

TRABAJO FIN DE MÁSTER

CARTOGRAFÍA DE CORTAS A HECHO MEDIANTE TRATAMIENTO DIGITAL DE IMÁGENES SENTINEL 2 Y DATOS LIDAR PNOA

Alberto Marco Asenjo

Director: Juan de la Riva

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

septiembre de 2016



Universidad
Zaragoza

**Departamento de Geografía
y Ordenación del Territorio**



*A mi familia y amigos.
A todo el equipo de **föra forest technologies**.
Todos habéis contribuido en uno u otro
momento a que esta nave llegue a puerto.
De manera muy especial a Santi, que me ha
guiado en las distintas fases de este proyecto,
haciendo gala de gran capacidad didáctica, así
como de buenas dosis de paciencia; y a Paco,
por las muchas facilidades que me ha dado y
por la confianza que me
mostró desde el primer día.*

Resumen

El presente documento pretende contribuir a solucionar una carencia asociada a la gestión silvícola en España, en concreto, pretende desarrollar un proceso metodológico que permita localizar y cartografiar cortas a hecho en una superficie boscosa mediante teledetección. Para ello se empleará información geográfica de libre acceso. En particular, se utilizará una imagen Sentinel 2A en conjunto con datos LiDAR del PNOA. En una fase posterior se realizará una segmentación del área de estudio a partir de un software de clasificación orientada a objetos para, finalmente, clasificar empleando el algoritmo Random Forest.

Palabras Clave: Cortas a hecho, Sentinel 2, LiDAR, segmentación, clasificación.

Abstract

The current document attempts to solve a need associated to forest management in Spain, in particular, it attempts to develop a methodologic process for locating and mapping clear-cuts in forest areas using open source remote sensing data. More specifically, it uses a Sentinel 2 image combined with PNOA LiDAR data. In a subsequent stage, object based image analyst software is used to carry out a segmentation. Finally, the algorithm Random Forest is used to develop a classification model.

Key Words: Clear-cuts, Sentinel 2, LiDAR, segmentation, classification.

Índice general

1. INTRODUCCIÓN.....	1
2. MARCO CONTEXTUAL.....	3
2.1. Objetivos.....	3
2.2. Antecedentes.....	4
2.3. Área de estudio.....	5
3. PROCESO METODOLÓGICO.....	8
3.1. Procesado de datos.....	9
3.1.1. Procesado de datos LiDAR.....	9
3.1.2. Procesado de las Ortofotos.....	13
3.1.3. Procesado de datos Sentinel 2A.....	13
3.2. Segmentación del área de estudio.....	17
3.3. Toma de datos <i>verdad-terreno</i>	25
3.4. Clasificación mediante Random Forest.....	27
4. RESULTADOS.....	32
4.1. Análisis del resultado final.....	32
4.2. Discusión de la utilidad del modelo.....	34
4.3. Vías descartadas.....	35
5. CONCLUSIONES.....	36
6. BIBLIOGRAFÍA.....	38

Índice de figuras

Figura 1. Secuencia de cortas por fajas sobre Eucaliptus globulus. A la derecha con P95 LiDAR PNOA.	3
Figura 2. Mapa con la ubicación de las principales masas de coníferas en la provincia de Soria.	6
Figura 3. Mapa del área de estudio con núcleos de población cercanos.	7
Figura 4. Diagrama de flujos del proceso metodológico.	8
Figura 5. Segmentación generada por el algoritmo “Multiresolution segmentation” de eCognition; a la derecha, ampliación sobre el sector inferior izquierdo.	19
Figura 6. Efecto del algoritmo “Merge region” de eCognition: a la izquierda, fusión de la clase “Bosque”; a la derecha, fusión de la clase “No Bosque”	21
Figura 7. Segmentación generada por el algoritmo “Multiresolution segmentation” de eCognition sobre el primer nivel la derecha, ampliación sobre un sector.	22
Figura 8. Vista de la capa vectorial extraída de eCognition. A la derecha, detalle de los polígonos residuales.	23
Figura 9. Vista de puntos incorrectamente clasificados por efecto de los datos LiDAR sobre el terreno abrupto.	25
Figura 10. Gráfico de importancia de las variables.	29
Figura 11. Cartografía de cortas a hecho.	32

Índice de tablas

Tabla 1. Bandas originales de Sentinel 2A seleccionadas para el multibanda.	14
Tabla 2. Bandas Sentinel 2 integrantes de la imagen multibanda con nivel digital mínimo e índice de ND mínimo.	15
Tabla 3. Capas integradas en el proceso de segmentación de eCognition con indicación de su resolución espacial.	18
Tabla 4. Parámetros del algoritmo “Multiresolution segmentation” de eCognition.	19
Tabla 5. Parámetros del algoritmo “Assign Class” de eCognition.	20
Tabla 6. Parámetros del algoritmo “Merge region” de eCognition.	20
Tabla 7. Parámetros del algoritmo “Multiresolution segmentation” de eCognition.	21
Tabla 8. Parámetros del algoritmo “Export vector layer” de eCognition.	22
Tabla 9. Número de registros verdad-terreno por cobertura.	25
Tabla 10. Diagrama cruzado entre observaciones y predicciones.	30
Tabla 11. Variables explicativas del modelo con su índice MIR ordenadas de menor a mayor relevancia.	31
Tabla 12. Matriz de confusión a nivel de pixel.	33
Tabla 13. Medidas simétricas obtenidas de IBM SPSS 22.	33
Tabla 14. Matriz de confusión a nivel de segmentos.	34
Tabla 15. Medidas simétricas obtenidas de IBM SPSS 22.	34

Índice de cuadros

Cuadro 1. Instalación e invocación de bibliotecas para Random Forest.....	27
Cuadro 2. Definición de directorio principal, creación de objeto con los datos y visualización de nombres de las columnas.	27
Cuadro 3. Lista completa de las variables.	27
Cuadro 4. Definición del campo “CLASE” como factor y resumen de este.....	28
Cuadro 5. Resumen del campo “CLASE”.....	28
Cuadro 6. Fórmula de Random Forest.	28
Cuadro 7. Sentencias para clasificar las variables explicativas según su importancia.....	29
Cuadro 8. Variables explicativas en fase inicial del proceso de ajuste-descarte.....	29
Cuadro 9. Creación del campo "pred_rf_Segmentacion" y creación de la tabla con las predicciones.	30
Cuadro 10. Generación del diagrama cruzado	30
Cuadro 11. Creación de un objeto de lectura a partir de la tabla de la segmentación.....	31
Cuadro 12. Creación del campo rf_Segmentacion y resumen.	31
Cuadro 13. Generación del resultado final fusionando "tabla_segmentacion" con "tabla_segmentacion_prob"	31

El presente Trabajo de Fin de Máster, en adelante TFM, constituye, en el marco académico de Bolonia, por sí mismo, una asignatura de carácter práctico, aplicado y transversal, en la que se utilizarán las competencias adquiridas a lo largo del “Máster Universitario en Tecnologías de la Información Geográfica para la Ordenación del Territorio: SIG y Teledetección” para satisfacer el objetivo principal de este proyecto. Se incluye en la modalidad C, que agrupa “Trabajos específicos realizados como resultado de prácticas en empresas o instituciones”. La empresa en la que se desarrolló el proceso es *föra forest technologies* (<http://fora.es/es/>), consultoría forestal especializada en tecnologías de la información geográfica, ubicada en la capital soriana, *spin-off* de la Universidad de Valladolid, en funcionamiento desde 2014.

1. INTRODUCCIÓN

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

Las diferentes resoluciones -temporal, radiométrica, espacial y espectral-están condicionadas tanto por el sensor que registra la información, como por la plataforma que lo porta. A este respecto, encontramos plataformas terrestres, aerotransportadas y satelitales. Tradicionalmente se asumía que, cuanto más próxima estuviera una plataforma del suelo, mayor era su resolución espacial y, posiblemente, también su resolución radiométrica y espectral, si bien es cierto que desde hace algún tiempo se dispone de sensores satelitales con resoluciones radiométrica y espectral considerables. Hasta el punto de que, hoy en día, los avances técnicos en cuanto a la resolución espacial han dado como resultado sensores satelitales con resoluciones espaciales anteriormente propias de plataformas aerotransportadas. En consecuencia, esta evolución en la técnica facilita una evolución metodológica en los distintos campos de aplicación de la teledetección.

Uno de esos campos de aplicación es el forestal, en el que, tradicionalmente, se han venido empleando imágenes de satélites multiespectrales de resolución espacial media, tales como Landsat y SPOT, debido a la aceptable relación que ofrece su disponibilidad, periodicidad, resolución espacial y espectral (Wulder y Franklin, 2003). La evolución tecnológica experimentada por algunos sensores abre un nuevo abanico de posibilidades al respecto.

En este sentido, resulta especialmente interesante la irrupción en el panorama de la observación remota del satélite Sentinel 2A, puesto en órbita en junio de 2015 en el marco de la constelación Copernicus, a la espera de que su hermano gemelo Sentinel 2B sea puesto en órbita a lo largo de 2017, y que alberga un sensor multiespectral denominado MSI que recoge y codifica información en 12 bandas del espectro visible e infrarrojo. Sin embargo, no es esta mejora en la resolución espectral en relación a Landsat y SPOT lo que lo hace tan interesante de cara a propiciar un avance metodológico. Su verdadera importancia radica en i) su gran resolución temporal, con un periodo de revisita que alcanzará los 5 días, gracias a la acción coordinada de los satélites gemelos, frente a los 16 de Landsat o los 26 de SPOT; ii) la accesibilidad a las imágenes, disponibles a toda la población de manera gratuita con solo registrarse en el servidor de la Agencia Espacial Europea (ESA); y iii) en la gran resolución espacial que ofrece en tres de sus bandas del visible y en una de las del infrarrojo próximo, que logra un pixel de 10x10 m, frente a los 30x30m de Landsat (USGS, 2015; ESA, 2015).

