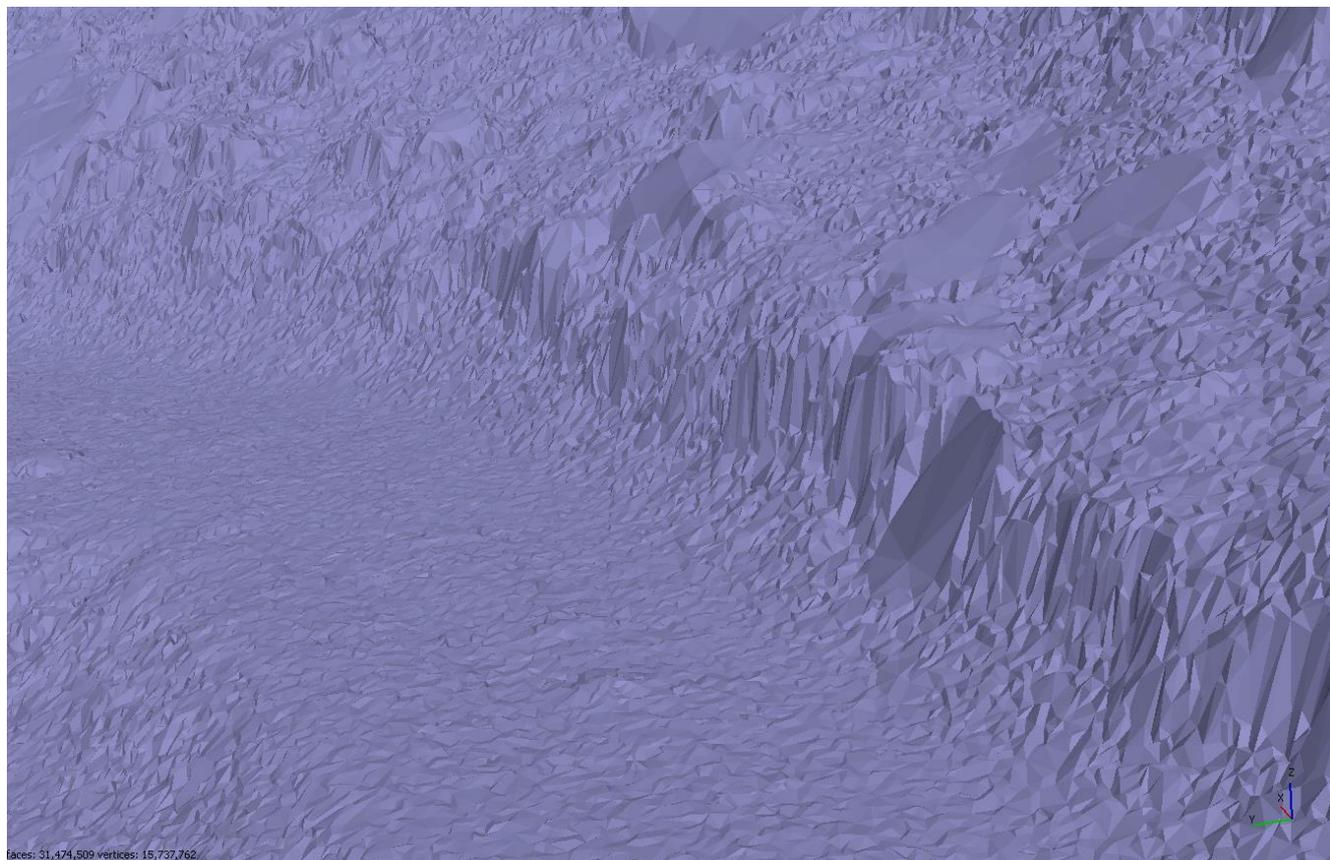


Figura 10. Detalle de malla triangulada

Una vez generada la malla triangulada, se genera sobre esta un modelo digital de elevaciones en formato raster, con una resolución espacial configurable, que en el caso de los modelos digitales de elevaciones generados para la nube de puntos LiDAR del UAV ha sido de un tamaño de celda de 30 centímetros de lado.

Para los datos LiDAR de 2 cobertura del PNOA, la resolución espacial del modelo digital de elevaciones generado ha sido un tamaño de celda de 50 centímetros de lado.

Los modelos digitales de elevaciones generados se exportan en formato raster geotiff y sistema de referencia ETRS89 UTM Huso 30 Norte con altura respecto al geoide EGM08, referencias utilizadas para todas las capas del proyecto.

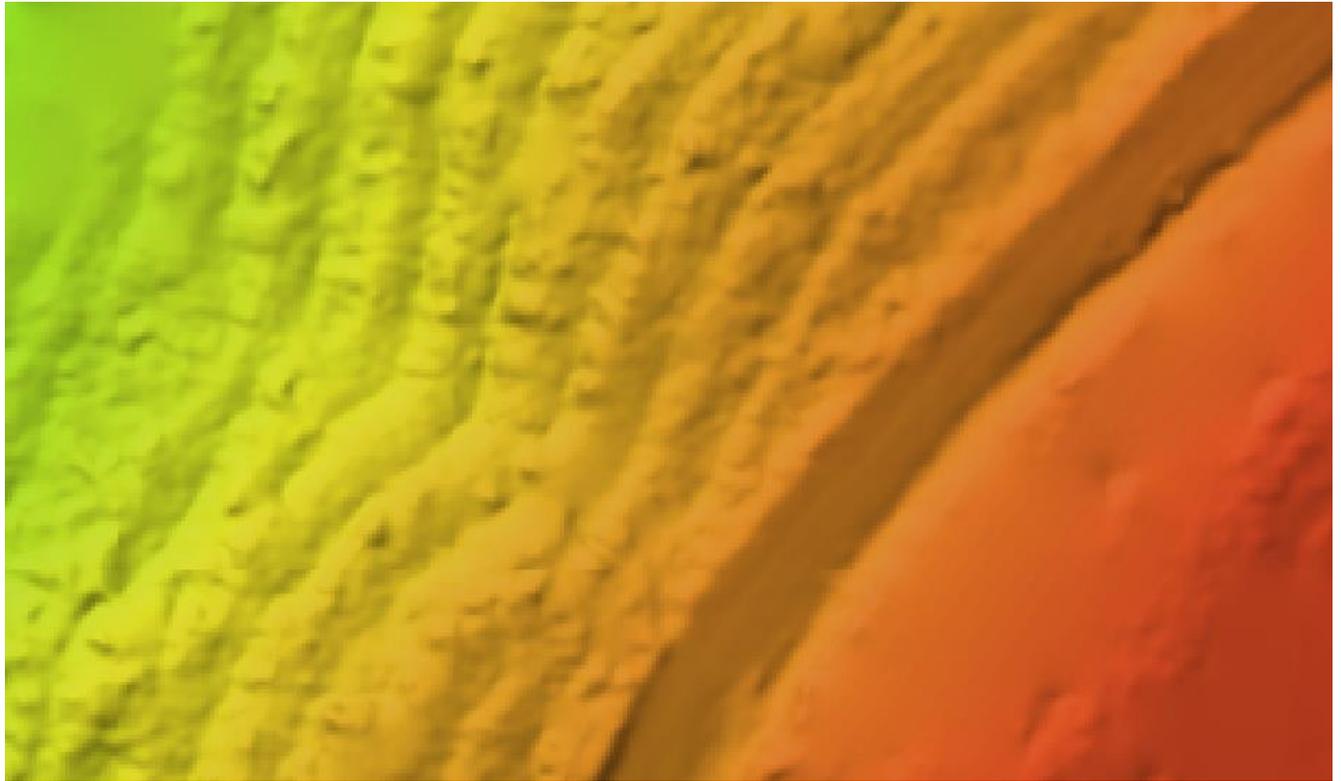


Figura 11. Detalle de uno de los modelos digitales de elevaciones

5.4.3. Normalización de nubes de puntos

La normalización de alturas consiste en un procedimiento por el que se resta al valor de cota de los puntos de la nube LiDAR el valor del terreno en el píxel del modelo digital de elevaciones sobre el que queda proyectado cada punto según el eje Z, obteniendo un valor de altura para cada punto relativo al terreno.

Para este procesamiento se ha programado una herramienta en lenguaje Python que realiza el proceso a partir de un número indeterminado de archivos de nubes de puntos alojados en una misma carpeta en formato .las, que son normalizados respecto a un modelo digital de elevaciones dado, por lo que, para reducir la potencia de procesamiento requerida, en el caso de los datos LiDAR obtenidos con el UAV se han agrupado en carpetas aquellas nubes de puntos provenientes de los vuelos y correspondientes a un mismo conglomerado de nubes, según la distribución detallada anteriormente, para los que se ha generado un único modelo digital de elevaciones, ejecutando la herramienta para cada conglomerado de forma independiente.

La herramienta genera una lista con los archivos de formato “las” contenidos en una ruta definida, y los carga uno a uno, obteniendo las coordenadas x e y de cada punto de la nube para identificar el valor correspondiente del modelo digital de elevaciones. Una vez obtenido el valor del terreno para cada punto LiDAR, la herramienta resta dicho valor al valor z de cada punto, obteniendo el valor normalizado, tras lo que sustituye el valor original z de cada punto por el valor normalizado calculado, para finalmente exportar cada nube de puntos en una ruta definida, nombrándolo por su nombre original con el término “_normalizado” añadido para su identificación.

5.4.4. Recorte de nubes de puntos

Una vez normalizados por conglomerados las nubes de puntos, estos han de ser separados en archivos individuales que contengan los puntos pertenecientes a cada uno de los transectos por separado, los cuales serán archivos de entrada en FUSION para ejecutar la herramienta “Cloudmetrics” que genera las métricas LiDAR por transecto.

Para realizar el recorte de las nubes de puntos para cada transecto, se ha programado una herramienta en lenguaje de programación R en la que se genera un bucle según el cual, cargada una capa

vectorial de polígonos y la nube de puntos a recortar, la herramienta selecciona consecutivamente cada uno de los polígonos de la capa vectorial y lo utiliza para recortar la nube de puntos de manera que conserva los puntos cuya proyección sobre el eje Z está contenida en el polígono, para cada uno de los recortes la herramienta exporta un archivo en formato “las” que contiene los puntos de un transecto, y queda renombrado según el nombre del transecto correspondiente, tomando dicha nomenclatura del valor correspondiente a para el polígono utilizado de un campo de la capa vectorial que contiene el nombre de la parcela.

Esta nomenclatura es importante, dado que el nombre de cada nube de puntos, al ser procesada mediante la herramienta “Cloudmetrics”, quedará guardado en la tabla de datos de salida por transecto, en un campo dedicado a incluir el nombre del archivo de origen, de manera que tendremos identificados cada uno de los transectos en la tabla de métricas LiDAR.

5.4.5. Reducción de densidad de nubes de puntos

Con el objetivo de analizar la viabilidad del producto LiDAR PNOA, además de la importancia de una alta resolución espacial, para estimar valores de LAI, se plantea la reducción de la densidad de las nubes de puntos obtenidas a partir de los vuelos del UAV DJI Matrice hasta alcanzar densidades semejantes a las obtenidas por un vuelo LiDAR del programa PNOA.

Para realizar dicha reducción de densidad se ha utilizado el paquete de herramientas Lastools para ArcMap. La herramienta lasthin de dicho paquete permite generar una rejilla con cuadrículas de un tamaño determinado y extensión similar a la de la nube de puntos cuya densidad se quiere reducir, conservando para cada cuadrícula un único punto de todos aquellos cuya proyección sobre el eje Z está contenida sobre dicha cuadrícula.

En cuanto al punto a conservar, el programa permite seleccionar un punto aleatorio, el punto más bajo o el punto más alto, en este caso se ha seleccionado un punto aleatorio, única opción lógica para simular una nube de puntos obtenida con un sensor de menos resolución espacial.

Para aumentar la eficiencia del procedimiento, se ha aplicado un procesado por lotes para las nubes de puntos normalizadas de los 76 transectos.

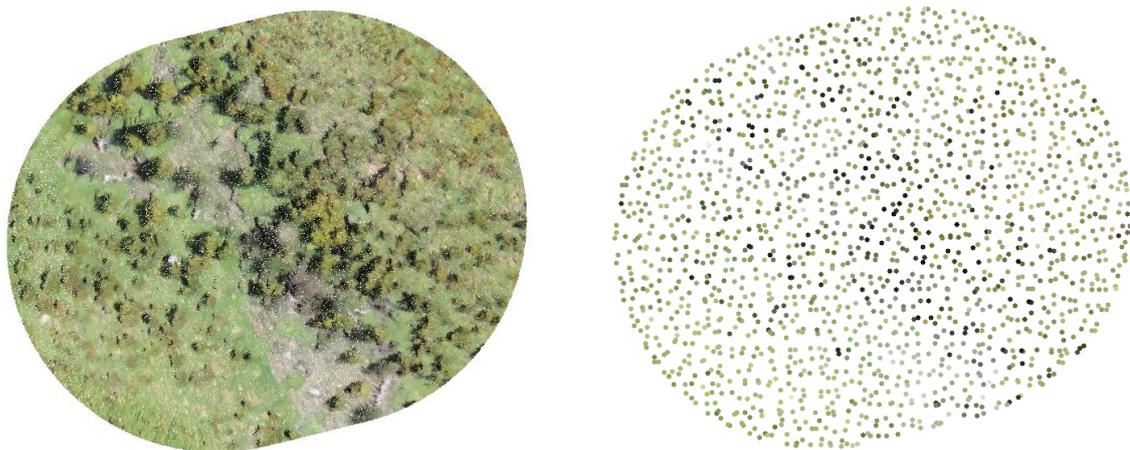


Figura 12. Izquierda, nube de puntos completa, derecha, nube con densidad reducida a 1 punto por m²

5.4.6. Cálculo de métricas LiDAR

Una vez disponemos de las nubes de puntos normalizadas e individualizadas para cada transecto podemos ejecutar la herramienta Cloudmetrics del software FUSION. La herramienta funciona a partir de comandos que se ejecutan sobre el símbolo de sistema de Windows, interprete de comandos nativo de dicho sistema operativo.