Las dos últimas plataformas SPOT -SPOT 6 y SPOT 7- sin embargo, ofrecen una mayor resolución espacial -6x6 m en el visible y NIR y 1,5x1,5m en el PAN-pero,por el contrario,sus productos resultan menos accesibles, al no ofrecer sus imágenes de manera gratuita para el gran público.Esta es la gran desventaja de SPOT a la hora de considerarlo una alternativa viable para el monitoreo de superficies boscosas, no tanto la resolución temporal, ya que es capaz de reducir su periodo de revisita inicial ajustando el ángulo de observación y combinando los satélites SPOT 6 y 7 (Astrium, 2012). Por su parte, la principal desventaja de Landsat radica en su resolución espacial, primero, y en su resolución temporal, segundo, ya que la resolución espectral y la distribución de las bandas de Landsat 8 son similares a las de Sentinel 2. Por el contrario, tanto Landsat como SPOT permiten la realización de estudios multitemporales, por ser misiones de largo recorrido -Landsat desde 1972 y SPOT desde 1986- mientras que Sentinel, actualmente, no puede ofrecer esa posibilidad. Sin embargo, en vista de la relativa similitud entre las bandas Landsat y Sentinel, es posible emplear información Landsat a este efecto. Por todo esto, Sentinel 2 se posiciona como una alternativa de gran funcionalidad para tomar el relevo de sus competidores francés y americano como herramienta fundamental en la gestión forestal.

El éxito de la aplicación de la teledetección en la gestión forestal se debe a la amplia cobertura del territorio que proporcionan las imágenes de satélite, lo que permite trabajar con extensas zonas de manera simultánea y no invasiva, permitiendo recabar información de zonas remotas o de acceso difícil (Álvarez-Bermúdez, 2015), por lo que se ajusta a la idiosincrasia de las masas boscosas. Además, gracias a la información multiespectral es posible obtener índices para evaluar aspectos relativos al estado de los bosques tales como la superficie foliar, el vigor vegetal, la severidad de un incendio, etc. De acuerdo con Koch (2011) es posible estimar el volumen de madera y la biomasa superficial a partir de la reflectividad. Como afirman García-Martín *et al.* (2009), la estimación de variables biofísicas es una línea exitosa de investigación en teledetección forestal. Otra de las posibilidades que ofrece la teledetección forestal es la realización de estudios multitemporales de cambios en la cobertura boscosa, que permiten conocer la evolución de una masa forestal a lo largo del tiempo.

Combinando la información multiespectral de los sensores ópticos con la información tridimensional de una nube de puntos LiDAR es posible clasificar las masas forestales de acuerdo a las especies que la componen, estructura -vertical u horizontal-o grado de desarrollo. Puede servir para determinar la altura de los árboles, la altura a la que comienza la copa, para discretizar los distintos estratos, para estimar el índice de área foliar o la fracción de cavidad cubierta (Riaño *et al.*, 2004). La tecnología RADAR también es útil en este tipo de estudios, si bien se centra en la rugosidad de la superficie (Chuvieco, 2008), aportando además la capacidad de atravesar coberturas nubosas, algo de gran utilidad ante bosques tropicales frecuentemente cubiertos por nubes (Koch, 2011).En definitiva, se trata de una herramienta eficaz que permite trabajar con grandes superficies de manera rápida y reduciendo considerablemente los costes económicos.

En cualquier caso, la teledetección no debe sustituir por completo el registro de datos fiables sobre el terreno (Koch, 2011), sino complementarse mutuamente a fin de combinar la rapidez, el abaratamiento de costes y la capacidad para trabajar a escalas considerables de la teledetección con la precisión y calidad de los datos verdad-terreno. Actualmente, se están realizando mediante teledetección trabajos de inventario forestal y de pastizales (Chuvieco, 2008), análisis de la severidad de incendios y estudios la regeneración posterior del ecosistema; desarrollo de cartografía de biomasa enfocados tanto a su extracción y aprovechamiento energético (García-Martín *et al.*, 2011) como hacia la predicción del nivel de severidad del fuego en incendios forestales (García-Martín *et al.*, 2009), entre otras aplicaciones.

Con este TFM se pretende contribuir a solucionar una carencia asociada a la gestión silvícola en España; en concreto, se pretende desarrollar un proceso metodológico que permita localizar y cartografiar cortas a hecho en una superficie boscosa, dado que en determinadas comunidades con bosques fragmentados de titularidad privada en los que se realiza explotación forestal, no se ha llevado históricamente un registro de este tipo de actuaciones y, por tanto, carecen de la trazabilidad deseable de cara a su correcto manejo y gestión, tanto por parte del titular del monte, como por parte de la Administración pública en determinados casos.

De acuerdo con la definición facilitada por el *Diccionario Forestal* de la Sociedad Española de Ciencias Forestales, una corta a hecho es un tipo de corta de regeneración que consiste en la extracción total y en una vez de todos los pies que forman el rodal. (SECF, 2005)

Son, por tanto, cortas continuas que dan lugar a masas regulares y que suponen un cambio brusco en las condiciones ambientales de la zona. Las alternativas de regeneración posteriores son varias, según especie y estación, y sirven para establecer tipos de corta a hecho. De este modo se puede hablar de cortas a hecho en un tiempo, cuando se extraen todos los pies de un rodal en la misma actuación, o en dos tiempos, cuando se deja una reserva de árboles padres sin talar para facilitar la regeneración, que se cortarán en una actuación posterior. Este tipo de corta contradice la definición que ofrece el *Diccionario Forestal* y, sin embargo, es un concepto asentado en el sector.

En la actualidad, el modelo de corta a hecho más usual es la corta a hecho en fajas, que supone una suerte de modelo mixto ya que consiste en la extracción de una vez de todos los pies de un tranzón -faja-, enmarcado en una actuación que comprende varias fajas alternándose con rodales que se dejan en pie para favorecer el regenerado posterior y que se talarán en un turno posterior. Estas fajas han de tener una anchura inferior al doble de la altura media de los pies a talar y han de realizarse en perpendicular al viento dominante para favorecer la regeneración gracias a la acción del viento al esparcir las semillas.

En la provincia de Soria, desde hace décadas, la corta a hecho por fajas es la técnica que se realiza en la mayoría de los casos. Este factor puede condicionar tanto el proceso como el resultado de este proyecto, puesto que la anchura de estas fajas llevará al límite la resolución espacial del satélite Sentinel 2.

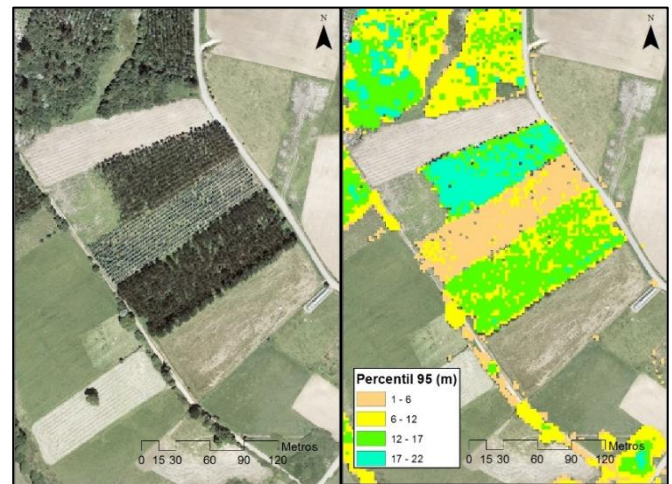


Figura 1. Secuencia de cortas por fajas sobre *Eucalyptus globulus*. A la derecha con P95 LiDAR PNOA.

2. MARCO CONTEXTUAL

2.1. Objetivos

El objetivo general de este trabajo es proponer una metodología precisa y eficaz para detectar y cartografiar áreas cuya explotación responda al modelo silvícola de la corta a hecho en pinares de silvestre (*Pinus sylvestris* L.), resinero (*Pinus pinaster* Ait.) y mixtos mediante teledetección, en concreto una imagen óptica de Sentinel 2 e información adicional aportada por un conjunto de datos LiDAR-PNOA. Esta metodología debe ser extensible a otras masas forestales con otras especies arbóreas.

Como objetivos específicos, en relación con su naturaleza de TFM realizado en conexión con una estancia de prácticas en empresa, se plantean los siguientes: i) abordar un recorrido integrado por las distintas fases a realizar en un proyecto técnico de teledetección (adquisición, corrección, composición, transformación, interpretación, clasificación...); ii) poner en práctica diversas habilidades adquiridas a lo largo del Máster, tales como la edición cartográfica, que habrá de servir de soporte gráfico a lo largo del desarrollo del proyecto, el empleo del lenguaje de programación R, el manejo de datos LiDAR, la segmentación de capas ráster empleando software de clasificación orientada a objetos o la generación de neocanales; iii) acercarse a la realidad operativa en un entorno profesional, mediante la gestión de volúmenes considerables de datos, lo que conlleva un esfuerzo importante en su procesado, y, en definitiva, familiarizarse con el entorno forestal y los conceptos y métodos que lo relacionan con el mundo de la teledetección y los sistemas de información geográfica.

2.2. Antecedentes

El mundo forestal lleva ya algunas décadas sirviéndose de la teledetección para optimizar su gestión. Primero a partir de la fotointerpretación y el análisis visual de las ortofotos, después, con el surgimiento de las imágenes de satélite, mediante su análisis visual primero y mediante su tratamiento digital posteriormente. Dentro del tratamiento digital encontramos técnicas como la generación de neocanales para sintetizar la información, el empleo de índices específicos para aislar una determinada variable biofísica, como puede ser el vigor vegetal o la severidad de un incendio, o la clasificación de imágenes mediante procesos no supervisados o supervisados; siendo estos últimos aquellos procesos en los que el usuario determine las unidades de referencia de cara al entrenamiento. Estas unidades reciben generalmente el nombre de área de interés -AOI, por sus siglas en inglés- o región de interés -ROI- en función del software a partir del cual sean procesadas.

La clasificación de imágenes pretende transformar la información cuantitativa procedente de las imágenes captadas por los sensores, compuestas por píxeles con un valor asignado en niveles digitales en relación a la resolución radiométrica del sensor; en imágenes temáticas compuestas por clases categóricas (González-Ferreiro *et al.*, 2013).

Dentro de la clasificación de imágenes, encontramos procesos que se fundamentan exclusivamente en las propiedades espectrales de los píxeles y otros que, además, también consideran sus propiedades geométricas y textuales, así como las de su entorno. Estos últimos son los denominados clasificadores orientados a objetos -OBIA, Object Based Image Analysis- que segmentan la imagen en regiones separadas, denominadas objetos, de acuerdo con su heterogeneidad espacial y espectral, para posteriormente asignarlos a una serie de clases previamente definidas. Este tipo de clasificación ofrece mejores resultados que la clasificación convencional orientada a las propiedades espectrales de los píxeles cuando se dispone de imágenes de alta resolución espacial pero menor resolución espectral (Huang y Ni, 2008).

Desde hace algún tiempo, varios autores han tratado de detectar y cartografiar cortas a hecho a partir de técnicas de observación remota procesadas en sistemas de información geográfica -SIG o GIS-, si bien han empleado una metodología diferente a la que aquí se propone. También ha habido otros autores que han trabajado en temas similares cuyos resultados pueden ser de utilidad para la realización de este trabajo.

Malila (1980) realizó un estudio enfocado en detectar cambios en los bosques mediante imágenes Landsat a través de un análisis de vectores de cambio; esta técnica, en palabras de Ruiz *et al.* (2007), “consiste en el cálculo del módulo y la dirección del vector que une los valores de los píxeles en el espacio definido por dos bandas espectrales, en dos fechas diferentes”. Empleando este método, Malila (1980) logró detectar y cartografiar cortas a hecho con resultados razonablemente positivos.

Skole y Tucker (1993) desarrollaron una técnica para cartografiar el avance de la deforestación en la Amazonia brasileña empleando el infrarrojo próximo de imágenes Landsat -MSS y TM-. Gracias a esta técnica lograron cuantificar el número de hectáreas afectadas en su área de estudio, así como identificar los patrones de explotación y, finalmente, demostrar que las anteriores estimaciones realizadas a través de imágenes obtenidas por satélites meteorológicos de baja resolución espacial sobreestimaban la deforestación en torno a un 50%.

Cohen *et al.* (1998) realizaron un estudio comparativo entre dos formas de cartografiar cortas a hecho empleando imágenes Landsat. Por un lado, realizaron una clasificación no supervisada a cada una de las cinco imágenes correspondientes a una misma zona de estudio en una secuencia temporal para después unificarlas en un mapa de cambios. El otro método que compararon seguía un proceso inverso, realizando primero la cartografía de evolución temporal en la masa boscosa para posteriormente realizar una única clasificación no supervisada. En ambos casos habían realizado un proceso previo de síntesis de la información mediante la generación de los ejes de *Tasseled Cap*. De este modo concentraron la información espectral de la que disponían en tres nuevas dimensiones con sentido físico explícito: brillo, verdor y humedad; si bien, en el caso de las imágenes captadas mediante el sensor

MSS, generaron solamente el brillo y el verdor, por carecer de información espectral propia del infrarrojo medio de onda corta. Tras la realización del estudio concluyeron que el segundo de los métodos resultaba más eficiente para cartografiar cortas a hecho.

Chuvienco *et al.* (2001) compararon la capacidad de Landsat, NOAA y SPOT Vegetation para estimar el contenido hídrico de la vegetación; para ello emplearon índices de vegetación (NDVI y NDII) e información de la región térmica del espectro con la que realizar el cociente entre el NDVI y la temperatura en superficie.

Álvarez (2005) inventarió la superficie ocupada por *Pinus radiata* D. Don en el Bierzo a partir de imágenes Landsat, a las que aplicó una clasificación orientada a objetos obteniendo resultados que presentaban cierto grado de confusión entre formaciones vegetales.

Gonçalves *et al.* (2008) demostraron la mayor eficacia de la clasificación orientada a objetos frente a la más convencional orientada a píxeles mediante los resultados obtenidos a partir de métodos diversos.

Gozález-Ferreiro *et al.* (2013) realizaron un estudio similar al que aquí se propone titulado “Clasificación digital de zonas forestales usando datos LiDAR, ortofoto y análisis orientado a objetos”, si bien no pretendían clasificar cortas a hecho ni tampoco explorar-como sí se plantea en el presente trabajo- las posibilidades que ofrece Sentinel 2 a tal efecto. El resultado de su estudio fue altamente satisfactorio logrando una fiabilidad global del 95%.

Lambert *et al.* (2015) desarrollaron un proyecto destinado a detectar cortas a hecho, pero, a diferencia de nuestro trabajo, lo hicieron mediante un estudio multitemporal con datos de un sensor de resolución media-baja -Modis- empleando el NDVI para detectar cambios en la vegetación sobre el terreno.

2.3. Área de estudio

El presente estudio se enmarca en el ámbito forestal de la provincia de Soria, la cual es sin duda una de las referencias en la materia a nivel nacional. Buena prueba de ello es que cuenta en su territorio con bosques con Certificación de Gestión Forestal Sostenible de acuerdo a los estándares del PEFC (Pan-European Forest Certification), que garantizan que el producto que llega hasta el consumidor lo hace desde bosques cuyos recursos son explotados de manera responsable y sostenible de acuerdo a las directrices contempladas en el correspondiente Plan de Ordenación (Pinillos y Moro, 2004).

La ordenación forestal en la provincia de Soria viene de lejos. En la Edad Media existían diversas figuras jurídicas relativas a la titularidad de los bosques en el territorio actual de Soria. Montes *propios*, *comunales* o *realengos*, obedecían cada uno a un sistema de gestión y propiedad específico. Durante siglos, los bosques sorianos fueron aprovechados por los vecinos de forma irregular, dando pie a prácticas como el huroneo, que debilitaba la estructura de la masa forestal; o a quemas por parte de pastores que buscaban ampliar la superficie de pasto aprovechable por los rebaños. (Santolaya, s.f)

Este tipo de actuaciones ocasionaron que entre finales del S.XIX y principios del XX la salud y estado de conservación de los bosques sorianos estuviera lejos de lo deseable. La elaboración en 1859 del primer inventario de patrimonio forestal de titularidad pública, la aprobación de la Ley de Montes de 1863 y la creación de una Administración Forestal de concepción moderna supusieron un punto de inflexión en la gestión de estos montes.

En el monte Pinar Grande, una masa boscosa compuesta de *Pinus sylvestris* y *Pinus pinaster* a caballo entre las provincias de Soria y Burgos, la Ordenación data de 1907. Esta figura compatibilizó los usos del monte y reguló la extracción de madera organizándola en el tiempo, considerando la capacidad de regeneración del bosque a la hora de establecer las secciones, cuarteles y tramos de corta. En un primer momento la extracción se realizaba mediante entresacas por bosquetes, si bien a partir de la

4º revisión del Plan, en 1954, comienza a implantarse la corta a hecho como tratamiento silvícola principal fomentando la regeneración con siembras y plantaciones en caso de ser necesario. Este método se propone de forma explícita en la 5ª revisión, diez años después. (Santolaya, s.f)

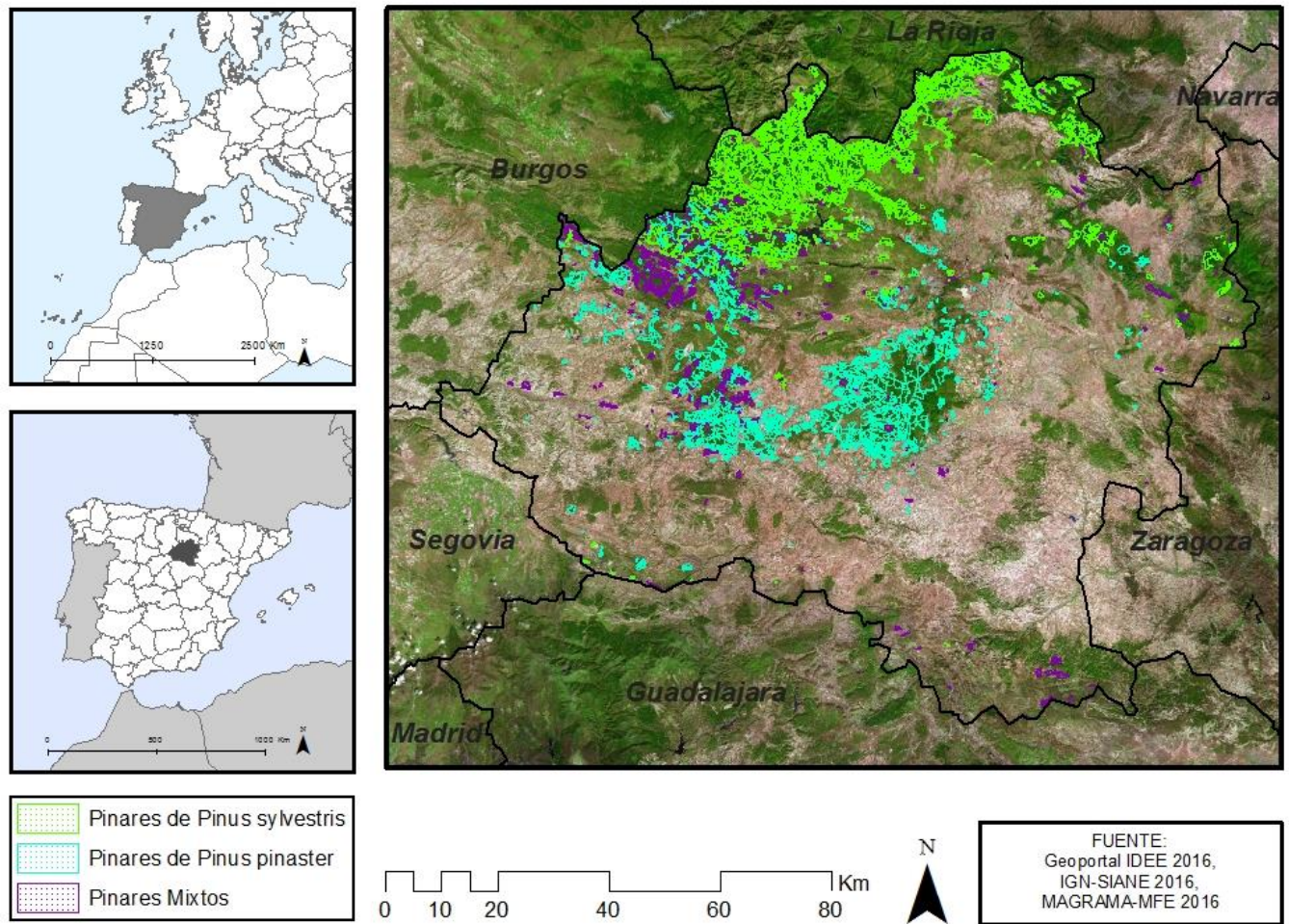


Figura 2. Mapa con la ubicación de las principales masas de coníferas en la provincia de Soria.

Como en el Pinar Grande, en otros pinares de la provincia el proceso ha sido paralelo. Los beneficios que reporta tanto al medio ambiente como al medio socioeconómico son considerables: Por un lado, ha integrado los aprovechamientos pascícolas, cinegéticos, micológicos, apícolas, turísticos y madereros; ha generalizado los tratamientos silvícolas arraigando entre la población local la concienciación para con el bosque como patrimonio natural, ambiental y material; ha supuesto una mejora radical en el estado de salud de las masas forestales disminuyendo hasta niveles inapreciables la incidencia de plagas e incendios forestales y ha creado una red de caminos y vías de saca que estructuran el bosque proporcionando un alto grado de accesibilidad para la prevención de incendios y plagas. (Santolaya, s.f).