La línea de comando a construir llama a la herramienta concreta a utilizar de FUSION, en este caso “Cloudmetrics”, estipula la altura máxima y mínima de los puntos, que hemos definido en 30 metros y 0 metros respectivamente, por lo que la herramienta descartará aquellos que se sitúen fuera de ese rango, y también debemos definir la altura sobre la cual los puntos se consideran cobertura vegetal a efectos de definición de fracción de cabida cubierta, valor que hemos establecido en un metro, considerando la altura media sobre la que se sitúa el sensor LI-COR LAI2200c a lo largo de los transectos. Finalmente, se indica la ruta a un archivo txt ubicado en la misma carpeta en la que se encuentran los archivos de nubes de puntos normalizados y que contiene un listado con la ubicación completa de cada archivo, y se indica una ruta de salida para los archivos csv con los estadísticos calculados, generándose dos archivos, un archivo de estadísticas de los valores de elevación de los puntos, y un archivo de estadísticas de los valores de intensidad.

```
C:\FUSION\cloudmetrics.exe /above:1 /minht:0 /maxht:30
E:\MATRICE\Normalizados\Parcela\ListadoParcelas.txt
E:\MATRICE\Normalizados\Cloudmetrics\Matrice_parcelas_cloudmetrics.csv
```

Figura 12. Línea de comandos utilizada para la ejecución de la herramienta cloudmetrics

5.5. Cálculo de LAI a partir de imágenes Sentinel 2

5.5.1. Selección y descarga de imágenes

La selección de imágenes satelitales multiespectrales que nos permitieran obtener mediciones de LAI ha sido especialmente costosa en tiempo por la necesidad de obtener imágenes con fechas de captura similares a las fechas de visita de las distintas parcelas, lo que nos permite valorar las correlaciones con los datos de campo en ventanas de desarrollo de la vegetación similares.

Las condiciones climatológicas propias de la primavera y finales del invierno, junto con un mes de junio especialmente lluvioso han condicionado la selección de la plataforma a utilizar, descartando las imágenes Landsat 8 por no disponer de imágenes libres de nubes sobre algunas de las parcelas visitadas durante el mes de junio, escogiéndose un único programa espacial para realizar el contraste entre datos de campo y datos de imagen satélite, el programa Sentinel 2.

La plataforma utilizada para la descarga ha sido el ecosistema de datos espaciales Copernicus (Copernicus Data Space Ecosystem), que ofrece acceso gratuito e instantáneo a un amplio rango de datos a través de su navegador, y que viene a sustituir al antiguo Centro de acceso abierto de Copernicus (Copernicus Open Access Hub).

En dicho navegador se han configurado distintos tipos de filtrado para identificar las imágenes apropiadas para el trabajo, incluyendo la delimitación geográfica de las superficies de interés, el rango de fechas de adquisición de las imágenes, el máximo porcentaje de imagen ocupado por nubosidad y la plataforma satelital seleccionada.

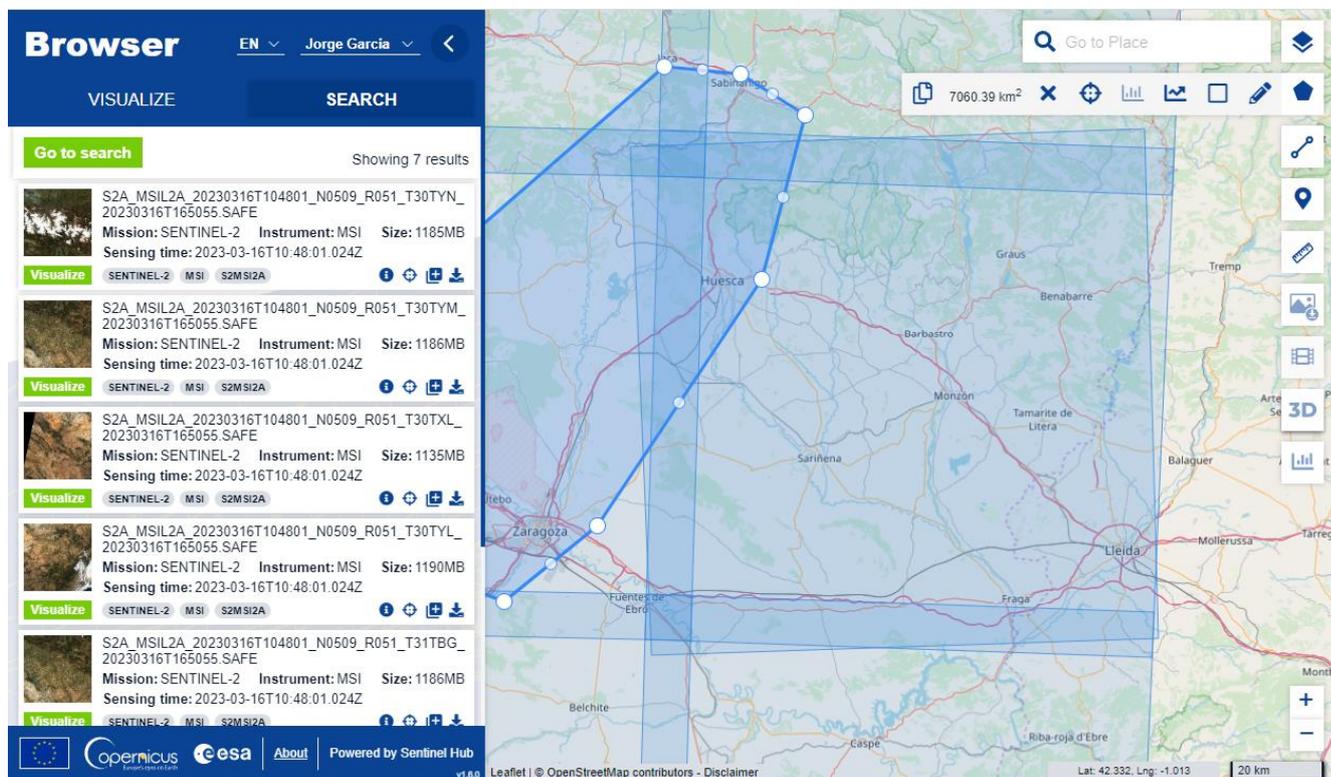


Figura 13. Vista del navegador de Copernicus Data Space Ecosystem. (Fuente: Programa Copernicus)

Ha sido necesaria en algunos casos la selección de distintas imágenes para transectos ejecutados en fechas similares, dado que la presencia de nubes no permitía obtener una única imagen válida para todos los transectos, concretamente durante el mes de junio.

Imagen Sentinel	Transectos
S2A_MSIL2A_20230316T104801_N0509_R051_T30TXN_20230316T165055	Visitados marzo 2023
S2A_MSIL2A_20230518T105621_N0509_R094_T30TYM_20230518T182451	Visitados mayo 2023
S2A_MSIL2A_20230617T105621_N0509_R094_T30TXN_20230617T184156	Visitados 16 junio 2023
S2A_MSIL2A_20230627T105621_N0509_R094_T30TXM_20230627T184856	Resto junio 2023

Tabla 3. Correspondencia imagen Sentinel utilizada y transectos de LAI asignados

5.5.2. Tratamiento de imágenes

El tratamiento de las imágenes Sentinel se realiza a partir del software SNAP de la ESA, utilizando la herramienta Biophysical Processor, el algoritmo de esta herramienta está basado en modelos radiativos que permiten estimar variables de estructura de la cubierta forestal observada, como el LAI, la fracción de cubierta cubierta o la fracción de radiación fotosintéticamente activa absorbida (FAPAR).

Se ha utilizado la herramienta en su configuración NNET, que permite la obtención del índice de área foliar (LAI), el contenido de clorofila en hoja (Cab), el contenido de agua de la cubierta vegetal (CWC), la fracción de radiación fotosintéticamente activa absorbida (FAPAR) y la fracción de cubierta cubierta (FVC). Para obtener dichos índices utiliza las bandas del satélite Sentinel: B3, B4, B5, B6, B7, B8a, B11 y B12.

Previamente a la ejecución de la herramienta, se cargan cada imagen Sentinel y se utiliza el software SNAP para generar una nueva imagen a partir de un subconjunto de bandas con las bandas utilizadas, y se realiza un remuestreo ajustando la resolución espacial a 20 metros, dado que todas las bandas deben tener la misma resolución para la ejecución del Biophysical Processor.

Mediante la ejecución del Biophysical Processor se genera una capa raster con tantas bandas como productos de salida solicitados, en este caso, una banda con el valor de LAI.

5.5.3. Extracción de valores de LAI

Generados los raster con valores de LAI para cada imagen Sentinel, se utiliza la capa vectorial de ROIs, a la que además del campo identificador de cada transecto, se agrega en ArcMap un campo de tipo numérico que contendrá los valores de LAI, tras lo que se procede a la selección de los elementos de dicha capa correspondientes a cada imagen Sentinel.

Con la selección activa, se utiliza la herramienta de análisis espacial “Zonal statistics as table” para generar una tabla que contenga un identificador común con la capa de ROIs y un campo con el valor medio del raster de LAI para cada uno de los polígonos de la capa vectorial. Dicha tabla se une a la capa ROIs a través del campo identificador común y se copian los valores de LAI extraídos sobre el campo generado anteriormente, incorporando los valores medios de LAI de cada transecto a la capa vectorial de ROIs original.

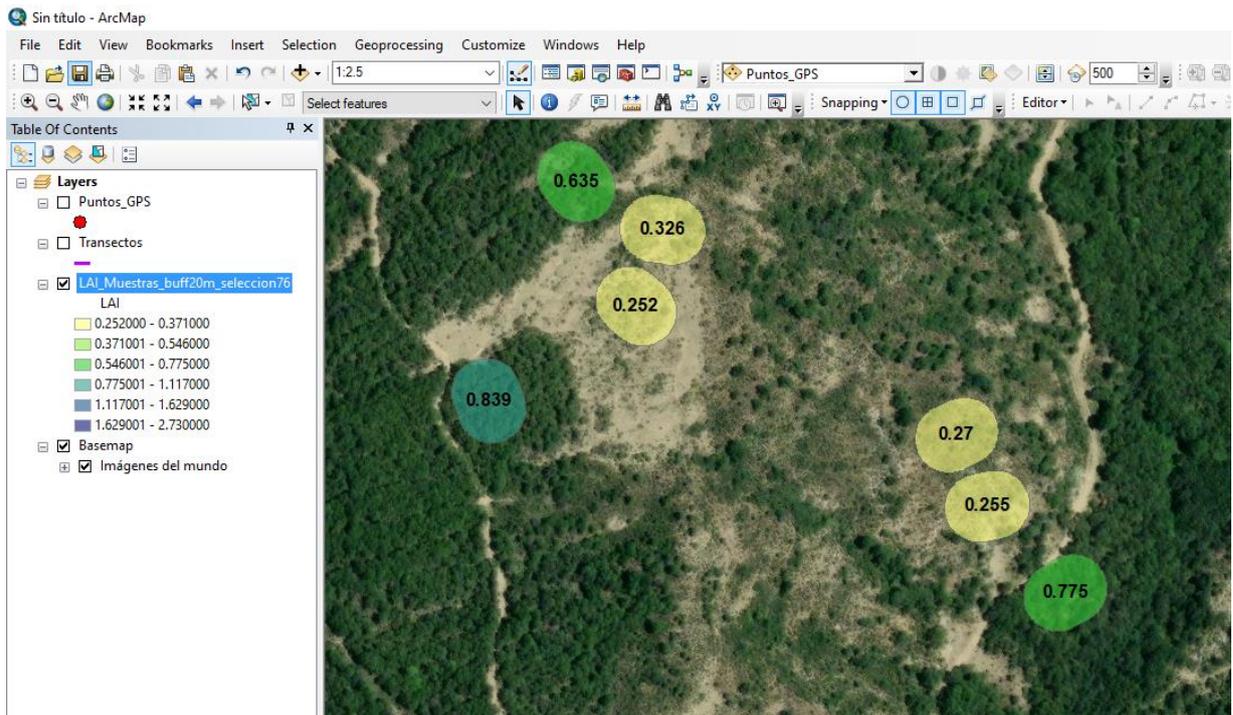


Figura 14. Vista de Arcmap con valores de LAI de distintos transectos

5.6. Unión de tablas con variables dependientes y estadísticos a analizar

A partir de los procedimientos nombrados anteriormente, se han generado, por un lado, una tabla de valores de LAI obtenidos a partir del sensor LI-COR y valores de LAI obtenidos a partir del procesamiento de las imágenes Sentinel 2, con un campo dedicado a la identificación de cada transecto. Por otro lado, se han generado las tablas de métricas LiDAR de los datos LiDAR PNOA, y de los datos LiDAR obtenidos a partir de los vuelos UAV tanto de alta resolución espacial como los de resolución espacial reducida para asemejarse a la resolución de los datos LiDAR PNOA, en todos los casos, igualmente, con un campo común que identifica cada uno de los transectos.

Con el objetivo de realizar el análisis estadístico de los datos, se ha realizado la unión, a partir del campo común, de la tabla de valores de LAI con cada una de las tablas de métricas LiDAR de manera independiente.

La unión se ha ejecutado mediante un script programado en R, con el que se cargan las dos tablas a unir, se convierten los campos numéricos de formato carácter a formato número para evitar problemas de formato posteriores en Excel, se eliminan elementos de la tabla que pudieran contener campos sin valor (NA), se unen mediante el campo común y se exportan en formato Excel xlsx.

5.7. Análisis estadístico

Durante el análisis estadístico se aplican distintos análisis sobre los datos generados con el objetivo de, en base a la naturaleza de los datos, seguir un procedimiento adecuado para obtener modelos de regresión apropiados para la estimación del valor de LAI utilizando como variables dependientes tanto la medición de LAI del sensor LI-COR LAI 2200c como la medición obtenida a partir de imágenes Sentinel 2. Las variables independientes que analizar serán, en cada caso, las métricas LiDAR obtenidas del vuelo UAV, las obtenidas del vuelo UAV con densidad reducida, y las obtenidas del vuelo LiDAR PNOA de segunda cobertura.

5.7.1. Filtrado de datos

Se realiza un primer análisis mediante el que se pretende describir los datos y filtrar aquellas variables independientes que no van a ser de utilidad para el modelo de regresión por su propia naturaleza, concretamente por presentar valores invariables para los distintos transectos o una cantidad significativa de ceros que reduzca notablemente la variabilidad de la variable, descartando su utilidad en el análisis de modelos de regresión.

Para realizar una primera aproximación a los datos, se ha generado un script en R mediante el que se cargan los datos y se genera una tabla reducida con los 6 primeros valores de cada campo para su observación, y un resumen estadístico para cada variable dependiente mediante el que podemos observar sus valores mínimo, medio y máximo, así como los cuartiles primero, segundo (mediana) y tercero.

```
11 - ##### Resumen estadístico de tablas de datos #####
12 file_path <- "F://TFM/ANALISIS_ESTADISTICO/TABLAS_BASE/LICOR_SENTINEL2_MATRICE_OK_CSV_R.CSV"
13
14 # Cargamos los datos desde el archivo csv usando punto y coma (;) como delimitador
15 data <- read.csv(file_path, sep=';')
16
17 # Configuramos opciones para mostrar todas las columnas
18 options(max.print=9999)
19
20 # Mostramos las primeras filas
21 data_head <- head(data)
22
23 # Resumen estadístico con más detalle
24 data_description <- summary(data)
25
26 # Transponemos el resumen estadístico para una mejor visualización
27 data_description_transposed <- t(data_description)
28
29 # Mostramos las primeras filas y el resumen estadístico transpuesto
30 print(data_head)
31 print(data_description_transposed)
--
```

Figura 15. Fragmento de código de aproximación a los datos

Tras este primer análisis, ejecutamos en R una serie de comandos de limpieza con el objetivo de eliminar de los datos todos aquellos campos que tengan un valor único para todos los transectos, así como todos aquellos que contengan un número significativo de valores igual a 0, para lo que se establece como valor límite que el 25 por ciento de los valores sean igual a 0 (primer cuartil = 0). Obteniendo una tabla de datos en la que se han descartado dichos campos.

```

34 - ##### Limpiamos los datos eliminando las columnas con valores únicos o que contienen ceros significativos#####
35
36 # verificar si alguna columna tiene un único valor único o una alta cantidad de ceros.
37 unique_value_columns <- colnames(data)[sapply(data, function(col) length(unique(col)) <= 1)]
38 high_zero_columns <- colnames(data)[sapply(data, function(col) sum(col == 0) > (0.25 * nrow(data)))]
39
40 # Columnas a eliminar: la unión de ambas listas anteriores
41 columns_to_drop <- unique(c(unique_value_columns, high_zero_columns))
42
43 # Eliminar las columnas identificadas
44 data_cleaned <- data[, !names(data) %in% columns_to_drop]
45
46 # verificar la nueva tabla y las columnas que han sido eliminadas
47 data_cleaned_shape <- dim(data_cleaned)
48 dropped_columns <- columns_to_drop
49
50 # Imprimir la nueva forma del dataframe y las columnas eliminadas
51 cat("Nueva forma del dataframe:", data_cleaned_shape, "\n")
52 cat("Columnas eliminadas:", dropped_columns, "\n")

```

Figura 16. Descarte de datos con variabilidad reducida

5.7.2. Detección de valores atípicos y estadísticas descriptivas de variables dependientes

Previamente a la selección del método estadístico apropiado para obtener un modelo de regresión que nos permita estimar las variables dependientes, es necesario realizar un análisis exploratorio de los datos que nos permita orientar los próximos pasos del análisis estadístico.

Para avanzar en el conocimiento de los datos y la elección de los próximos pasos, vamos a estudiar la existencia de valores atípicos (outliers) y estadísticas descriptivas que nos permitan obtener una visión global de las variables dependientes que estamos analizando.

Para alcanzar este objetivo, se considera la creación de diagramas de cajas para ambas variables dependientes, considerando además la principal característica de los transectos al objeto del presente estudio, que hubieran sido afectados o no por incendios forestales. Por otra parte, se obtendrán los cuartiles y estadísticos básicos de ambas variables para profundizar el análisis de variabilidad de los datos.

Ambas herramientas se obtendrán a partir de un script programado en R para tal fin, utilizando la librería “ggplot2” para la representación de diagramas.

```

55 - #####
56 - ##### DIAGRAMAS DE CAJA PARA LAS VARIABLES DEPENDIENTES #####
57 # Cargar bibliotecas necesarias
58 library(ggplot2)
59
60 # Establecer el estilo de los gráficos
61 theme_set(theme_minimal())
62
63 # Crear un diagrama de cajas para las variable dependiente 'LAI'
64 ggplot(data_cleaned, aes(x = Sector, y = LAI, fill = Sector)) +
65   geom_boxplot(alpha = 0.7) +
66   labs(title = "Diagrama de cajas variable LAI") +
67   scale_fill_manual(values = c("C" = "blue", "Q" = "orange")) +
68   theme_minimal() +
69   theme(plot.title = element_text(hjust = 0.5))
70
71 # Crear un diagrama de cajas para las variable dependiente 'LAIS2'
72 ggplot(data_cleaned, aes(x = Sector, y = LAIS2, fill = Sector)) +
73   geom_boxplot(alpha = 0.7) +
74   labs(title = "Diagrama de cajas variable LAI sentinel 2") +
75   scale_fill_manual(values = c("C" = "blue", "Q" = "orange")) +
76   theme_minimal() +
77   theme(plot.title = element_text(hjust = 0.5))
78
79 # Obtener un resumen estadístico de las variables dependientes
80 summary(data_cleaned$LAI)
81 summary(data_cleaned$LAIS2)
82
83
84
85 - #####

```

Figura 17. Obtención de diagramas de caja y estadísticos básicos

5.7.3. Análisis de distribución normal y prueba de Shapiro-Wilk

Además de la presencia de valores atípicos no identificados con fallos en los procesos de medición o debidos a otros condicionantes conocidos, un análisis de la normalidad de la distribución de los datos es necesario para seleccionar un método de regresión adecuado.

Con el objeto de realizar dicho análisis se han generado histogramas de frecuencias para las variables dependientes, y se ha ejecutado una prueba de Shapiro-Wilk para contrastar la normalidad de ambos conjuntos de datos.

De nuevo tanto la elaboración de histogramas como la ejecución de la prueba de Shapiro-Wilk se han realizado en un entorno de programación en R, mediante un script desarrollado a tal fin.

```
95 - #####
96 - ##### ANÁLISIS DE DISTRIBUCIÓN NORMAL - TEST SHAPIRO-WILK #####
97
98 # Cargar las bibliotecas necesarias
99 library(ggplot2)
100 library(dplyr)
101 library(e1071) # Para la prueba de shapiro-wilk
102
103 # Función para probar la normalidad
104 test_normality <- function(data, column_name) {
105   shapiro_test_result <- shapiro.test(data[[column_name]])
106   return(shapiro_test_result$p.value)
107 }
108
109 # Realizar pruebas de normalidad para las variables dependientes
110 p_value_LAI <- test_normality(data_cleaned, 'LAI')
111 p_value_LAIS2 <- test_normality(data_cleaned, 'LAIS2')
112
113 # Crear histogramas para las variables dependientes
114 hist_LAI <- ggplot(data_cleaned, aes(x = LAI)) +
115   geom_histogram(binwidth = 0.1, fill = "lightblue", color = "blue", alpha = 0.7) +
116   geom_density(aes(y = 0.1 * ..count..), linetype = "solid", color = "black") +
117   labs(title = "Histograma de frecuencias para la variable LAI", y = "Cuenta") +
118   theme(plot.title = element_text(hjust = 0.5))
119
120 hist_LAIS2 <- ggplot(data_cleaned, aes(x = LAIS2)) +
121   geom_histogram(binwidth = 0.1, fill = "lightblue", color = "blue", alpha = 0.7) +
122   geom_density(aes(y = 5 * ..scaled..), linetype = "solid", color = "black") +
123   labs(title = "Histograma de frecuencias para la variable LAI Sentinel 2", y = "Cuenta") +
124   theme(plot.title = element_text(hjust = 0.5))
125
126 # Mostrar los resultados de las pruebas de normalidad y los histogramas
127 cat("P-value for normality test (LAI):", p_value_LAI, "\n")
128 cat("P-value for normality test (LAIS2):", p_value_LAIS2, "\n")
129
130 # Imprimir los histogramas (esto debe hacerse en una sesión interactiva de R)
131 print(hist_LAI)
132 print(hist_LAIS2)
133
134
135 - #####
```

Figura 18. Script para construcción e histogramas y prueba de Shapiro-Wilk

5.7.4. Análisis de correlaciones y colinealidad de variables independientes

Las obtención de métricas LiDAR a partir del software FUSION, que constituyen la totalidad de las variables independientes a utilizar en los modelos de regresión para la estimación de valores de LAI, tienen la particularidad de que se generan un alto número de variables que están íntimamente relacionadas, dado que parten de la obtención de estadísticos similares a partir de un único dato, la posición respecto de un sistema de referencia con coordenadas X, Y y Z, de cada uno de los puntos de una nube LiDAR.

Esta característica hace totalmente indispensable un análisis detallado de las correlaciones y posible colinealidad que puedan existir entre todas estas variables, que permita una selección de variables linealmente no dependientes para la obtención de un modelo robusto que no sufra de una multicolinealidad que pueda generar un sobre ajuste del modelo.

De este modo, tras la valoración de la hipótesis de normalidad de la distribución de las variables dependientes, se establecerá una metodología para evaluar las correlaciones, generando una matriz de correlaciones a partir del método de Pearson, en el caso de distribución normal de la muestra, o Spearman en el caso de descartar la hipótesis de normalidad para la mayoría de las muestras.

Para la eliminación de variables altamente correlacionadas, se establecerá un valor de corte de la matriz de correlaciones de 0.75.

```

142 ▾ #####
143 ▾ ##### ANÁLISIS DE CORRELACIONES #####
144
145 library(caret)
146 library(corrplot)
147
148 # Preparar los datos
149 dependent_var <- data_cleaned[, 4]
150 independent_vars <- data_cleaned[, 6:93]
151
152 ▾ ###Análisis de correlaciones y colinealidad####
153 # calculamos la matriz de correlación
154 cor_matrix <- cor(independent_vars, method = "spearman")
155
156 # Identificamos las variables altamente correlacionadas y las eliminamos
157 highlyCorrelated <- findCorrelation(cor_matrix, cutoff = 0.75) |
158 independent_vars_filtered <- independent_vars[, -highlyCorrelated]
159
160 ▾ #####

```

Figura 19. Análisis de correlaciones mediante método de Spearman

Respecto de la colinealidad de las variables independientes, una vez descartadas aquellas con una alta correlación, se calcula para las restantes el Factor de Inflación de la Varianza (VIF) que indica cuánto se infla la varianza de un coeficiente de regresión lineal debido a la correlación entre las variables independientes en el modelo.

$$VIF_i = \frac{1}{1 - R_i^2}$$

Donde R^2 es el coeficiente de determinación (R-cuadrado) de una regresión de la variable independiente i sobre todas las otras variables independientes en el modelo. Un VIF que excede 5 o 10 sugiere que la regresión de la variable independiente sobre las otras variables independientes es fuerte, y, por tanto, esa variable independiente está altamente coalineada con las otras, por lo que se ha establecido el valor de corte para la selección de variables en VIF igual a 5.

```

160 ▾ #####
161 ▾ ##### ANÁLISIS DE COLINEALIDAD DE VARIABLES INDEPENDIENTES #####
162
163 library(car)
164 library(caret)
165
166 # Calculamos VIF para las variables restantes
167 vif_results <- vif(lm(dependent_var ~ ., data = independent_vars_filtered))
168 print(vif_results)
169 vif_filtered <- independent_vars_filtered[, which(vif_results < 5)]
170
171 # Actualizamos las variables independientes a las filtradas con coeficiente de inflacion de la varianza < 5
172 independent_vars <- vif_filtered
173
174
175 ▾ #####

```

Figura 20. Análisis de colinealidad con Factor de Inflación de la Varianza (VIF)

5.7.5. Análisis de regresión LASSO

El modelo de regresión LASSO (Least Absolute Shrinkage and Selection Operator), Tibshirani, R. (1996), es una técnica de regresión que implica la penalización de los coeficientes de la regresión. Esta técnica es útil para la selección de variables, ayudando a prevenir el sobreajuste y a mejorar la interpretabilidad del modelo al producir un modelo más simple y comprensible.

Este modelo produce una reducción de variables inherente al método, de gran utilidad en el presente análisis dado que, pese al filtrado de variables independientes realizado en el paso anterior, es previsible que un número importante de ellas lleguen a esta etapa, donde el modelo seleccionado simplificará nuevamente la cantidad de variables a utilizar y, por tanto, el modelo resultante y su interpretación.

Por otra parte, dado que LASSO penaliza el módulo de los coeficientes de la regresión, la escala de las distintas variables dependientes puede suponer una penalización indeseada sobre aquellas variables que por una cuestión de escala requieran coeficientes mas altos de entrada en el modelo de regresión, para evitar este efecto, se aplicara una normalización previa de las variables independientes.

Tras la ejecución del modelo obtendremos las variables independientes seleccionadas junto con los coeficientes asignados para el modelo de regresión, así como el coeficiente de correlación múltiple R² y el R² ajustado.

```
175 ~ #####
176 ~ ##### APLICAMOS EL MODELO DE REGRESIÓN LASSO #####
177
178 library(glmnet)
179 library(MASS) # para shapiro.test
180
181 # Normalizamos las variables independientes
182 preProcValues <- preProcess(independent_vars, method = c("center", "scale"))
183 independent_vars_scaled <- predict(preProcValues, independent_vars)
184
185 # Convertimos a matriz ya que glmnet no acepta dataframes
186 independent_vars_matrix <- as.matrix(independent_vars_scaled)
187
188 # Aplicamos la regresión Lasso
189 set.seed(123) # Para reproducibilidad
190 cv_fit <- cv.glmnet(independent_vars_matrix, dependent_var, alpha = 1)
191
192 # Obtenemos los coeficientes al mejor lambda y excluimos el intercepto
193 best_coef <- coef(cv_fit, s = "lambda.min")
194 best_coef <- best_coef[-1,] # Excluye el intercepto
195
196 # Identificamos las variables con coeficientes diferentes de cero
197 selected_features <- names(best_coef[best_coef != 0])
198
199 # Si hay variables seleccionadas, procedemos a extraerlas del dataframe
200 selected_independent_vars <- independent_vars_scaled[, selected_features, drop = FALSE]
201
202 # Ajustamos un modelo de regresión lineal con las variables seleccionadas por Lasso
203 selected_independent_vars <- independent_vars_scaled[, selected_features, drop = FALSE]
204
205 # Ajustar el modelo de regresión lineal
206 final_linear_model <- lm(dependent_var ~ ., data = selected_independent_vars)
207
208 # Resumen del modelo
209 summary(final_linear_model)
210
211 ~ #####
```

Figura 21. Aplicación del modelo de regresión LASSO

5.7.6. Prueba de normalidad y homocedasticidad de los residuos

Para confirmar la robustez de los modelos de regresión generados, debe comprobarse que los residuos presentan una distribución normal y que la varianza de los residuos es constante con independencia del valor que adquieren las variables independientes (homocedasticidad).