No menos importante resulta su repercusión sobre el medio humano. Los puestos de trabajos que genera el sector resultan vitales para mantener la población en los núcleos rurales de Soria, en los que la despoblación causó estragos a lo largo de la segunda mitad de siglo XX. Dentro del sector destaca la explotación maderera, la industria transformadora y la extracción de productos no maderables, entre los que cobra especial relevancia la biomasa.

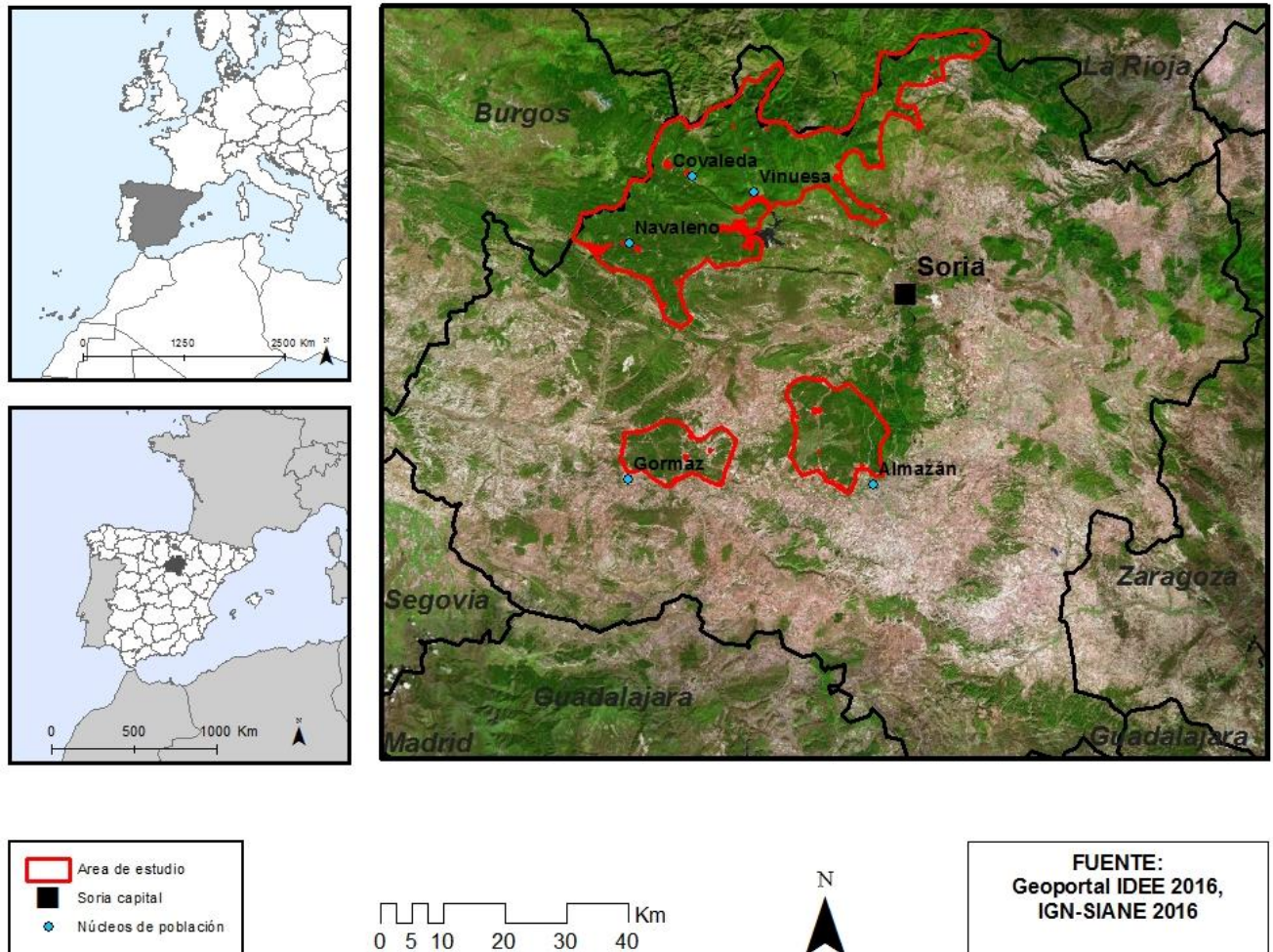


Figura 3. Mapa del área de estudio con núcleos de población cercanos.

Para delimitar el área de estudio sobre la que se focaliza este TFM se buscó definir zonas forestales de las especies de coníferas significativas en Soria y del suficiente tamaño para trabajar con ellas a escala provincial. Para ello, se partió de las teselas de *Pinus sylvestris*, *Pinus pinaster* y formaciones mixtas de ambas especies, circunscritas a la provincia de Soria y clasificadas por el *Mapa Forestal de España* (MFE, MAGRAMA, 2016). Naturalmente, muchas de estas teselas presentaban una distribución sobre el territorio desagregada, por lo que, a través de una serie de geoprocursos desde el software ArcGIS 10.3 se fueron desechando aquellas teselas aisladas o que conformaban conglomerados sin la suficiente entidad espacial; mientras que aquellas que, junto con sus vecinas representaban agregados compactos con superficie suficiente, se unieron formando el área de estudio. Al perímetro resultante se le sustrajo la superficie correspondiente a láminas de agua y suelo urbano.

Como se aprecia en la Figura 2, la zona norte de la provincia, alberga bosques de *Pinus sylvestris*, la zona noroeste, alberga formaciones mixtas y en la zona centro, es donde se ubican los bosques de *Pinus pinaster*. Tras geoprocusar las teselas del MFE y agregarlas definiendo el área de estudio (Figura 3), esta quedó constituida por tres sub-áreas, que se tratarán en adelante y en todo momento como una sola entidad. La mayor de estas áreas es la más septentrional, limítrofe con La Rioja y la provincia de Burgos, que incluye en su territorio elementos destacados de la geografía soriana tales como la Laguna Negra y los Picos de Urbión o el mentado Pinar Grande. Supone una superficie total de 94.438 ha. La segunda de estas áreas por extensión se ubica inmediatamente al noroeste de Almazán y al sur de la capital, y se extiende a lo largo de 20.847 ha. Finalmente, el área de menor extensión, próximo a la localidad de Gormaz, alcanza las 13.445 ha. En total, la superficie del área de estudio suma 128.730 ha., lo que representa el 12,5% de las 1.029.765 ha. de la provincia.

3. PROCESO METODOLÓGICO

Para la consecución del objetivo general de este TFM se seguirá un proceso metodológico estructurado en fases:

- La primera de estas fases será la de selección, adquisición y procesado de los datos que se emplearán durante el proceso. En adelante -y para sintetizar- esta fase se nombrará como “Procesado de datos”. Dentro de esta, se detallará el proceso empleado para transformar los datos brutos en información específica de las tres principales fuentes de información a emplear: LiDAR, ortofotos y Sentinel 2A.
- La segunda fase a realizar será la de “Segmentación” en la que, empleando el software eCognition, se dividirá el área de estudio en polígonos homogéneos, a partir de la información aportada por las capas previamente generadas que participarán del proceso de la segmentación. Estos polígonos servirán de soporte para estructurar las fases posteriores y recibirán información temática heredada de los estadísticos LiDAR y los productos Sentinel previamente generados, además de información sobre sus propiedades geométricas aportada desde el entorno del software eCognition.
- Sobre estos polígonos se realizará una toma de puntos *verdad-terreno*, una selección de puntos de entrenamiento cuya cobertura es conocida gracias a las fuentes de información de las que se dispone.
- La última fase del proceso metodológico será la clasificación de los segmentos creados a partir de la *verdad-terreno*, con el objeto de aislar estadísticamente aquellos que se correspondan con cortas a hecho. Para esto se empleará el paquete Random Forest del lenguaje de programación estadístico R. Previamente habrá sido necesario un proceso de depuración y complementación de la información temática asociada a la capa vectorial resultante de la segmentación.