Con este objetivo, una vez calculados los modelos, se ejecuta la prueba de Shapiro-Wilk sobre los residuos, considerando que no se puede rechazar la hipótesis de normalidad e los residuos cuando el p-value obtenido sea igual o mayor que 0.05.

En cuanto a la hipótesis nula de homocedasticidad, se ejecuta la prueba de homocedasticidad de Breusch-Pagan, según la cual, si se obtiene un p-value mayor que 0.05, no se puede rechazar la hipótesis de homocedasticidad.

```

211 - #####
212 - ### Pruebas de normalidad de los residuos y homocedasticidad de Breusch-Pagan ###
213
214 # Prueba de normalidad de los residuos con el test de Shapiro-wilk
215 shapiro_test <- shapiro.test(residuals(final_linear_model))
216
217 # Mostrar los resultados de la prueba de Shapiro-wilk
218 print(shapiro_test)
219
220 # si la prueba de Shapiro-wilk tiene un p-value < 0.05, los residuos no se consideran normalmente distribuidos
221
222 # Prueba de homocedasticidad con la prueba de Breusch-Pagan
223 bptest_result <- bptest(final_linear_model)
224
225 # Mostrar los resultados de la prueba de Breusch-Pagan
226 print(bptest_result)
227
228 # si la prueba de Breusch-Pagan tiene un p-value < 0.05, se sugiere heterocedasticidad
229
230 # Visualizar los residuos
231 plot(residuals(final_linear_model))
232
233 - #####

```

Figura 22. Pruebas de distribución normal y homocedasticidad de los residuos

6. RESULTADOS

6.1. Resumen estadístico de variables dependientes

En el contexto de un primer análisis estadístico de los datos con los que se trabaja, se realiza una primera valoración de la distribución de los datos de las dos variables dependientes, contenidas en los campos 3 y 4, que corresponden respectivamente a los valores de LAI calculados a partir de datos Sentinel 2 y el sensor LI-COR.

Mediante el resumen estadístico obtenido en R podemos observar una notable disparidad entre ambos datos, con un valor máximo de 2.73 para los datos Sentinel 2 y 6.8 para los datos del sensor LI-COR, o una mediana de 0.83 y 1.82 para los datos Sentinel 2 y LI-COR respectivamente.

Transecto	Length:76	Class :character	Mode :character					
Sector	Length:76	Class :character	Mode :character					
LAI_S2	Min. :0.2520	1st Qu.:0.4280	Median :0.8325	Mean :0.9470	3rd Qu.:1.3920	Max. :2.7300		
LAI	Min. :0.050	1st Qu.:1.215	Median :1.820	Mean :2.069	3rd Qu.:2.688	Max. :6.800		

Figura 23. Resumen estadístico de variables dependientes

6.2. Filtrado de datos UAV Matrice

Un primer resumen estadístico de los datos nos permite observar que disponemos de una tabla de datos con 105 variables o campos para un total de 76 transectos analizados. Los dos primeros campos pertenecen a variables descriptivas del estudio, de tipo texto y que describen el código identificativo de cada transecto y si este fue realizado sobre superficie afectada por incendios forestales o superficie de control.

Transecto	Length:76	Class :character	Mode :character					
Sector	Length:76	Class :character	Mode :character					
LAI52	Min. :0.2520	1st Qu.:0.4280	Median :0.8325	Mean :0.9470	3rd Qu.:1.3920	Max. :2.7300		
LAI	Min. :0.050	1st Qu.:1.215	Median :1.820	Mean :2.069	3rd Qu.:2.688	Max. :6.800		
Totalreturncount	Min. : 672219	1st Qu.:1151147	Median :1409939	Mean :1451653	3rd Qu.:1704933	Max. :3359604		
Totalreturncountabove0andbelow30	Min. : 299646	1st Qu.: 632824	Median : 974300	Mean :1006890	3rd Qu.:1274466	Max. :2862696		
Return1countabove0andbelow30	Min. : 299004	1st Qu.: 617387	Median : 928725	Mean : 976115	3rd Qu.:1223296	Max. :2862696		
Return2countabove0andbelow30	Min. : 0	1st Qu.: 0	Median : 8408	Mean : 29659	3rd Qu.: 23642	Max. :324372		
Return3countabove0andbelow30	Min. : 0.0	1st Qu.: 0.0	Median : 5.5	Mean : 1116.9	3rd Qu.: 103.8	Max. :35555.0		
Return4countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return5countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return6countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return7countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return8countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return9countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Otherreturncountabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Elevminimum	Min. :0.001	1st Qu.:0.001	Median :0.001	Mean :0.001	3rd Qu.:0.001	Max. :0.001		

Figura 24. Descarte de datos con variabilidad reducida

Como podemos observar también en la figura 23, existe un número significativo de variables independientes para las que todos sus valores son iguales, o un gran número de ellos son igual a cero.

Una vez ejecutado el script de descarte de variables, conservamos una tabla con 93 variables para 76 elementos, tras la eliminación de 12 de las variables independientes pretendidas.

Entre las variables eliminadas encontramos:

Return 2 count above 0 and below 30

Return 3 count above 0 and below 30

Return 4 count above 0 and below 30

Return 5 count above 0 and below 30

Return 6 count above 0 and below 30

Return 7 count above 0 and below 30

Return 8 count above 0 and below 30

Return 9 count above 0 and below 30

Other return count above 0 and below 30

Elev minimum

Int minimum (Mínimo del rango intercuartilico)

Int P01 (Rango intercuartilico P01)

6.3. Filtrado de datos UAV Matrice densidad reducida

En el caso de los datos Matrice de densidad reducida a 1 punto por metro cuadrado también partimos de una tabla de 105 variables con 76 transectos, con la misma estructura que la anterior.

Transecto	Length:76	Class :character	Mode :character					
Sector	Length:76	Class :character	Mode :character					
LAI52	Min. :0.2520	1st Qu.:0.4280	Median :0.8325	Mean :0.9470	3rd Qu.:1.3920	Max. :2.7300		
LAI	Min. :0.050	1st Qu.:1.215	Median :1.820	Mean :2.069	3rd Qu.:2.688	Max. :6.800		
Totalreturncount	Min. :1643	1st Qu.:1725	Median :1740	Mean :1745	3rd Qu.:1749	Max. :1962		
Totalreturncountabove0andbelow30	Min. : 449	1st Qu.: 974	Median :1293	Mean :1197	3rd Qu.:1477	Max. :1697		
Return1countabove0andbelow30	Min. : 448.0	1st Qu.: 949.2	Median :1230.5	Mean :1163.1	3rd Qu.:1442.8	Max. :1631.0		
Return2countabove0andbelow30	Min. : 0.00	1st Qu.: 0.00	Median : 8.00	Mean : 32.21	3rd Qu.: 31.25	Max. :288.00		
Return3countabove0andbelow30	Min. : 0.000	1st Qu.: 0.000	Median : 0.000	Mean : 1.211	3rd Qu.: 0.000	Max. :36.000		
Return4countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return5countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return6countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return7countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return8countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Return9countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Otherreturncountabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0		
Elevminimum	Min. :0.001000	1st Qu.:0.001000	Median :0.001000	Mean :0.001039	3rd Qu.:0.001000	Max. :0.002000		
Elevmaximum	Min. : 1.323	1st Qu.: 5.098	Median : 7.382	Mean : 7.651	3rd Qu.: 9.571	Max. :18.174		
Elevmean	Min. :0.07824	1st Qu.:0.46635	Median :1.14647	Mean :1.76020	3rd Qu.:2.66611	Max. :6.50649		
Elevmode	Min. :0.001000	1st Qu.:0.001000	Median :0.001000	Mean :0.001039	3rd Qu.:0.001000	Max. :0.002000		
Elevstddev	Min. :0.1592	1st Qu.:0.7066	Median :1.5742	Mean :1.7265	3rd Qu.:2.2099	Max. :4.6784		
Elevvariance	Min. : 0.02536	1st Qu.: 0.49930	Median : 2.47806	Mean : 4.27921	3rd Qu.: 4.88371	Max. :21.88738		
ElevIQ	Min. :0.3965	1st Qu.:0.9362	Median :1.3220	Mean :1.4155	3rd Qu.:1.7357	Max. :3.1380		
ElevIQ	Min. : 0.0560	1st Qu.: 0.3581	Median : 1.4901	Mean : 2.4486	3rd Qu.: 3.8333	Max. :10.1225		
Elevskewness	Min. :-1.4422	1st Qu.: 0.5224	Median : 1.6727	Mean : 1.9230	3rd Qu.: 2.8238	Max. : 9.9859		
Elevkurtosis	Min. : 1.262	1st Qu.: 2.069	Median : 5.014	Mean : 11.288	3rd Qu.: 12.442	Max. :146.866		
ElevAAD	Min. :0.07999	1st Qu.:0.46368	Median :1.17485	Mean :1.40234	3rd Qu.:1.92463	Max. :4.24736		

Figura 25. Vista reducida de resumen estadístico de la tabla Matrice con densidad de 1 pto/m2

Si bien la tabla de datos también contiene diversos campos con una variabilidad reducida debido a valores iguales para un campo y campos con un porcentaje de 0 superior al 25% de las mediciones, en este caso se han descartado 11 variables, de las cuales 7 contaban con el mismo valor para todos sus elementos, y 4 con un porcentaje de 0 igual o superior al 25%.

Entre las variables eliminadas encontramos:

Return 2 count above 0 and below 30

Return 3 count above 0 and below 30

Return 4 count above 0 and below 30

Return 5 count above 0 and below 30

Return 6 count above 0 and below 30

Return 7 count above 0 and below 30

Return 8 count above 0 and below 30

Return 9 count above 0 and below 30

Other return count above 0 and below 30

Int minimum (Mínimo del rango intercuartílico)

Int P01 (Rango intercuartílico P01)

6.4. Filtrado de datos LiDAR PNOA segunda cobertura

La estructura de tabla de datos vista anteriormente se repite para los datos LiDAR PNOA de segunda cobertura, con 105 variables y 76 transectos.

Transecto	Length:76	Class :character	Mode :character						
Sector	Length:76	Class :character	Mode :character						
LAI52	Min. :0.2520	1st Qu.:0.4280	Median :0.8325	Mean :0.9470	3rd Qu.:1.3920	Max. :2.7300			
LAI	Min. :0.050	1st Qu.:1.215	Median :1.820	Mean :2.069	3rd Qu.:2.688	Max. :6.800			
TotalReturncount	Min. : 1047	1st Qu.: 2308	Median : 3290	Mean : 4380	3rd Qu.: 5630	Max. :17406			
TotalReturncountabove0andbelow30	Min. : 359	1st Qu.: 1406	Median : 2819	Mean : 3456	3rd Qu.: 3697	Max. :16614			
Return1countabove0andbelow30	Min. : 359	1st Qu.: 1156	Median : 1954	Mean : 2696	3rd Qu.: 3146	Max. :11264			
Return2countabove0andbelow30	Min. : 0.0	1st Qu.: 65.5	Median : 267.0	Mean : 625.1	3rd Qu.: 856.8	Max. :4026.0			
Return3countabove0andbelow30	Min. : 0.00	1st Qu.: 1.75	Median : 20.50	Mean : 119.04	3rd Qu.: 147.75	Max. :1189.00			
Return4countabove0andbelow30	Min. : 0.00	1st Qu.: 0.00	Median : 0.00	Mean : 14.42	3rd Qu.: 12.00	Max. :207.00			
Return5countabove0andbelow30	Min. : 0.000	1st Qu.: 0.000	Median : 0.000	Mean : 1.171	3rd Qu.: 1.000	Max. :32.000			
Return6countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0			
Return7countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0			
Return8countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0			
Return9countabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0			
OtherReturncountabove0andbelow30	Min. :0	1st Qu.:0	Median :0	Mean :0	3rd Qu.:0	Max. :0			
Elevmininum	Min. :0.001000	1st Qu.:0.001000	Median :0.001000	Mean :0.001329	3rd Qu.:0.001000	Max. :0.008000			
Elevmaxinum	Min. : 1.350	1st Qu.: 4.787	Median : 7.435	Mean : 7.893	3rd Qu.: 9.794	Max. :20.353			
Elevmean	Min. :0.1159	1st Qu.:0.6471	Median :1.5493	Mean :1.9730	3rd Qu.:2.6463	Max. :7.1611			
Elevmode	Min. :0.001	1st Qu.:0.001	Median :0.001	Mean :0.206	3rd Qu.:0.156	Max. :9.676			
Elevstddev	Min. :0.1772	1st Qu.:0.7122	Median :1.6082	Mean :1.7682	3rd Qu.:2.4826	Max. :4.1002			
Elevvariance	Min. : 0.03139	1st Qu.: 0.50727	Median : 2.58700	Mean : 4.40620	3rd Qu.: 6.16312	Max. :16.81134			
ElevCV	Min. :0.5679	1st Qu.:0.8372	Median :0.9854	Mean :1.0498	3rd Qu.:1.2052	Max. :2.2243			
ElevIQ	Min. :0.1165	1st Qu.:0.5975	Median :2.0925	Mean :2.7416	3rd Qu.:3.9872	Max. :8.3153			
Elevskewness	Min. : -0.4986	1st Qu.: 0.4767	Median : 1.2414	Mean : 1.4863	3rd Qu.: 2.2388	Max. : 6.8640			
Elevkurtosis	Min. : 1.322	1st Qu.: 1.906	Median : 3.608	Mean : 7.265	3rd Qu.: 9.481	Max. :79.613			
ElevAAD	Min. :0.09802	1st Qu.:0.53153	Median :1.31521	Mean :1.47765	3rd Qu.:2.17454	Max. :3.71716			
ElevMADmedian	Min. :0.0450	1st Qu.:0.2315	Median :0.5675	Mean :1.0756	3rd Qu.:1.6780	Max. :3.8195			
ElevMADmode	Min. :0.0630	1st Qu.:0.2633	Median :0.7157	Mean :1.5255	3rd Qu.:1.9534	Max. :7.9230			
ElevL1	Min. :0.1159	1st Qu.:0.6471	Median :1.5493	Mean :1.9730	3rd Qu.:2.6463	Max. :7.1611			
ElevL2	Min. :0.06741	1st Qu.:0.35346	Median :0.86200	Mean :0.93627	3rd Qu.:1.35604	Max. :2.34362			
ElevL3	Min. : -0.32415	1st Qu.: 0.07879	Median : 0.15970	Mean : 0.17942	3rd Qu.: 0.29084	Max. : 0.69424			
ElevL4	Min. : -0.25090	1st Qu.: -0.02540	Median : 0.03846	Mean : 0.02377	3rd Qu.: 0.08565	Max. : 0.21301			
ElevLCV	Min. :0.3273	1st Qu.:0.4611	Median :0.5188	Mean :0.5131	3rd Qu.:0.5742	Max. :0.6832			
ElevLskewness	Min. : -0.1419	1st Qu.: 0.1689	Median : 0.3097	Mean : 0.2889	3rd Qu.: 0.4258	Max. : 0.6843			
ElevLkurtosis	Min. : -0.13044	1st Qu.: -0.01957	Median : 0.09588	Mean : 0.11036	3rd Qu.: 0.21768	Max. : 0.54742			
ElevP01	Min. :0.001580	1st Qu.:0.006918	Median :0.010420	Mean :0.017977	3rd Qu.:0.020250	Max. :0.183440			

Figura 26. Vista reducida de resumen estadístico de la tabla LiDAR PNOA

Para esta fuente de datos, la mayor capacidad de penetración y rebote del haz LiDAR montado en los vuelos tripulados se hace patente en un menor número de campos con valor 0, especialmente aquellos dedicados al conteo de puntos presentes en la nube de puntos y pertenecientes a los primeros rebotes, encontrando valores distintos a 0 para más de un 75% de la muestra de los segundos y terceros rebotes captados por el sensor.

En este caso, establecidos los mismos criterios de filtrado que en análisis anteriores, se han descartado un total de 7 campos, 5 campos por contener un único valor para todos los transectos, y 2 campos por contener valores iguales a cero para un 25% o más de la muestra.

Entre las variables descartadas encontramos:

Return 4 count above 0 and below 30

Return 5 count above 0 and below 30

Return 6 count above 0 and below 30

Return 7 count above 0 and below 30

Return 8 count above 0 and below 30

Return 9 count above 0 and below 30

Other return count above 0 and below 30

6.5. Detección de valores atípicos y análisis de estadísticas descriptivas

Los diagramas de cajas para las variables dependientes “LAI” y “LAIS2” muestran la distribución de los datos por sector, variable que determina si los transectos fueron afectados por incendios forestales (Q) o son transectos de control que no fueron afectados por los incendios (C).

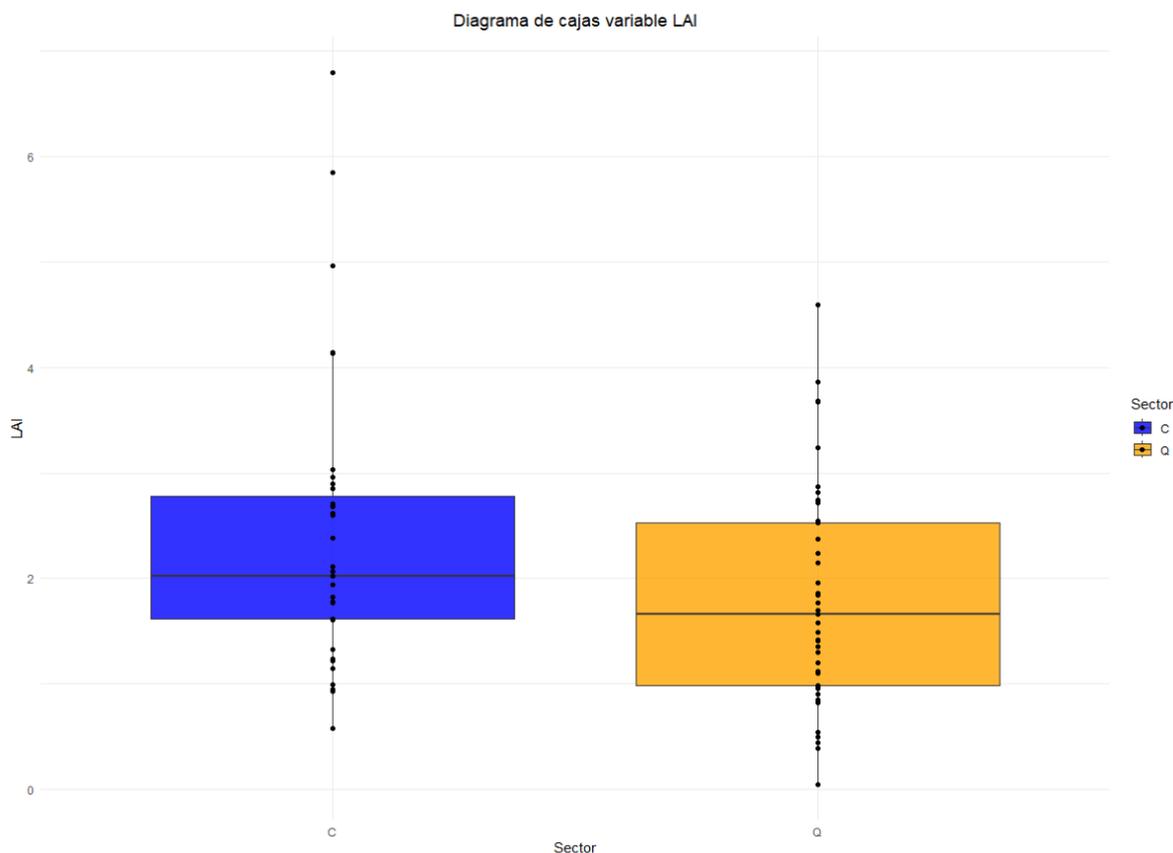


Figura 27. Diagrama de cajas para la variable LAI proveniente de datos de campo

Como se puede observar, la variable dependiente LAI presenta una gran dispersión, con numerosos valores atípicos y una mayor dispersión hacia valores altos de LAI. Por otra parte, la mediana y los cuartiles parecen diferir entre los dos sectores para la variable, lo que parece indicar que la distribución de la variable LAI se ve afectada por los incendios forestales.

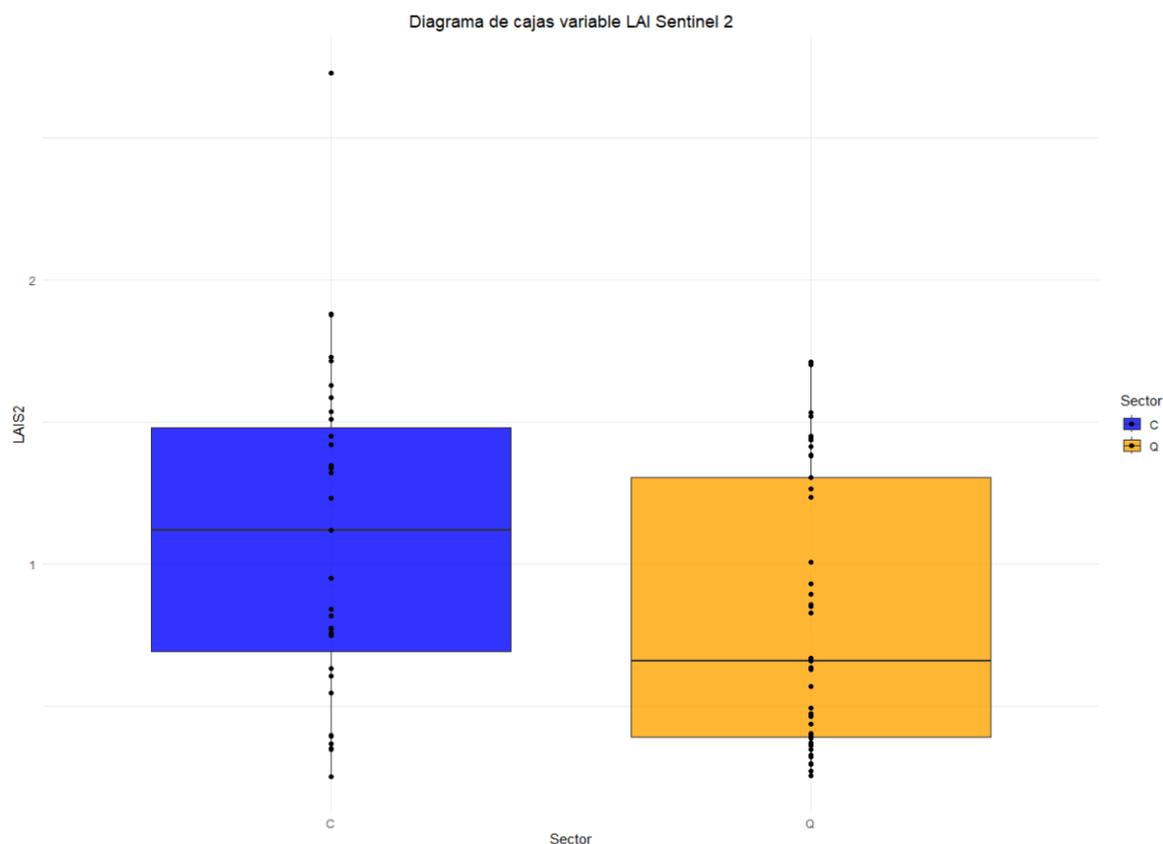


Figura 28. Diagrama de cajas para la variable LAI proveniente de datos Sentinel 2

En cuanto a la variable dependiente LAI proveniente de datos Sentinel 2, también observamos la presencia, en menor medida, de valores atípicos y dispersión de los datos. Igualmente, se confirma la tendencia observada para la variable LAI proveniente de datos de campo, donde la mediana y los cuartiles parecen indicar que la distribución de los datos está relacionada con la afección de los incendios forestales.

Las estadísticas descriptivas confirman las observaciones anteriores, con una desviación estándar notablemente mayor en la variable LAI respecto a LAIS2, o una media también notablemente mayor

Estadístico	LAI	LAI S2
Mínimo	0.050	0.252
Primer cuartil	1.215	0.428
Mediana	1.820	0.833
Tercer cuartil	2.688	1.392
Máximo	6.800	2.730
Media	2.069	0.947
Desviación estándar	1.246	0.535

Tabla 4. Estadísticos descriptivos para ambas variables dependientes

6.6. Análisis de distribución normal y prueba de Shapiro-Wilk

La visualización de los histogramas de frecuencia y las pruebas de Shapiro-Wilk para ambas variables dependientes, permiten descargar la hipótesis de distribución normal para ambas variables dependientes.

En cuanto a la variable dependiente LAI procedente de datos de campo, el p-value obtenido para la prueba de normalidad es de 6.981×10^{-5} , valor muy por debajo del límite inferior de 0.05 para la consideración de un conjunto de datos con distribución normal. Respecto al histograma, podemos caracterizar la distribución de frecuencias como unimodal leptocúrtica con sesgo positivo.

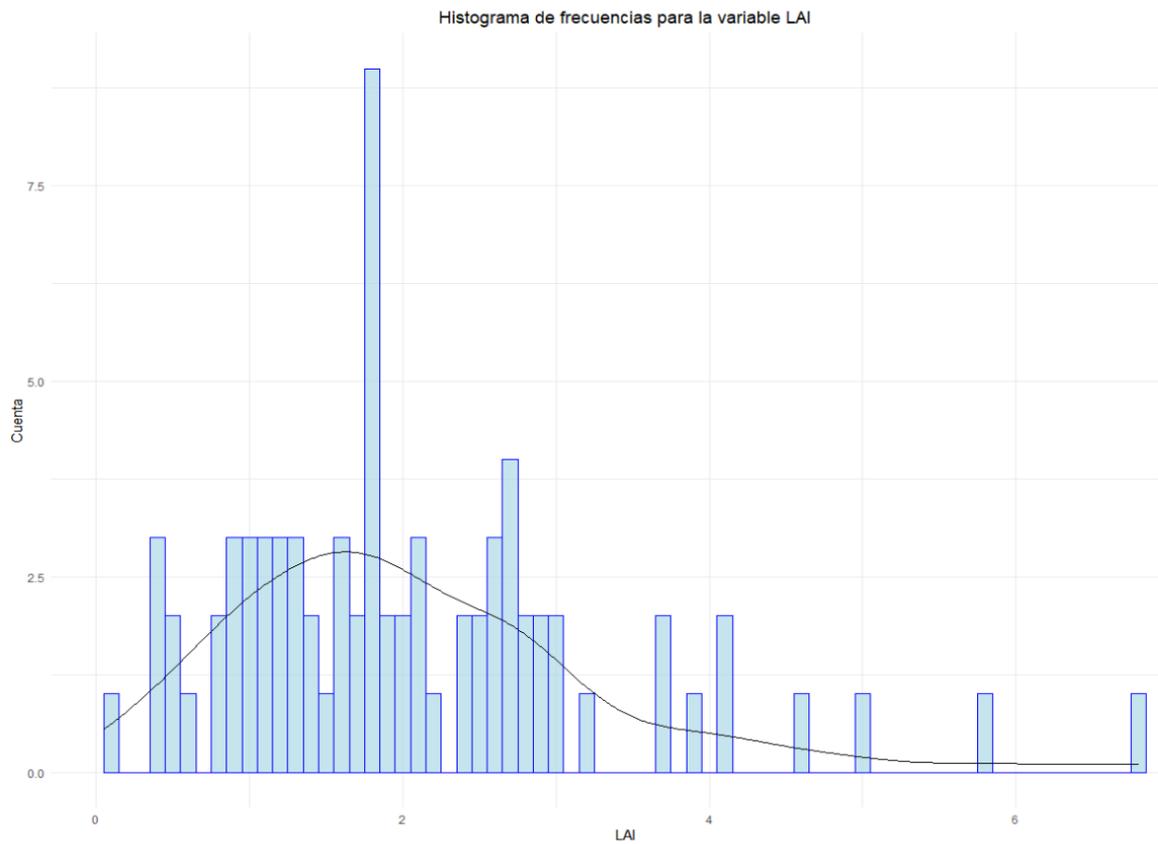


Figura 29. Histograma de frecuencias de la variable LAI

La variable dependiente LAIS2 presenta un p-value obtenido para la prueba de Shapiro-Wilk de 0.000130 que también nos permite descartar la hipótesis de distribución normal. El histograma de frecuencias para el conjunto de datos de Sentinel 2 nos permite caracterizar la distribución como bimodal leptocúrtica con sesgo positivo.