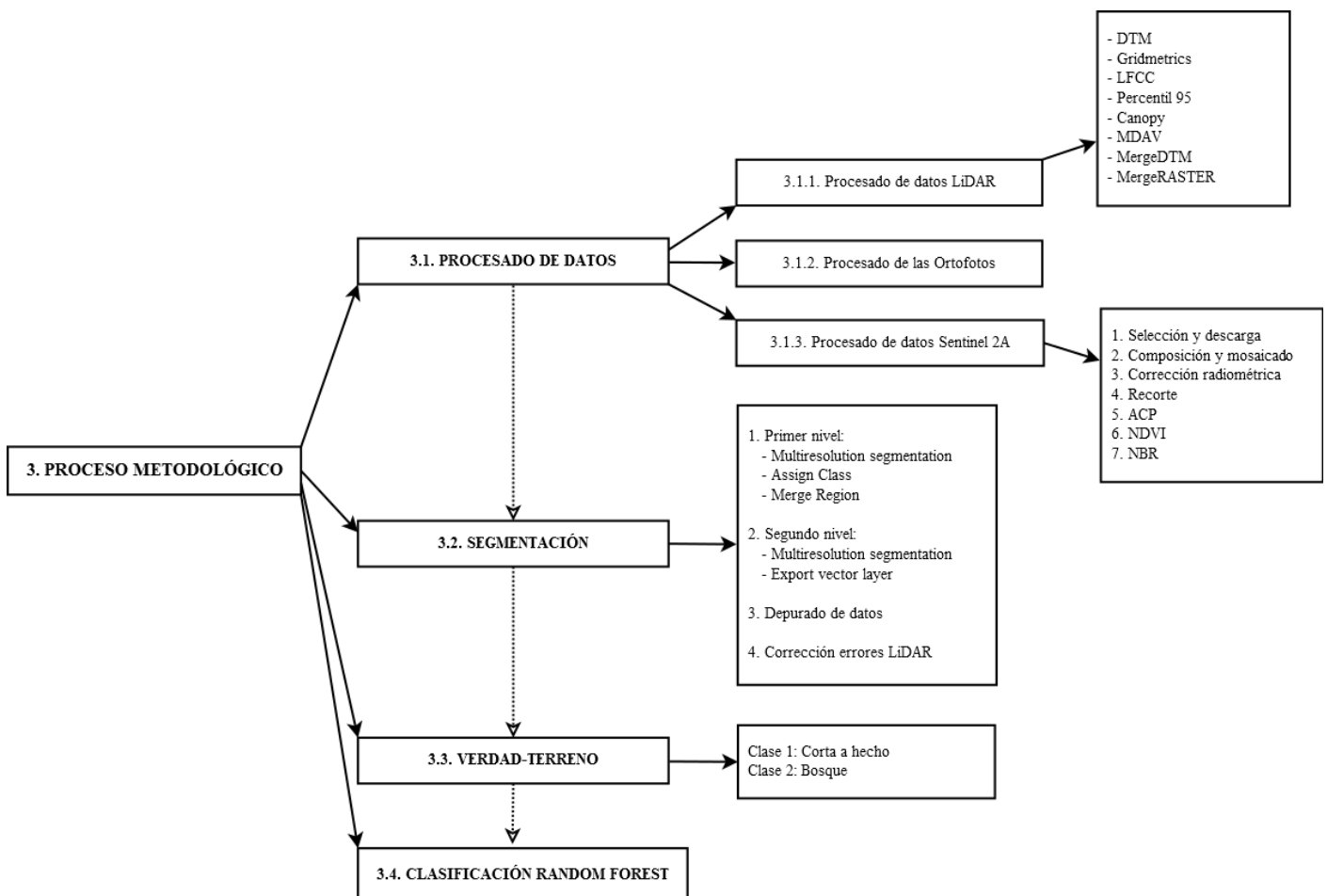


Figura 4. Diagrama de flujos del proceso metodológico

3.1. Procesado de datos

El procesado de los datos es una de las fases más extensas del proceso metodológico por tres motivos. El primero, que ya se adelantó en el punto anterior, responde a que, además del procesado estricto de los datos, previamente requiere de su selección y adquisición y, por tanto, del conocimiento de su naturaleza. El segundo, obedece a que se habrán de procesar datos obtenidos de tres fuentes diferentes: Datos LiDAR del PNOA de 2010, ortofotos de máxima actualidad también del PNOA y datos procedentes del satélite Sentinel 2^a, puestos a disposición del público por la Agencia Espacial Europea (ESA). Finalmente, el tercer motivo que explica la extensión en la duración de esta primera fase es la gran cantidad de tiempo de procesado que requieren algunas de las operaciones, especialmente cuando se han de procesar volúmenes de información de la magnitud necesaria para abarcar el área de estudio antes propuesta.

A continuación, se detallarán las operaciones realizadas para preparar los datos obtenidos de cada una de estas fuentes para su utilización en las fases sucesivas del proyecto.

3.1.1. Procesado de datos LiDAR

LiDAR (Laser Imaging Detection and Ranging) es una tecnología que permite conocer la distancia entre el sensor y un punto de la superficie terrestre midiendo el desfase temporal entre la emisión y la recepción del pulso electromagnético. Se trata, por tanto, de una tecnología de las consideradas activas dentro del mundo de los sensores de teledetección. Gracias a la acción combinada del sensor laser con sistemas de posicionamiento global (GPS) es posible localizar estos puntos sobre el terreno y, conociendo la altura del sensor en el momento de la captación de la información, es posible determinar la altitud de cada uno de estos puntos con precisiones que van desde centímetros hasta milímetros, en función de la naturaleza del sensor que, como en el caso de la teledetección óptica, pueden ser terrestres, aéreos o satelitales.

El producto resultante es una nube de puntos georreferenciada a partir de la cual se pueden realizar una serie de productos derivados. Concretamente, de cara al desarrollo de este proyecto, se generará un modelo digital de elevaciones -DTM, por sus siglas en inglés-, un modelo digital de altura de la vegetación -MDAV-, un ráster de la fracción de cavidad cubierta -LFCC-, un Canopy y un percentil 95 de la altura de la vegetación. Más adelante se profundizará en la naturaleza de cada uno de estos elementos.

La fuente de la que se han obtenido los datos LiDAR ha sido el Plan Nacional de Ortofotografía Aérea -PNOA- a través del portal de descargas del Centro Nacional de Información Geográfica -CNIG- haciendo uso del visor de descargas que ofrece la posibilidad de descargar las hojas de LiDAR por superposición de una capa vectorial, para lo que se empleó el shape del área de estudio (Figura 3). Los archivos que se descargan del visor vienen en formato LAZ (formato de compresión de ficheros LAS) y contienen información altimétrica de la nube de puntos LiDAR, distribuidos en ficheros de 2x2 km de extensión. Las nubes de puntos han sido capturadas mediante vuelos con sensor LiDAR con una densidad de 0,5 puntos/m² -o superior-, y posteriormente clasificadas de manera automática y coloreadas mediante RGB obtenido a partir de ortofotos del PNOA con tamaño de pixel de 25 o 50cm. (CNIG, 2016). La fecha de captura de la información varía en función a la comunidad autónoma, en el caso de Castilla y León, data de 2010 (PNOA, 2016).

➤ Generación de rásters a partir de datos LiDAR en bruto.

▪ DTM

Un modelo digital del terreno -DTM o MDT- es *un conjunto de datos numéricos que describe la distribución espacial de una característica del territorio* (Doyle, 1978), en este caso, concretamente, la cota sobre el nivel del mar, por lo que se puede denominar específicamente modelo digital de elevaciones (MDE), al margen de la denominación genérica para este tipo de archivos en el entorno del software de código libre Fusion. Se trata de un archivo de naturaleza ráster empleado para normalizar los valores de los demás productos derivados de la nube de puntos LiDAR. También se empleará posteriormente para generar los ráster de pendientes y orientación.

El proceso para obtenerlo se realizó mediante uno de los algoritmos implementados en Fusion, ejecutados desde la ventana de Símbolo del Sistema. En concreto, se empleó el algoritmo *GridSurfaceCreate*, articulado mediante la siguiente sentencia¹:

```
C:\FUSION\GridSurfaceCreate.exe /class:2 D:\alberto_2016\Datos_Lidar\dtm\490-4632.dtm 5 m m 1 0 0 0 D:\alberto_2016\Datos_Lidar\PNOA_2010_Lote5_CYL-RIO_490-4632_ORT-CLA-COL.LAZ
```

Donde:

C:\FUSION\GridSurfaceCreate.exe es el algoritmo; */class:2* es un parámetro (switch) que indica que solo los puntos de clase 2 (ground) se emplearán para generar el ráster;

D:\alberto_2016\Datos_Lidar\dtm\490-4632.dtm es el archivo de salida;

5 m m 1 0 0 0 son una serie de parámetros cuya razón (en orden) es: indica el tamaño de celda; unidades de medida latitud/longitud; unidades de medida en altura; sistema de coordenadas UTM; zona del sistema de coordenadas desconocida, datum desconocido, y datum vertical desconocido;

D:\alberto_2016\Datos_Lidar\PNOA_2010_Lote5_CYL-RIO_490-4632_ORT-CLA-COL.LAZ es el archivo de salida.

▪ Gridmetrics

El siguiente paso a realizar tras la generación de los DTM consistió en obtener una serie de estadísticas descriptivas sobre la composición y estructura de la nube de puntos sobre las que se sustentarán procesos posteriores de creación de rásters. El algoritmo empleado se denomina *gridmetrics* y requiere de los DTM para normalizar los valores altitudinales de la nube de puntos, es decir, para transformar la altitud elipsoidal en la que vienen los datos en origen a altura de los puntos sobre el terreno. Para realizar este paso, se utilizó la siguiente sentencia:

```
C:\FUSION\gridmetrics /minht:2 D:\alberto_2016\dtm\490-4632.dtm 2 10 D:\alberto_2016\gridmetrics\490-4632.csv D:\alberto_2016\datos_Lidar\PNOA_2010_Lote5_CYL-RIO_490-4632_ORT-CLA-COL.LAZ
```

Donde:

C:\FUSION\gridmetrics es el algoritmo; */minht:2* es el parámetro (switch) que define el umbral de puntos por unidad de superficie por debajo del cual no genera resultados.

D:\alberto_2016\dtm\490-4632.dtm es el archivo dtm empleado para normalizar valores altitudinales.

2 10, son, por este orden, indicador de la altura de corte por debajo de la cual no considera registros y tamaño de celda.

D:\alberto_2016\gridmetrics\490-4632.csv es el archivo de salida.

D:\alberto_2016\datos_Lidar\PNOA_2010_Lote5_CYL-RIO_490-4632_ORT-CLA-COL.LAZ es el archivo de entrada.

¹ Aquí, así como en casos similares, se debe generar una sentencia por cada archivo de entrada o bien realizar un proyecto batch.

- LFCC (Fracción de cabida cubierta)

La fracción de cabida cubierta es un índice que expresa el grado de recubrimiento del suelo por la proyección vertical de las copas de arbolado. Por tanto, se trata de un indicador que aporta información sobre la espesura del dosel vegetal. Se expresa en tanto por ciento. Para obtenerla es preciso haber generado previamente las estadísticas de la nube de puntos mediante el comando *gridmetrics* y aplicar el comando del entorno Fusion *csv2grid*. La sentencia que articula la orden es la siguiente:

```
C:\FUSION\csv2grid ndzero:0 D:\alberto_2016\gridmetrics\490-4632_all_returns_elevation_stats.csv 49
D:\alberto_2016\raster\LFCC\490-4632.asc
```

Donde:

C:\FUSION\csv2grid es el algoritmo; *ndzero:0* asigna valores 0 a los NoData;
D:\alberto_2016\gridmetrics\490-4632_all_returns_elevation_stats.csv es el archivo de entrada
49 es el parámetro que determina que el archivo de salida será la fracción de cabida cubierta.
D:\alberto_2016\raster\LFCC\490-4632.asc es el archivo de salida.

- Percentil 95 de la altura del dosel vegetal (P95)

Se trata de un estadístico que representa el valor bajo el cual se encuentran el 95% de los valores medidos. Su objetivo es filtrar valores extremos -ruido- que pueden haber sido causados por diversos motivos. En el caso de datos LiDAR, un pájaro o un avión pueden interferir en el pulso electromagnético devolviendo al sensor un valor anómalo. Gracias al P95, este valor no se consideraría y no alteraría la estadística de la superficie a parametrizar.

Al igual que en el caso de la fracción de cabida cubierta, requiere para su elaboración de las estadísticas obtenidas mediante el comando *gridmetrics*. La sentencia del entorno Fusión que articula la orden es la siguiente:

```
C:\FUSION\csv2grid D:\alberto_2016\Datos_Lidar\gridmetrics\490-4632_all_returns_elevation_stats.csv 37
D:\alberto_2016\Datos_Lidar\raster\P95\490-4632_P95.asc
```

Donde:

C:\FUSION\csv2grid es el algoritmo
D:\alberto_2016\Datos_Lidar\gridmetrics\490-4632_all_returns_elevation_stats.csv es el archivo de entrada
37 es el parámetro que determina que el archivo de salida será el percentil 95.
D:\alberto_2016\Datos_Lidar\raster\P95\490-4632_P95.asc es el archivo de salida.