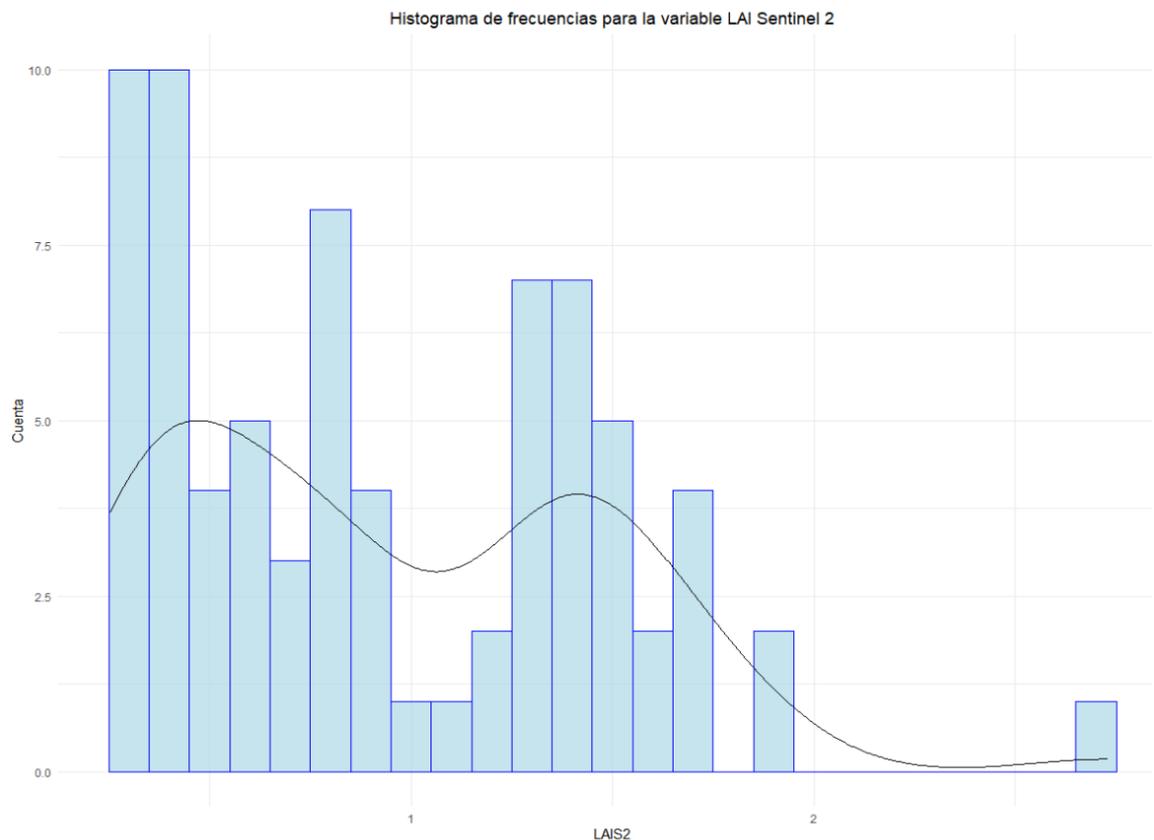


Figura 30. Histograma de frecuencias de la variable LAI Sentinel 2

En cuanto a la valoración de la hipótesis de normalidad de la muestra para muestras separadas en función de la fecha de adquisición de los datos, valorando distintos estados vegetativos de la cubierta vegetal, se observa una aproximación de las distribuciones hacia la hipótesis de distribución normal, cumpliéndose incluso la condición de p-value para la muestra de valores de LAI obtenidos con el sensor LI-COR durante el mes de marzo.

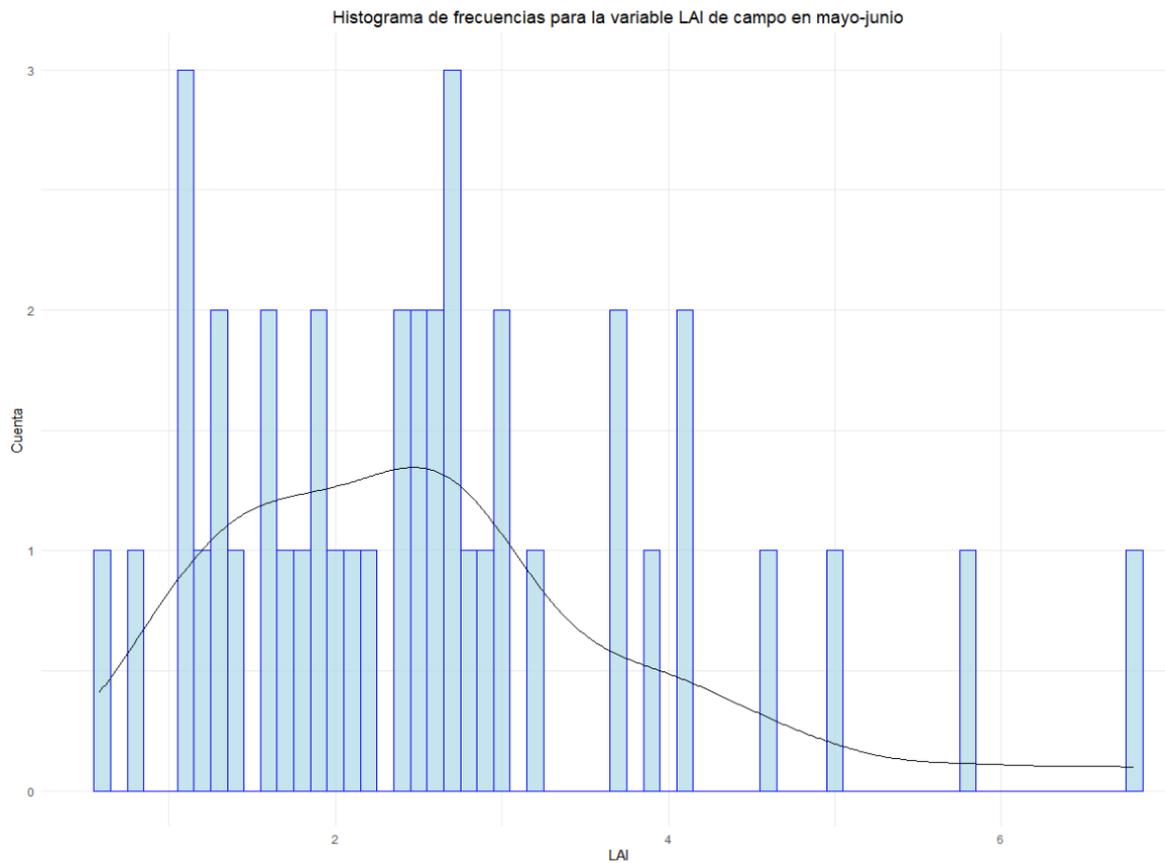


Figura 31. Histograma de la variable LAI LI-COR de las muestras de mayo-junio, p-value =0.00102

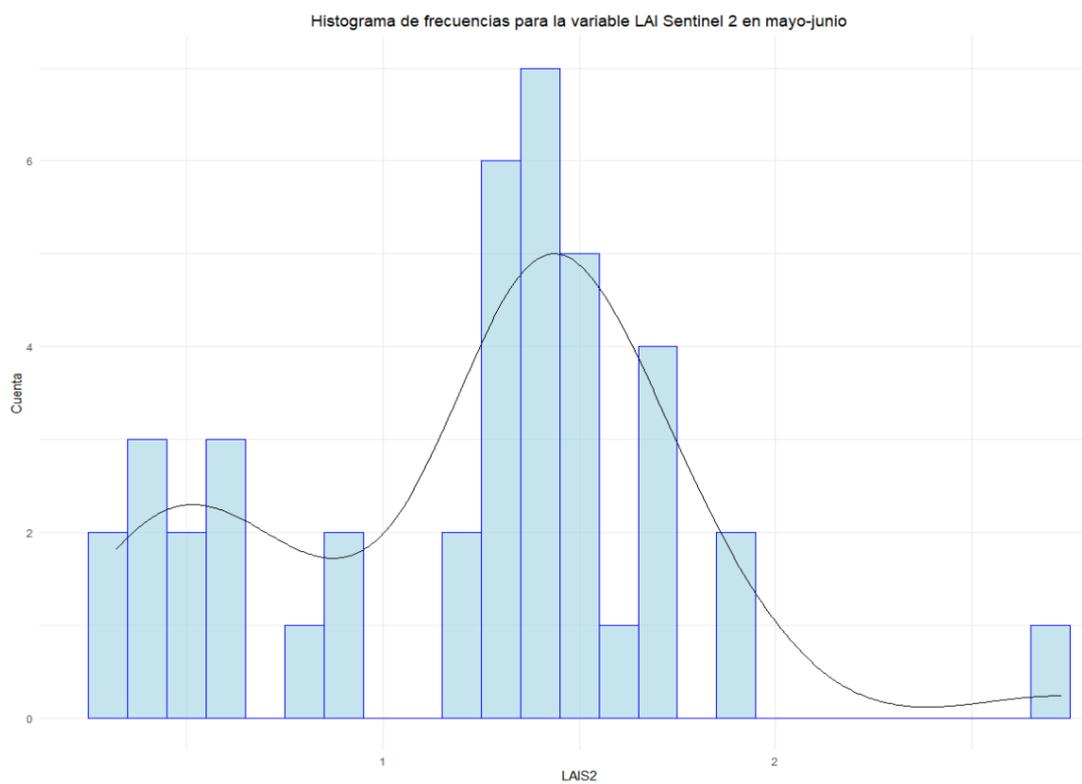


Figura 32. Histograma de la variable LAI Sentinel 2 de las muestras de mayo-junio, p-value =0.0056

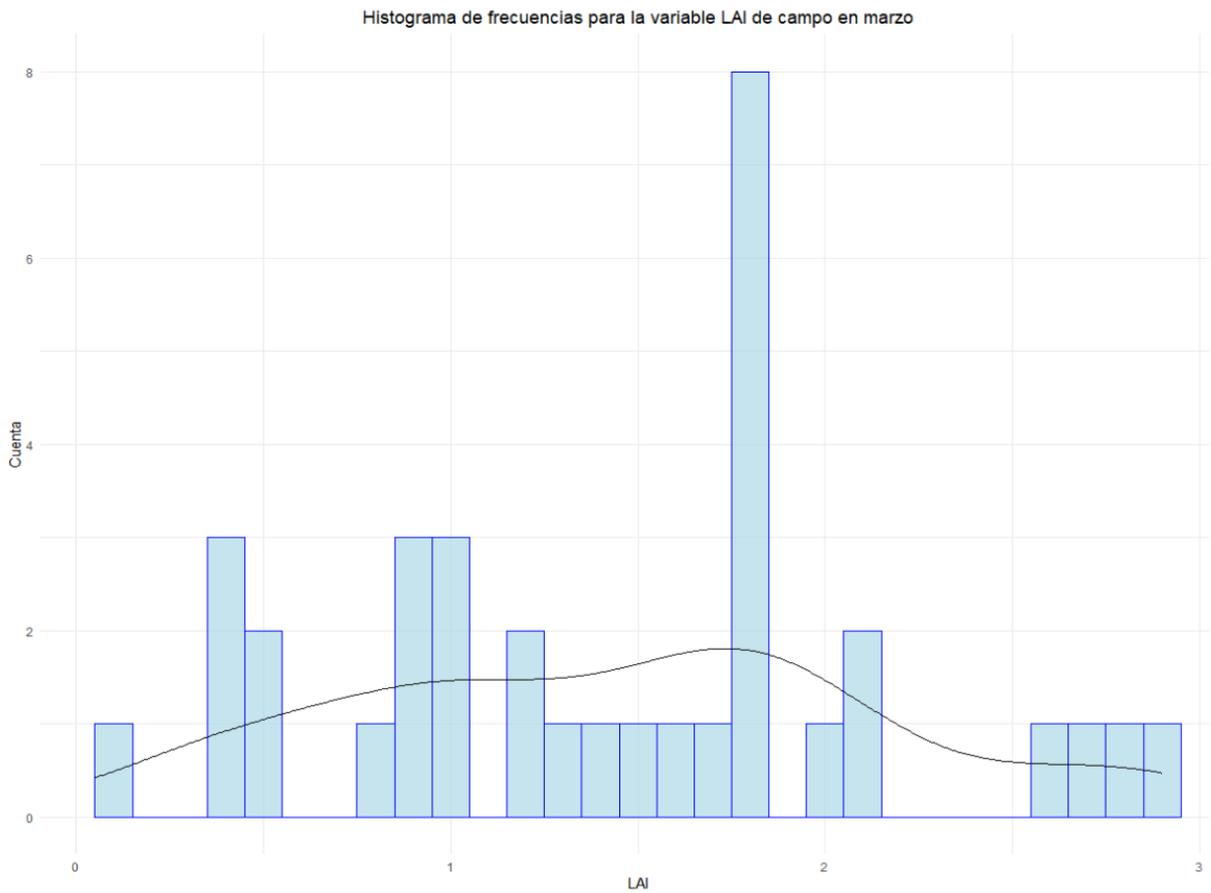


Figura 33. Histograma de la variable LAI LI-COR de las muestras de marzo, p-value =0.4045