- Canopy

Se trata de otro estadístico comúnmente usado en gestión forestal que hace referencia a la mayor altura en el dosel a la que la densidad del combustible es mayor de 0.011 kg / m³ (Andersen et al., 2005). Como en los casos anteriores requiere de las estadísticas del *gridmetrics* y se genera en el entorno Fusión a través de la siguiente sentencia:

```
C:\FUSION\csv2grid.exe D:\alberto_2016\Datos_Lidar\gridmetrics\490-4632_all_returns_elevation_stats.csv
68 D:\alberto_2016\Datos_Lidar\raster\CANOPY\490-4632_canopy.asc
```

Donde:

C:\FUSION\csv2grid es el algoritmo
D:\alberto_2016\Datos_Lidar\gridmetrics\490-4632_all_returns_elevation_stats.csv es el archivo de entrada
69 es el parámetro que determina que el archivo de salida será el canopy.
D:\alberto_2016\Datos_Lidar\raster\CANOPY\490-4632_canopy.asc es el archivo de salida.

▪ MDAV

El modelo digital de altura de la vegetación -MDAV- puede expresarse como la diferencia entre el modelo digital de la superficie en un terreno con presencia de elementos vegetales, y el modelo digital del terreno (MDS-MDE). Se trata, por tanto, de una parametrización del dosel vegetal sin incluir el suelo desnudo. Para generarlo se ha de introducir la siguiente sentencia al entorno Fusion:

```
C:\FUSION\Canopymodel.exe /smooth:3 /outlier:0,40 /ground:D:\alberto_2016\Datos_Lidar\dtm\490-4632.dtm D:\alberto_2016\Datos_Lidar\mdav\490-4632_mdav.dtm 2 m m 1 0 0 0 D:\alberto_2016\datos_Lidar\PNOA_2010_Lote5_CYL-RIO_490-4632_ORT-CLA-COL.LAZ
```

Donde:

C:\FUSION\Canopymodel.exe es el algoritmo; */smooth:3* (switch) que aplica un filtro estadístico a partir de la media de los 3 valores vecinos; */outlier:0,40* es el parámetro (switch) que define el umbral por debajo del cual no se considerará ningún punto; */ground:D:\alberto_2016\Datos_Lidar\dtm\490-4632.dtm* es el parámetro (switch) que define el archivo empleado para extraer los valores del suelo desnudo.

D:\alberto_2016\Datos_Lidar\mdav\490-4632_mdav.dtm es el archivo de salida

2 m m 1 0 0 0, indican en orden: el tamaño de celda; unidades de medida latitud/longitud; unidades de medida en altura; sistema de coordenadas UTM; zona del sistema de coordenadas desconocida, datum desconocido, y datum vertical desconocido.

D:\alberto_2016\datos_Lidar\PNOA_2010_Lote5_CYL-RIO_490-4632_ORT-CLA-COL.LAZ es el archivo de entrada.

➤ Mosaicado y recorte de los rásters generados

Una vez se han creado los rásters, el siguiente paso consiste en realizar un mosaico con ellos para fusionarlos en un único archivo. Para ello se recurrió al comando *Merge* del entorno Fusion. Este comando ofrece dos variantes en función del tipo de archivos a unificar; por ello, en la librería de Fusion se encuentran los algoritmos *MergeDTM*, que permite unificar los archivos con extensión (.*dtm*) que, a efectos de este proyecto, son los DTM y los MDAV, y *MergeRASTER*, a partir del cual se mosaicaron los archivos de LFCC, P95 y Canopy, con extensión (.*asc*).

La sentencia que ordena el mosaicado de archivos (.*dtm*) en el entorno Fusion, que funciona del mismo modo tanto para los DTM como para los MDAV, es la siguiente:

```
C:\FUSION\MergeDTM.exe D:\alberto_2016\Datos_Lidar\dtm\dtm5x5.dtm D:\alberto_2016\Datos_Lidar\dtm\*.dtm
```

Donde:

C:\FUSION\MergeDTM.exe es el algoritmo;

D:\alberto_2016\Datos_Lidar\dtm\dtm5x5.dtm es el archivo de salida;

D:\alberto_2016\Datos_Lidar\dtm.dtm* es el archivo de entrada.

Por su parte, la sentencia que ordena el mosaicado de archivos (.*asc*) en el entorno Fusion es la siguiente, e igualmente funciona de la misma manera para unificar los archivos de LFCC, P95 y Canopy:

```
C:\FUSION\MergeRASTER D:\alberto_2016\raster\Merges\LFCC.tif D:\alberto_2016\Datos_Lidar\raster\LFCC\*.asc
```

Donde:

C:\FUSION\MergeRASTER es el algoritmo;

D:\alberto_2016\raster\Merges\LFCC.tif es el archivo de salida;

D:\alberto_2016\Datos_Lidar\raster\LFCC.asc* es el archivo de entrada.

Una vez quedaron realizados los mosaicos, se procedió a recortarlos empleando como perímetro el shape del área de estudio. El proceso se realizó en el entorno del software ArcGIS 10.3. a través de la herramienta del paquete *Spatial Analyst Tools, Extract by Mask* ubicada en el módulo *Extraction*. A continuación, a partir del ráster del DTM ya recortado, se generaron los rústers de orientación y pendientes mediante las herramientas *Aspect* y *Slope*, respectivamente, ubicadas en el módulo *Surface* del mismo paquete de herramientas.

3.1.2. Procesado de las Ortofotos

La segunda fuente que se ha empleado ha sido la ortofoto de máxima actualidad del PNOA, obtenida, de igual manera que los datos LiDAR, a través del portal de descargas del CNIG, mediante la búsqueda en visor por superposición con el shapefile del área de estudio. Las ortofotos se descargan en formato ECW, con sistema geodésico de referencia ETRS89 y proyección UTM en su huso correspondiente, a efectos de este proyecto, el 30. La unidad de distribución y descarga es la hoja del MTN50. Las ortofotos correspondientes al área de estudio datan de julio y agosto de 2014. Se trata de archivos multibanda -RGB- con una resolución espacial de 0,25 m (CNIG, 2016).

Su procesado es sencillo en comparación con los datos LiDAR puesto que tan solo es necesario realizar un mosaico con ellas y, posteriormente, recortar el mosaico resultante tal y como se hizo anteriormente. Para realizar el mosaico se recurrió al software ERDAS IMAGINE 2013, concretamente a la herramienta *Mosaic Pro*, de la sección *Geometry* del menú *Ráster*. Para recortar el mosaico se utilizó, de nuevo, la herramienta *Extract by Mask* del *Spatial Analyst* de ArcGIS 10.3.

3.1.3. Procesado de datos Sentinel 2A

Como se anticipaba en la Introducción, Sentinel 2A es el primero de una constelación de satélites ópticos de la ESA al que pronto se sumará su hermano gemelo, Sentinel 2B, y, a partir de 2021, los Sentinel 2C y 2D (Airbus, 2016).

Ofrece de manera gratuita y con una periodicidad de 10 días -a la espera de reducirla a la mitad con el lanzamiento de Sentinel 2B- imágenes con información de 13 bandas entre el ultravioleta y el infrarrojo de onda corta (SWIR), con resoluciones espaciales comprendidas entre los 10 m de 3 de las bandas del visible y uno de los NIR, es decir, las bandas 2, 3, 4 y 8, y los 60 m de la banda 1, en la región ultravioleta -diseñada para observar aerosoles-, de la banda 9 -diseñada para monitorear el vapor de agua atmosférico- y la banda 10 -denominada cirrus por ser su función la observación de las nubosidades de la atmósfera-.

El procesado de las imágenes Sentinel incluye varias fases. En primer lugar, es necesario seleccionar y descargar las imágenes pertinentes. A continuación, es preciso componer una imagen multi-banda con ellas, ya que se descargan las bandas individualmente. Después se deben unir las escenas de la imagen descargadas y corregirlas radiométricamente. Posteriormente se deben recortar al perímetro del área de estudio. Además, de cara a posteriores procesos, se generaron una serie de neocanales con la esperanza de que aportaran información relevante a la hora de cartografiar las cortas a hecho. Concretamente se realizó un análisis de componentes principales (ACP), un índice de vegetación de diferencia normalizada (NDVI) y una ratio normalizado de quema (NBR).

➤ Selección y descarga de las imágenes

La descarga de imágenes Sentinel es gratuita con solo registrarse a través del portal de descargas del programa Copernicus de la ESA “Sentinels Scientific Data Hub”. El primer paso para descargarlas es realizar una selección del área del que se pretenden las imágenes. A continuación, se debe especificar la fecha o rango temporal de captura o de procesamiento de las imágenes, el satélite del que se desean las imágenes -actualmente Sentinel 1, 2 ó 3- y algunos parámetros específicos de cada satélite; en el caso de Sentinel 2, la cobertura nubosa tolerada.

Las imágenes vienen divididas en escenas y éstas, a su vez, en cada una de las bandas espectrales, por lo que es preciso componerlas *a posteriori*. Para el desarrollo de este estudio se seleccionaron las escenas del centro y este de la imagen S2A_OPER_PRD_MSIL1C_PDMC_20160703T183124_R094_V20160703T1106, registrada el 3 de julio de 2016 y que cubre el área de estudio.

➤ Composición y mosaicado de las imágenes

De las 13 bandas Sentinel 2, se seleccionaron para el desarrollo de este proyecto las bandas 2, 3 y 4, que registran en las longitudes de onda del visible con una resolución espacial de 10 m, y la banda 8, correspondiente con el infrarrojo próximo -NIR-, con idéntica resolución espacial cuyo centro de banda se ubica en torno a los 840 nm. Junto con ellas se seleccionaron también las bandas 11 y 12, en adelante denominadas SWIR1 y SWIR2 a efectos de este proyecto, que registran en la región del infrarrojo de onda corta, en torno a los 1600 nm la primera, y a los 2200 nm la segunda.

Tabla 1. Bandas originales de Sentinel 2A seleccionadas para el multibanda.