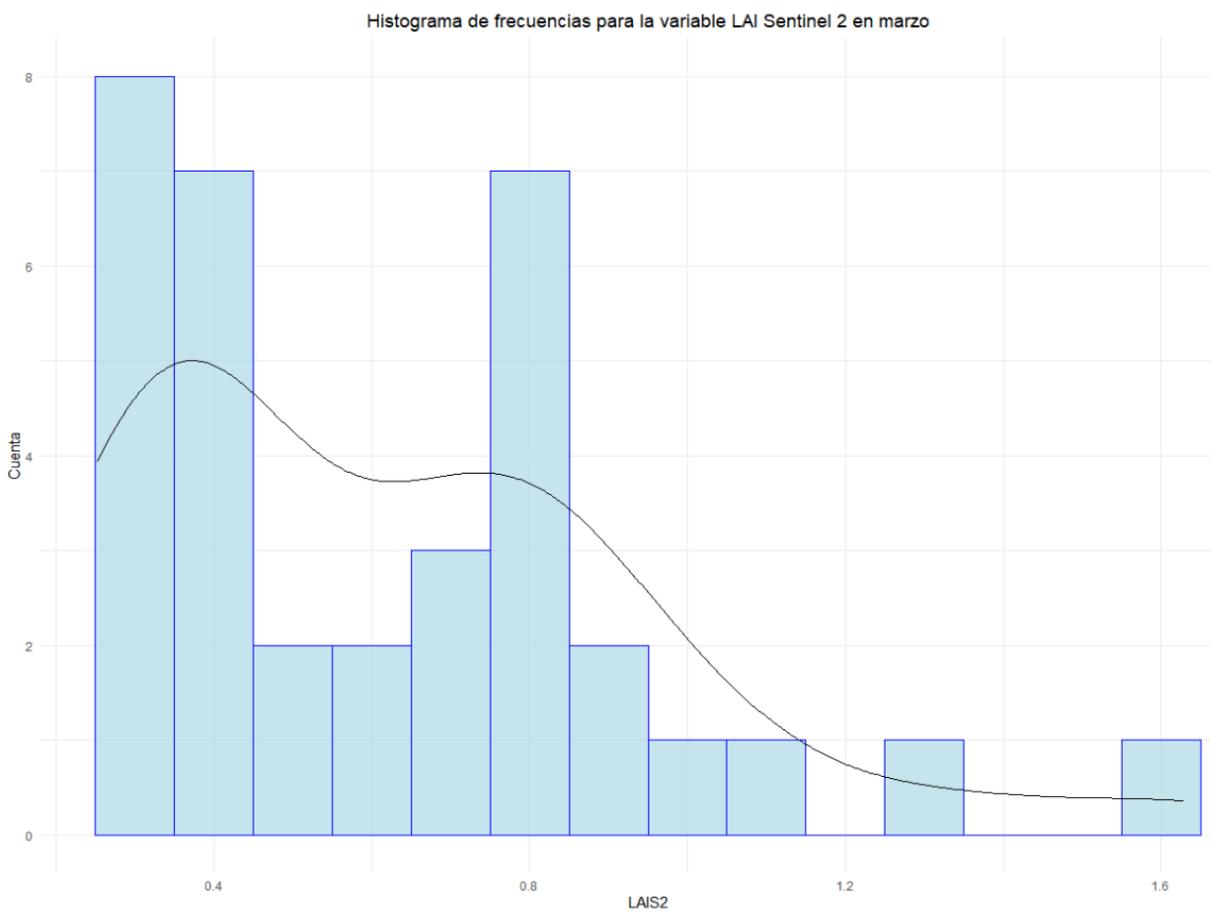


Figura 34. Histograma de la variable LAI Sentinel 2 de las muestras de marzo, p-value =0.0026

6.7. Descarte de variables independientes por correlación y colinealidad

Los umbrales establecidos en cuanto a los valores de las matrices de correlación y valores del Factor de inflación de la varianza nos permiten realizar un cribado sobre las variables independientes sobre las que se ejecutara el modelo de regresión.

Se describen para cada muestra de métricas LiDAR los resultados del cribado en cada una de sus fases, así como las variables seleccionadas en cada caso para la ejecución del modelo LASSO

6.7.1. Descarte de variables independientes de datos LiDAR Matrice de alta densidad

Análisis de variables independientes de datos LiDAR Matrice de alta densidad			
Muestra	Completa	Marzo	Mayo-junio
Descarte correlación de Spearman	73 variables	63 variables	77 variables
Descarte factor de inflación de la varianza	5 variables	7 variables	1 variable
Variabes de entrada en LASSO	10 variables	7 variables	6 variables

Tabla 4. Resumen de descarte de variables independientes

Variabes seleccionadas para la muestra completa: Elev maximum, Elev mode, Elev L3, Elev L4, Elev P01, Int kurtosis, Int L2, Int L3, Int L4, Int P05.

Variabes seleccionadas para la muestra de marzo: Elev maximum, Elev mode, Elev L3, Elev L4, Elev P01, Int kurtosis, Int L3.

Variabes seleccionadas para la muestra de mayo-junio: Elev L3, Elev L kurtosis, Elev P99, Int skewness, Int P95, Total all returns.

6.7.2. Descarte de variables independientes de datos LiDAR Matrice de baja densidad

Análisis de variables independientes de datos LiDAR Matrice de alta densidad			
Muestra	Completa	Marzo	Mayo-junio
Descarte correlación de Spearman	75 variables	75 variables	79 variables
Descarte factor de inflación de la varianza	4 variables	4 variables	2 variable
Variabes de entrada en LASSO	10 variables	10 variables	9 variables

Tabla 5. Resumen de descarte de variables independientes

VARIABLES SELECCIONADAS PARA LA MUESTRA COMPLETA: Total return count, Return 1 count above 0 and below 30, Elev minimum, Elev maximum, Elev L3, Elev L CV, Elev P01, Int mode, Int L3, Total first returns.

VARIABLES SELECCIONADAS PARA LA MUESTRA DE MARZO: Elev minimum, Elev maximum, Elev L3, Elev L4, Elev P01, Int kurtosis, Int L3, Percentage all returns above mode, Total first returns, Total all returns.

VARIABLES SELECCIONADAS PARA LA MUESTRA DE MAYO-JUNIO: Elev minimum, Elev L3, Elev L skewness, Int kurtosis, Int L skewness, Int P99, All returns above mode, Total first returns, Total all returns.

6.7.3. Descarte de variables independientes de datos LiDAR PNOA segunda cobertura

Para el caso de las variables dependientes LiDAR PNOA, el límite superior del factor de inflación de la varianza (VIF) se ha establecido en el límite recomendado, con un valor de 10, debido a que para el caso de los modelos de regresión para la variable dependiente LAI Sentinel 2 el descarte de variables independientes era demasiado agresivo, y el modelo de regresión LASSO descartaba todas las variables seleccionadas, causando error en la generación del modelo de regresión.

Análisis de variables independientes de datos LiDAR Matrice de alta densidad			
Muestra	Completa	Marzo	Mayo-junio
Descarte correlación de Spearman	78 variables	77 variables	81 variables
Descarte factor de inflación de la varianza	2 variables	7 variables	1 variable
VARIABLES DE ENTRADA EN LASSO	13 variables	9 variables	10 variables

Tabla 6. Resumen de descarte de variables independientes

VARIABLES SELECCIONADAS PARA LA MUESTRA COMPLETA: Elev minimum, Elev maximum, Elev mode, Elev kurtosis, ElevL3, Elev LCV, Elev P25, Int AAD, Int L3, Int L4, Int L kurtosis, Percentage all returns above mode, Total first returns.

VARIABLES SELECCIONADAS PARA LA MUESTRA DE MARZO: Elev minimum, Elev maximum, Elev mode, Elev L3, Elev P30, Int skewness, Int kurtosis, Int L4, Percentage first returns above mode.

VARIABLES SELECCIONADAS PARA LA MUESTRA DE MAYO-JUNIO: Return 1 count above 0 and below 30, Elev minimum, Elev mode, Elev CV, Elev MAD mode, Elev L3, Elev L4, Int kurtosis, Int AAD, Percentage all returns above mode.

6.8. Modelos de regresión LASSO y pruebas de normalidad y homocedasticidad de residuos

Las variables independientes seleccionadas a partir de las matrices de correlación y los valores de VIF son utilizadas, previa normalización de sus valores, para obtener 18 modelos de regresión, 6 modelos para cada muestra de métricas LiDAR, correspondientes a la estimación de valores de LAI utilizando como variable dependiente el valor LAI de datos de campo LI-COR LAI 2200c y los valores de LAI obtenidos a partir de datos de satélite y, en cada caso, para la muestra completa, la muestra correspondiente al mes de marzo y la muestra correspondiente a los meses de mayo-junio.

6.8.1. Modelos LASSO de datos LiDAR Matrice de alta densidad

ANÁLISIS DE DATOS LiDAR MATRICE DE ALTA DENSIDAD						
Variable dependiente	LAI LI-COR 2200c			LAI Sentinel 2		
Muestra	Completa	Marzo	Mayo-junio	Completa	Marzo	Mayo-junio
Variabes	6	4	4	7	4	6
Std error residuos	1.137	0.6062	1.048	0.4425	0.1554	0.2157
R ²	0.2339	0.3959	0.4564	0.3793	0.802	0.8576
R ² Ajustado	0.1673	0.3154	0.3961	0.3154	0.7756	0.8325
Normalidad residuos p-value	0.01089	0.5526	0.167	0.00298	0.01294	0.4281
Homocedasticidad p-value	0.03793	0.6928	0.4372	0.6726	0.6388	0.5242

Tabla 7. Resumen de resultados de los modelos de regresión de datos LiDAR Matrice de alta densidad

6.8.2. Modelos LASSO de datos LiDAR Matrice de baja resolución espacial

ANÁLISIS DE DATOS LiDAR MATRICE DE BAJA RESOLUCIÓN ESPACIAL						
Variable dependiente	LAI LI-COR 2200c			LAI Sentinel 2		
Muestra	Completa	Marzo	Mayo-junio	Completa	Marzo	Mayo-junio
Variabes	4	3	4	5	5	5
Std error residuos	1.105	0.6207	1.09	0.4395	0.1458	0.2449
R ²	0.2549	0.3455	0.4121	0.3697	0.8316	0.8111
R ² Ajustado	0.2129	0.2822	0.3468	0.3247	0.8026	0.7841
Normalidad residuos p-value	0.01415	0.6664	0.2746	0.01399	0.9738	0.8423
Homocedasticidad p-value	0.004649	0.498	0.2561	0.8389	0.2781	0.3952

Tabla 8. Resultados de los modelos de regresión de datos LiDAR Matrice de baja resolución espacial

6.8.3. Modelos LASSO de datos LiDAR PNOA de segunda cobertura

ANÁLISIS DE DATOS LiDAR PNOA DE SEGUNDA COBERTURA						
Variable dependiente	LAI LI-COR 2200c			LAI Sentinel 2		
Muestra	Completa	Marzo	Mayo-junio	Completa	Marzo	Mayo-junio
Variabes	6	3	3	10	3	9
Std error residuos	1.119	0.6173	1.027	0.4069	0.2282	0.2808
R ²	0.2581	0.3527	0.4776	0.4983	0.559	0.78
R ² Ajustado	0.1936	0.2901	0.4196	0.4212	0.5163	0.7162
Normalidad residuos p-value	0.003993	0.2024	0.7383	0.2699	0.3528	0.5352

Homocedasticidad p-value	0.473	0.7023	0.5835	0.1432	0.161	0.2262
--------------------------	-------	--------	--------	--------	-------	--------

Tabla 9. Resumen de resultados de los modelos de regresión de datos LiDAR PNOA de segunda cobertura

7. DISCUSIÓN

El análisis de fenómenos naturales a escala regional esta experimentando durante los últimos años un desarrollo importante de los datos accesibles en cuanto a la cantidad y resolución, tanto espacial como temporal. Numerosos estudios plantean objetivos cuyas hipótesis se asientan en la magnitud de los datos, o sensores y plataformas de última tecnología o reciente permeabilización a la comunidad científica.

El manejo de dichos datos a gran escala, si bien su creciente disponibilidad esta acompañada de nuevas herramientas y mejoradas capacidades de procesamiento informático, supone para el investigador un reto de constante actualización y descubrimiento de nuevas herramientas de software, cuando no lenguajes de programación que permiten la adaptación de las herramientas a los objetivos individuales del estudio.

Por otra parte, cabe preguntarse si no nos estamos dejando llevar por la magnitud de la información, cuando tal vez disponemos de datos mas ligeros de manejar, más accesibles o económicamente más rentables que proporcionen una mejor solución coste/beneficio. También si toda esa información permite descuidar una detallada planificación y selección de los datos, cuando unas condiciones de inestabilidad atmosférica persistentes o una excesiva extensión de temporal de las campañas de captura de datos pueden conllevar complicaciones inesperadas.

En España, los datos LiDAR han sido utilizados profusamente durante la última década en la descripción y caracterización de entornos forestales. El programa de vuelos LiDAR del Plan Nacional de Ortofotografía Aérea (PNOA) ha suministrado gratuitamente dos coberturas para todo el territorio nacional con una resolución espacial que ya se sitúa, en la mayoría de las comunidades autónomas, en el entorno de 1 punto por metro cuadrado.

Mientras tanto, la planificación de misiones LiDAR dedicadas ha llegado de la mano de los vehículos aéreos no tripulados y la miniaturización de sensores, si bien su capacidad para registrar rebotes está en cuestión, la resolución espacial alcanza densidades de puntos cuyo limite es el absurdo, véanse los 450 puntos por metro cuadrado del presente estudio, con un vuelo a 100 metros de altura respecto al terreno, cercano al limite legal de 120 metros. La realización de vuelos tripulados dedicados, con sensores que presentan altas capacidades, más allá de la resolución espacial, no ha terminado de despegar por sus altos costes operacionales.

El presente estudio ha pretendido analizar algunas de estas cuestiones, comparando el desempeño de distintos sensores, también satelitales y de campo, para el cálculo del índice de superficie foliar (LAI), para lo que se ha realizado un despliegue importante en cuanto a herramientas de software y diseño de programas que permitieran el tratamiento y gestión del volumen de datos LiDAR y de imágenes satelitales manejados, sobre una extensión que cubre la mitad de la Comunidad Autónoma de Aragón.