Nº banda Sentinel 2A	Denominación	Centro de banda (nm)	Resolución espacial (m)
2	BLUE	490	10
3	GREEN	560	10
4	RED	665	10
8	NIR	842	10
11	SWIR1	1610	20
12	SWIR2	2190	20

La razón de que se seleccionaron estas 6 bandas y no otras de entre las trece de las que dispone Sentinel 2 radica en dos cuestiones. Por una parte, son las que mayor resolución espacial ofrecen, considerando los 10 m de pixel de las tres bandas del visible y el infrarrojo próximo y los 20 m de los dos SWIR. Y por la otra, se consideró que estas seis bandas registraban las regiones del espectro óptico de mayor interés y provecho para la ejecución del proyecto. Así, las bandas 1 -aerosoles-, 9 -vapor de agua- y 10 -cirrus- quedaban descartadas tanto por resolución espacial como por su rango de registro. De entre los otros cinco infrarrojos -B5, B6, B7, B8, B8a- se seleccionó la 8 -en adelante y para efectos de este proyecto denominada NIR o B4- por ser la que mayor resolución espacial ofrecía.

Para unificar las bandas en una sola imagen se recurrió al software ERDAS IMAGINE 2013. Este programa dispone de una herramienta denominada *Model Maker* que permite diseñar funciones a partir de una serie de algoritmos que implementa en su biblioteca. Concretamente se empleó el algoritmo *Stack Layer* que dio como resultado dos imágenes multibanda, una por escena. A la hora de unir las dos imágenes resultantes en una sola, la herramienta utilizada fue *Mosaic Pro*, como en el caso de las ortofotos.

➤ Corrección radiométrica de las imágenes

Las imágenes Sentinel se descargan del servidor de la ESA con nivel de procesado 1C, o lo que es lo mismo, vienen corregidas al techo de la atmósfera (TOA). De cara a garantizar que se trabaja con valores no alterados por el efecto dispersor de la atmósfera, es decir, para corregirlas al suelo de la atmósfera (BOA), se aplicó el método de substracción de objeto oscuro (DOS), propuesto por Chaves (1988) y utilizado en numerosas aplicaciones por su aplicación simple. Postula que la bruma atmosférica incrementa los ND en toda la imagen, siendo detectable su magnitud en áreas de agua limpia, profunda y calma donde, por las características físicas, deben presentar reflectancia nula. El valor representativo de esa diferencia se sustrae, en cada banda, a todos los píxeles de la escena.

La aplicación de este método se realizó en el entorno de ERDAS IMAGINE 2013. El primer paso consiste en localizar aquellos píxeles que representen el valor mínimo en el histograma de cada una de las bandas al margen de valores correspondientes a ruido presentes en las imágenes. Para ello se abrió la imagen en pseudocolor y se seleccionó la primera de las bandas. En el menú *Table*, se activó el comando *Show Attributes* que ofrece la posibilidad de detectar en la imagen un determinado valor del histograma. Se ordenaron los valores de los niveles digitales de menor a mayor y se localiza el primer ND con representación en el histograma -frecuencia-. Esta operación ha de realizarse en cada una de las bandas.

Existe una relación de proporcionalidad entre el valor mínimo de cada una de las bandas que responde a la máxima que establece que, a menor longitud de onda, mayor dispersión atmosférica. Por lo tanto, la banda del azul debe presentar un valor superior a la del verde y esta a su vez superior a la del rojo, y así sucesivamente. Los valores sustraídos al mosaico recién generado figuran en la tabla 10.

Para realizar la corrección se recurrió de nuevo a la herramienta *Model Maker* que permite sustraer valores a una imagen de entrada, banda por banda, de una sola vez, mediante una tabla que recoja los valores de cada banda.

Tabla 2. Bandas Sentinel 2 integrantes de la imagen multibanda con nivel digital mínimo e índice de ND mínimo.

	BANDA	ND MIN	Bn/B1
B1	BLUE	588	1
B2	GREEN	381	0.65
B3	RED	193	0.33
B4	NIR	132	0.22
B5	SWIR1	35	0.06
B6	SWIR2	11	0.02

➤ Recorte de la imagen Sentinel

Una vez se dispuso de la imagen Sentinel 2 compuesta, mosaicada y corregida se recortó empleando el perímetro del área de estudio, tal y como se hizo con los ráster obtenidos de la nube de puntos LiDAR. Para ello, desde el entorno ArcGIS 10.3, se utilizó de nuevo la herramienta *Extract by Mask* del paquete *Spatial Analyst*. A partir de la imagen multibanda resultante se calcularon los neocanales mencionados con anterioridad: ACP, NDVI y NBR.

➤ Análisis de Componentes Principales (ACP)

Se trata de una técnica estadística entroncada en las técnicas multivariantes de síntesis de la información, cuyo objetivo es resumir un amplio abanico de variables en un nuevo conjunto de menor tamaño sin perder una parte significativa de la información original. La creación de los Componentes Principales se consigue a partir de la definición de unos nuevos ejes (y, por tanto, de unas nuevas coordenadas) en un espacio *n-dimensional* mediante la rotación y traslación de los ejes originales de las bandas (Chuvieco, 1996).

Las 3 primeras bandas resultantes de este proceso suelen asociarse a tres parámetros físicos: Brillo, verdor y humedad, si bien el resultado para cada imagen de satélite es independiente, pudiendo expresar la propiedad inversa, es decir, marchitez en lugar de verdor y sequedad en lugar de humedad. Del mismo modo, la interpretación de los componentes 2 y 3 puede invertirse en ocasiones, recogiendo el CP2 características espectrales relacionadas con la humedad/sequedad mientras que el CP3 recoge las relacionadas con el verdor/marchitez.

El número de componentes principales que se puede generar es, como máximo, igual al número de bandas de la imagen original. No obstante, de cara a la ejecución de este proyecto, se generaron solamente los tres primeros, que recogen la mayor parte de la información y se pueden relacionar de forma empírica con los factores biofísicos mencionados.

El proceso se realizó mediante ERDAS IMAGINE 2013, a través de la herramienta homónima del paquete *Spectral* en el menú *Raster*. Esta herramienta nos ofrece la posibilidad de generar una matriz con los eigenvectores, que expresan la relación direccional entre los viejos y los nuevos valores, y un archivo con los eigenvalores, que recogen la proporción de información original contenida en el nuevo canal generado.

El objetivo de la aplicación de este análisis en este proyecto era obtener una capa del área de estudio que sintetizara en tres neocanales el contenido de las 6 originales de la imagen Sentinel. Además, este nuevo multibanda tenía, en una resolución espacial de 10 m, la información de los SWIR, cuyo pixel era de 20 m de lado. Esto se debe a un proceso interno de remuestreo automático que, si bien es cierto que en superficies mayores puede generar ruido, no resulta así para un área de estudio como la de este proyecto.

➤ NDVI

Normalized Difference Vegetation Index (NDVI), es un índice de vegetación y, como tal, proporciona información sobre el estado de la vegetación relacionando las distintas bandas de las imágenes multispectrales. Se calcula a partir de las bandas del rojo y del NIR ya que es en las longitudes de onda que ambas comprenden donde se recoge la mayor parte de la información sobre la vegetación. La banda del rojo aporta información sobre la absorción clorofiliana mientras que la del infrarrojo próximo, sensible a las variaciones en la estructura interna de la hoja, informa sobre el vigor vegetal. El gran contraste que presenta la vegetación sana en estas dos bandas las hace especialmente interesantes para estudiar el estado de la vegetación. Responde a la siguiente fórmula:

$$NDVI = \frac{NIR - R}{NIR + R}$$

Se trata de la diferencia normalizada de las bandas del NIR y del rojo que genera un ráster cuyos valores oscilan entre -1 y +1. El objetivo de incluirlo en el proyecto es el de dotar al modelo que se elaborará con posterioridad de un índice con capacidad para identificar la vegetación y discriminarla de otras coberturas.

Para generarlo se recurrió de nuevo al *Model Maker* de ERDAS que permite operar como una calculadora ráster con las bandas de una imagen multibanda.

➤ NBR

Normalized Burn Ratio (NBR) o ratio normalizado de quema, es un índice diseñado para medir la severidad de los incendios. El objetivo de incluirlo entre las capas que posteriormente engrosarán la segmentación y la clasificación, radica en la similitud espectral entre un incendio y una corta a hecho, ya que ambos suponen un afloramiento del suelo desnudo subyacente al dosel vegetal cuando este desaparece. La ecuación que lo define es similar a la del NDVI, pero emplea el SWIR2 en lugar del rojo. Esta variación le permite recabar información sobre el suelo desnudo, cuya signatura espectral mantiene un signo ascendente a lo largo del espectro óptico por lo que, es en la banda del SWIR2 en la que más variabilidad ofrece de entre las disponibles. El rango de sus valores va desde -1 a +1.

$$NBR = \frac{NIR - SWIR2}{NIR + SWIR2}$$

La manera de generarlo es la misma que en el caso del NDVI, a través de la calculadora ráster del *Model Maker*, aplicando la ecuación de la fórmula.

3.2. Segmentación del área de estudio

Una vez estuvieron preparadas las capas resultantes del procesado de la información bruta extraída de las fuentes, el siguiente paso consistía en segmentar el área de estudio a partir de la información contenida en las capas. El objeto de la segmentación es dividir el área de estudio en zonas homogéneas interiormente y heterogéneas con respecto a sus vecinas. Cada una de estas zonas, denominadas objetos o segmentos, dentro del entorno del software empleado para su desarrollo -eCognition de Trimble-, debía ir acompañada de una tabla de atributos con información temática que permitiría posteriormente su clasificación en el entorno de R, empleando el paquete Random Forest, tratando de diferenciar las cortas a hecho del resto de superficies. Esta información temática se conformó a partir de la media del valor de los píxeles de las diferentes capas preparadas con anterioridad comprendidos dentro de cada segmento. Del mismo modo otros estadísticos relativos a estas capas, como la desviación estándar, participaron de los segmentos resultantes. Finalmente, eCognition aportará información respecto a las propiedades geométricas de los segmentos generados.

El proceso comienza con la selección de las capas que se emplearon durante la segmentación. En este caso, se seleccionaron por una parte todos los estadísticos LiDAR generados en formato ráster, incluyendo los rústers de orientación y pendiente; la imagen multibanda Sentinel 2, el NDVI, el NBR y los tres componentes principales generados.

Como se ha venido repitiendo con anterioridad, el objetivo del proyecto es detectar y cartografiar cortas a hecho, es decir, polígonos con formas regulares –principalmente fajas-, relativamente estrechos, que forman patrones reconocibles, emplazados dentro de masas boscosas, representando una discontinuidad abrupta en altura respecto a sus inmediaciones y con un comportamiento espectral semejante al del suelo desnudo o al del matorral, en función del estado de la regeneración.