En cuanto a los resultados obtenidos respecto del primer objetivo, el estudio de la viabilidad de los distintos sensores LiDAR, se han obtenido buenas correlaciones para modelos de regresión respecto de los valores de LAI estimados a partir de imágenes Sentinel 2. Con independencia de la resolución espacial utilizada, coeficientes de R^2 ajustado en torno a 0.8 han sido alcanzados para los datos LiDAR obtenidos con el vehículo aéreo no tripulado, independientemente de que la resolución espacial fuera de 1 o 450 puntos por metro cuadrado. La resolución temporal si parece determinante, obteniendo los datos LiDAR del PNOA, con 3 años de antigüedad respecto a las mediciones de LAI, unas correlaciones que son ligeramente inferiores, 0.72, en el caso de las parcelas cuyo LAI fue calculado en los meses de mayo y junio.

Las correlaciones alcanzadas por los datos LiDAR PNOA son sin embargo discretas, 0.52 en el caso de las parcelas visitadas durante marzo, quedando la cuestión la viabilidad de estos datos para la

estimación de LAI, factores como el estadio vegetativo respecto a la captura de los datos LiDAR o la variabilidad en el desarrollo de distintas masas forestales a lo largo del periodo de carencia entre la captura de los datos LiDAR y la fecha de cálculo de LAI pudieran ser objeto de estudios posteriores.

Se ha observado también, respecto de las capacidades de los distintos sensores, como la información respecto del número de rebotes del haz LiDAR del sensor Zenmuse L1 se ha probado incapaz de aportar datos de utilidad a los modelos de regresión obtenidos, cuando el 28 % de las parcelas voladas presentaba un conteo de segundos retornos igual a 0, descartando el uso de la variable como variable independiente en un modelo de regresión por su nula variabilidad para una fracción importante de la muestra. En el caso del LiDAR PNOA, por el contrario, los datos de segundos y terceros retornos han superado los primeros filtros del análisis estadístico.

Es necesario comentar el comportamiento de los datos de LAI capturados con el sensor LI-COR, que obtienen bajos coeficientes de correlación en todos los casos. La metodología de captura de los datos o las condiciones del entorno en el momento de la captura deberán ser objeto de análisis. Las condiciones de las parcelas afectadas por el fuego pudieran no ser las óptimas, con coberturas vegetales de menor altura o más dispersas, que pudieran estar provocando mediciones con excesiva contribución de factores lumínicos. Por otra parte, y tal vez relacionado, si se observa una tendencia de mejora de las correlaciones en todos los modelos de los datos capturados durante mayo y junio respecto de los capturados en marzo, con diferencias en torno a 0,1 puntos de R^2 , tal vez relacionada con el estadio vegetativo de las masas forestales en el momento de la captura.

Respecto del estadio vegetativo se alcanza una de las conclusiones importantes del estudio. Se observa consistentemente como los modelos de regresión calculados con la muestra completa, que incluye los datos capturados en marzo y los capturados en mayo y junio, son incapaces de superar coeficientes de correlación superiores a 0.42. El rango temporal de captura de los datos, de casi 4 meses, esta provocando que los modelos de regresión no sean capaces de explicar los valores de LAI de Sentinel 2 a partir de datos LIDAR para toda la muestra, cuando la fecha de captura de los datos de LAI y los datos LIDAR UAV es similar. Parece que la variabilidad estacional de los datos de LAI para la vegetación a lo largo de la primavera no se corresponde con una variación análoga de los datos estructurales de vegetación representados por las métricas LIDAR.

8. AGRADECIMIENTOS

Me gustaría hacer mención en este apartado a la universidad pública, cuya existencia me ha permitido, después de algunos años, volver para quitarme una espina que tenía clavada, que era poder realizar este Máster para profundizar académicamente en el mundo de la teledetección, al que, si bien estaba conectado por mi actividad profesional, hacia tiempo que no me acercaba sin más intención que aprender.

Pero sobre todo debo agradecer a mi familia, mi hijo Bruno y mi mujer, a los que no he podido dedicar el tiempo que me habría gustado durante este complicado año, y una vez más a mis padres, Manolo y Mari Carmen, que han tenido que asumir una carga importante para que pudiera compatibilizar los deberes familiares, laborales, y el estudio.

9. BIBLIOGRAFÍA

- Adrián Jiménez Ruano, Fernando Pérez-Cabello y Raquel Montorio Llovería (2016). Niveles de LAI/fPAR en superficies afectadas por incendios forestales en Aragón. Análisis mediante el producto MCD15A2 DE MODIS. Pirineos, 171, e019. doi: <http://dx.doi.org/10.3989/Pirineos.2016.171003>
- Cabello, F. P., De, J., & Fernández, L. R. (1998). EL EMPLEO DE IMÁGENES LANDSAT TM PARA LA DETECCIÓN Y CARTOGRAFÍA DE ÁREAS INCENDIADAS EN EL PREPIRINEO OCCIDENTAL OSCENSE. In GEOGRAPHICALIA (Vol. 36).
- Cerdà, A., Imeson, A. C., & Calvo, A. (1995). Fire and aspect induced differences on the erodibility and hydrology of soils at La Costera, Valencia, southeast Spain. In Catena (Vol. 24).

- Cerdà, A. (1998). Changes in overland flow and infiltration after a rangeland fire in a Mediterranean scrubland. *Hydrological Processes*, 12(7), 1031–1042. [https://doi.org/10.1002/\(SICI\)1099-1085\(19980615\)12:7<1031::AID-HYP636>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1099-1085(19980615)12:7<1031::AID-HYP636>3.0.CO;2-V)
- Chen, J. M. & black, t. A., 1992. Defining leaf area index for non-flat leaves. *Plant, Cell & Environment*, 15 (4): 421-429.
- Giovannini, G. and Lucchesi, S. 1983. `Effect of fire on hydrophobic and cementing substances of soil aggregates',*Soil Sci.*,136,231±236
- Giovannini, G. and Lucchesi, S. 1984. `Differential thermal analysis and infrared investigations on soil hydrophobic substances.*Soil Sci.*,137(6), 457±463.
- Giovannini, G., Lucchesi, S., and Giachetti, M. 1988. `Effect of heating on some physical and chemical parameters related to soilaggregation and erodibility',*Soil Sci.*,143, 220±226.
- Marqués, M. A., & Mora, E. (1992). The Influence of Aspect on Runoff and Soil Loss in a Mediterranean Burnt Forest (Spain) (Vol. 19).
- Montorio Llovería, R., Pérez-Cabello, F., vLassova, L., & de La Riva Fernández, J. (n.d.). LA SEVERIDAD DEL FUEGO: REVISIÓN DE CONCEPTOS, MÉTODOS Y EFECTOS AMBIENTALES.
- Pérez-Cabello, F., Echeverría, M., De La Riva, J., & Ibarra, P. (2011). APUNTES SOBRE LOS EFECTOS DE LOS INCENDIOS FORESTALES Y RESTAURACIÓN AMBIENTAL DE ÁREAS QUEMADAS. ESTADO DE LA CUESTIÓN Y PRINCIPIOS GENERALES. In *GEOGRAPHICALIA*.(Cabello et al., 1998)
- Tibshirani, Robert. “Regression Shrinkage and Selection via the Lasso.” *Journal of the royal statistical society series b-methodological* 58 (1996): 267-288.
- Wang, Y.; Fang, H. Estimation of LAI with the LiDAR Technology: A Review. *Remote Sens.* 2020, 12, 3457. <https://doi.org/10.3390/rs12203457>

10. ANEXO I: SCRIPT EN R UNION DE TABLAS EXCEL

```
--
36 - #####
37 - ##### JOIN DE TABLAS EXCEL #####
38 install.packages("openxlsx")
39
40 library(openxlsx)
41 library(dplyr)
42 library(readxl)
43
44 #Cargar tablas
45 #Excel1 <- read_excel("D://Documentos/01_MASTER/01_TFM/LAI_campo/Campo_LAI.xlsx", sheet = 'BASE_TODAS_LUZ')
46 Excel2 <- read_excel("F://TFM/PNOA_2Cober/cloudmetrics/CloudMetrics_above1m_TFM_PNOA_2COBER_excel.xlsx")
47 Excel1 <- read_excel("F://TFM/LAI_DATA/LAI_LICOR_Sentinel2.xlsx")
48
49 Excel1 <- Excel1[!is.na(Excel1$LAI_L),]
50
51 # Rango de columnas a convertir (de la 3 a la 103)
52 columnas_a_convertir <- 2:102
53
54 # Convertir las columnas en números
55 Excel2[, columnas_a_convertir] <- lapply(Excel2[, columnas_a_convertir], as.numeric)
56
57 #Unir tablas
58 Tabla <- Excel1 %>%
59   left_join(Excel2, by = c("LAI_L" = "FileTitle" ))
60
61 datos <- Tabla[!is.na(Tabla$LAI_L),]
62
63 # Guardar tabla resultante
64 ruta_salida <- "F://TFM/PNOA_2Cober/Cloudmetrics/resultados/"
65 write.xlsx(datos, file = file.path(ruta_salida, "LICOR_SENTINEL2_PNOA_2COBER_OK.xlsx"))
66
67
68
69 - #####
```

11. ANEXO II: SCRIPT EN R DE RECORTE DE NUBES DE PUNTOS CON PARCELAS

```
129 < #####
130 < ### RECORTAR NUBES DE PUNTOS NORMALIZADAS CON SHAPE POL PARCELA Y EXPORTAR INDIVIDUALMENTE###
131
132 < # Cargar paquetes necesarios
133
134 < library(lidr)
135 < library(rgdal)
136 < library(sf)
137 < library(raster)
138
139 < # Ruta al shapefile con los elementos para recortar
140 < ruta_shapefile <- "F://MATRICE/Normalizados/SHP/LAI_Muestras_buff20m.shp"
141
142 < # Cargamos el shapefile
143 < parcelas <- st_read(dsn = ruta_shapefile)
144
145 < #CRS LAS
146 < ruta_archivoLAS <- "D:/02_MASTER/01_TFM/Datos_base/LIDAR/LIDAR_PARCELAS_CLIP/LIDAR_2COBER_PARCELAS_LAI.las"
147 < archivoLAS <- readLAS(ruta_archivoLAS)
148 < crs <- st_crs(archivoLAS)
149
150 < # Buffer de seguridad a capa vectorial
151 < parcelas <- st_as_sf(parcelas)
152 < parcelas <- st_transform(parcelas,25830)
153
154 < # Función para recortar y exportar las nubes de puntos
155 < recortar_y_exportar_laz <- function(archivoLAS, shapefile) {
156 <   nube_puntos <- readLAS(archivoLAS)
157 <   cat("Procesando archivo LAS:", archivoLAS, "\n")
158
159 <   for (j in 1:nrow(shapefile)) {
160 <     LIDAR_recortados <- clip_roi(nube_puntos, shapefile[j,])
161 <     estadillo <- shapefile$inventario[j]
162 <     cat(estadillo)
163
164 <     nombre_archivo <- paste0("D:/02_MASTER/01_TFM/Datos_base/LIDAR/LIDAR_PARCELAS_CLIP/Parcelas/",
165 <                               estadillo, ".las")
166
167 <     writeLAS(LIDAR_recortados, nombre_archivo)
168 <   }
169 < }
170 < recortar_y_exportar_laz(archivoLAS, parcelas)
171
172 < #####
```

12. ANEXO 3: SCRIPT EN PYTHON DE NORMALIZACION DE NUBES DE PUNTOS

```
1 @'''
2 Created on 11 oct 2023
3
4 @author: Jorge García Hernández
5 '''
6 import os
7 import pylas
8 import rasterio
9 import numpy as np
10
11 # Ruta al directorio que contiene los archivos LAS de entrada
12 ruta_directorio_las = "D:/02_MASTER/01_TFM/Metashape/DEM/MATRICE/SO_ESTE"
13 # Ruta al directorio donde se guardarán los archivos LAS normalizados
14 ruta_salida_las = "D:/02_MASTER/01_TFM/Metashape/DEM/MATRICE/Normalizados"
15 os.makedirs(ruta_salida_las, exist_ok=True)
16 # Lista de archivos LAS en el directorio
17 lista_archivos_las = [archivo for archivo in os.listdir(ruta_directorio_las) if archivo.lower().endswith('.las')]
18
19 # Función para normalizar un archivo LAS
20 def normalizar_las(archivo_las, dataset, output_folder):
21     print(f"Procesando archivo: {archivo_las}")
22
23     # Cargar el archivo LAS
24     las = pylas.read(os.path.join(ruta_directorio_las, archivo_las))
25
26     # Obtener coordenadas XY de los puntos LAS
27     xy = np.vstack([las.x, las.y]).T
28
29     print(f"Cantidad de coordenadas XY: {len(xy)}")
30
31     # Obtener elevaciones del MDE en las coordenadas XY utilizando la función sample
32     elevations_mde_generator = dataset.sample(xy)
33     elevations_mde = [elevation[0] for elevation in elevations_mde_generator]
34
35     print(f"Cantidad de elevaciones MDE: {len(elevations_mde)}")
36
37     # Calcular la diferencia de altura entre el punto LAS y el MDE
38     difference = las.z - elevations_mde
39
40     # Actualizar las alturas de los puntos LAS normalizados
41     las.z = difference
42
43     # Guardar el archivo LAS normalizado
44     ruta_salida = os.path.join(output_folder, archivo_las.replace('.las', '_Normalizado.las'))
45     las.write(ruta_salida)
46     print(f"Archivo normalizado guardado: {ruta_salida}")
47
48 # Ruta al archivo GeoTIFF del Modelo Digital de Elevaciones (MDE)
49 ruta_mde_geotiff = "D:/02_MASTER/01_TFM/Metashape/DEM/MATRICE/SO_ESTE.tif"
50
51 # Leer el archivo GeoTIFF del Modelo Digital de Elevaciones (MDE)
52 with rasterio.open(ruta_mde_geotiff) as dataset:
53     # Normalizar y guardar cada archivo LAS
54     for archivo_las in lista_archivos_las:
55         normalizar_las(archivo_las, dataset, ruta_salida_las)
56         print("Proceso completado para:", archivo_las)
57
```