Bajo estas premisas parece lógico pensar que los datos LiDAR son los idóneos para soportar el peso de la segmentación y definir con la máxima precisión y eficacia las cortas a hecho. Por un lado, presentan una mayor resolución espacial que los productos derivados de Sentinel 2, lo que les permite ajustarse mejor a las dimensiones de las fajas, ocasionando en menor medida píxeles con mezcla de varias superficies. Por otro lado, en vista de la idéntica respuesta espectral entre las cortas y los cami-

nos u otras superficies de suelo desnudo, la modelización tridimensional de la realidad que posibilita LiDAR supone una alternativa a este problema.

Aparentemente la tecnología LiDAR es la más apropiada para dirigir la segmentación; sin embargo, los datos obtenidos del PNOA datan del año 2010, lo que supone que no es posible detectar con ellos cortas a hecho posteriores a esta fecha.

Por otro lado, tal y como se adelantó en la introducción, resulta de gran interés explorar las posibles aplicaciones de una tecnología de acceso gratuito y libre para toda la población, y cuyas mejoras técnicas frente a sus competidores no son en absoluto desdeñables, como es Sentinel 2.

Para resolver el problema de la similitud espectral entre fajas, caminos y otras superficies que pudieran confundirse, al mismo tiempo que se aprovechaba de la mejor manera posible la información procedente de Sentinel, el proceso de segmentación se estructuró en dos fases o niveles: Una primera fase en la que el peso de la segmentación recayó exclusivamente en el MDAV, cuya resolución espacial es de 2 m, y una segunda en la que la imagen multibanda Sentinel y sus productos derivados fueron los encargados de aportar la información que delimitó los segmentos.

El sentido de la primera segmentación a partir del MDAV consistía en diferenciar con gran precisión las zonas que en 2010 eran bosque y las que no. Esto permitió diferenciar las cortas a hecho de todas aquellas superficies presentes antes de 2010 sin vegetación, como caminos, roquedos, superficies agrícolas en torno a los núcleos rurales, etc. Sin embargo, esta opción, que funciona a modo de cribado masivo, también deja fuera todas aquellas cortas a hecho anteriores a la captación de los datos LiDAR, mientras que no cribará todas aquellas pistas forestales posteriores a 2010. Ambos casos se asumen como males necesarios. En el caso de las cortas con más de 6 años, se entiende que deben de encontrarse en una fase de regeneración avanzada. Por otro lado, se considera que ninguna masa boscosa ha podido desarrollarse lo suficiente en 6 años como para ser susceptible de soportar esta práctica silvícola a día de hoy.

Tabla 3. Capas integradas en el proceso de segmentación de eCognition con indicación de su resolución espacial.

CAPA	RESOLUCIÓN ESPACIAL (m)
Componente principal 1 (CP1)	10
Componente principal 2 (CP2)	10
Componente principal 3 (CP3)	10
Aspect	5
CANOPY	5
LFCC	10
MDAV	2
NBR	10
NDVI	10
Percentil 95 (P95)	10
SLOPE	5
Sentinel B1 (BLUE)	10
Sentinel B2 (GREEN)	10
Sentinel B3 (RED)	10
Sentinel B4 (NIR)	10
Sentinel B5 (SWIR1)	20
Sentinel B6 (SWIR2)	20

➤ Segmentación inicial sobre primer nivel: *Multiresolution segmentation*

Para la ejecución de esta primera fase, se implementaron las capas recogidas en la tabla 11 y se definió un subconjunto espacial, a fin de comprobar los resultados y poder realizar las pruebas necesarias sin que el tiempo de procesado se alargara.

A continuación, se abrió un árbol de procesos al que se añadió un nuevo apéndice -en el entorno eCognition, cada apéndice envuelve un algoritmo destinado a realizar una acción en base a umbrales y parámetros definidos por el criterio experto del usuario-. El algoritmo seleccionado para este primer apéndice fue *multiresolution segmentation*, diseñado para segmentar superficies a partir de información de entrada con diversas resoluciones espaciales. Se trata de un algoritmo que opera sobre píxeles para generar objetos. Sus principales parámetros a ajustar son, por un lado, el peso que tiene cada capa en la segmentación, para lo que se le atribuyó al MDAV un valor de 5 mientras que las demás capas no intervinieron, y, por otro lado, el factor escala, que regula el tamaño de los segmentos a generar. Este último es un parámetro empírico, dependiente de la resolución espacial de las capas implicadas, que ocasionará resultados distintos en función del entorno del proyecto. En este caso se le dio un valor de 40. El siguiente parámetro para ajustar es el de forma que, de manera dicotómica, define si la segmentación se ajustará a las propiedades de forma o de color de las capas, es decir, a las propiedades geométrico-contextuales o espectrales; el rango de ajuste va de 0 a 1, siendo 0 la máxima influencia de las propiedades espectrales y 1 la de las contextuales. El valor que se empleó fue de 0,2. El último parámetro a regular es la compacidad, que determina si los segmentos que se formarán tendrán formas compactas o no. Se le atribuyó un valor neutro, de 0,5, sobre su rango comprendido entre 0 y 1.

Tabla 4. Parámetros del algoritmo “*Multiresolution segmentation*” de eCognition.

Parámetro	<i>Multiresolution segmentation</i>
Domain	Pixel level
Level name	New level
Image Layer weights	MDAV (5), resto (0)
Scale parameter	60
Shape	0,2
Compactness	0,5

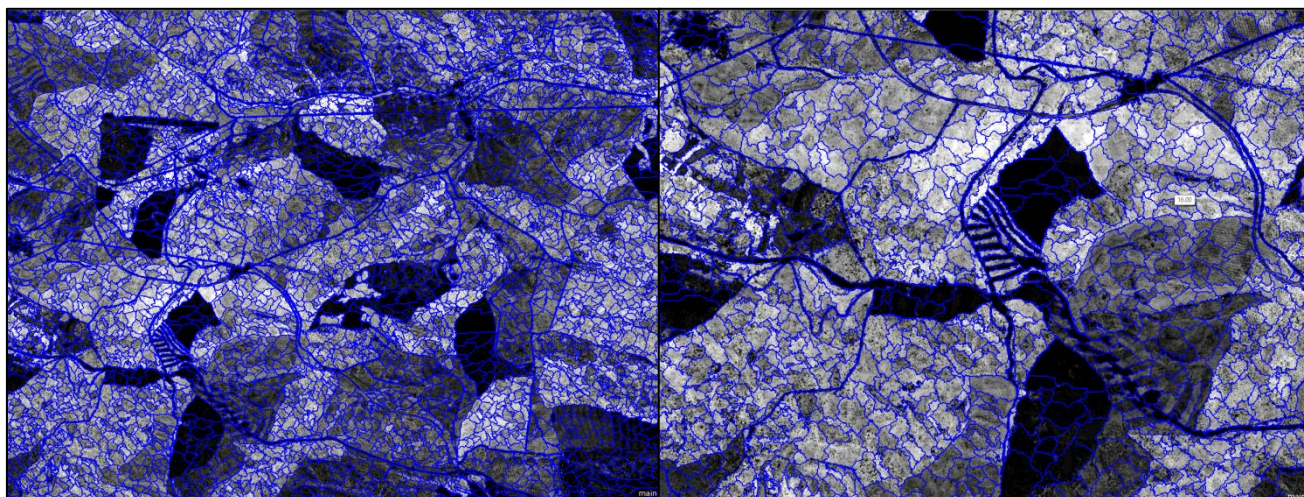


Figura 5. Segmentación generada por el algoritmo “*Multiresolution segmentation*” de eCognition; a la derecha, ampliación sobre el sector inferior izquierdo.

Como se aprecia en la figura 4, a partir del MDAV y el algoritmo *multiresolution segmentation* con la parametrización adecuada, es posible realizar una segmentación que define con gran precisión las cortas a hecho, así como otros elementos del entorno, tales como masas boscosas y caminos. A

partir de esta segmentación se constituyó el primer nivel del proceso, sobre ella se establecerán dos clases: bosque y no bosque. El segundo nivel del proceso de segmentación consistió, como se dijo con anterioridad, en realizar una segunda segmentación, exclusivamente sobre los objetos clasificados como bosque, empleando la fuente de información más actual de la que se disponía, Sentinel 2. De esta manera se excluyó toda superficie que pudiera confundirse con las cortas a hecho, principalmente caminos, claros de los bosques o superficies agrícolas.

➤ Definición de clases: *Assign Class*

Para separar los objetos segmentados se empleó el algoritmo *assign class*, que trabaja a nivel de objetos y no de píxeles como anteriormente. Este algoritmo ofrece la posibilidad de definir un umbral que clasifique en una nueva categoría generada a tal efecto a todos los objetos que cumplan la condición. En este caso, la clase creada fue “Bosque” y el umbral clasificaba como tal todos los objetos cuya media del MDAV fuera mayor o igual a 8 m y que, además, tuvieran una fracción de cabida cubierta de por lo menos el 60%. Acto seguido se generó la clase complementaria “No Bosque” mediante el mismo algoritmo agrupando todo segmento no clasificado como bosque. El umbral establecido sobre el MDAV garantizaba que en la clase “Bosque” solo hubiera segmentos correspondientes a masas que en 2010 alcanzaban los 8 m de altura y, por ende, susceptibles de haber sido objeto de corta desde entonces hasta la actualidad. El umbral definido para la fracción de cabida cubierta, por su parte, garantizaba que las masas clasificadas como bosque tuvieran una densidad de pies superior al 60%, evitando así la posterior clasificación como cortas de masas abiertas con relativa presencia de suelo desnudo.

Tabla 5. Parámetros del algoritmo “*Assign Class*” de eCognition.

Parámetro	<i>Assign Class</i>
Domain	Image object level
Threshold condition	Mean MDAV \geq 8; Mean LFCC \geq 60
Use class	Bosque
Parámetro	<i>Assign Class</i>
Domain	Image object level
Threshold condition	-
Class Filter	Unclassified
Use class	No Bosque

➤ Fusión de los segmentos de una misma clase: *Merge region*

Tras definir ambas clases, se procedió a agregar los segmentos generados en cada una de ellas. De nuevo, se empleó uno de los algoritmos disponibles, en este caso, *merge region*. Este algoritmo unifica los polígonos de la misma clase siempre y cuando exista continuidad espacial entre ellos. Como en el caso anterior, es necesario aplicarlo una vez a cada clase.

Tabla 6. Parámetros del algoritmo “*Merge region*” de eCognition.

Parámetro	<i>Merge region</i>
Domain	Image object level
Class filter	Bosque
Fusión super objects	Yes
Parámetro	<i>Merge region</i>
Domain	Image object level
Class filter	No Bosque
Fusión super objects	Yes

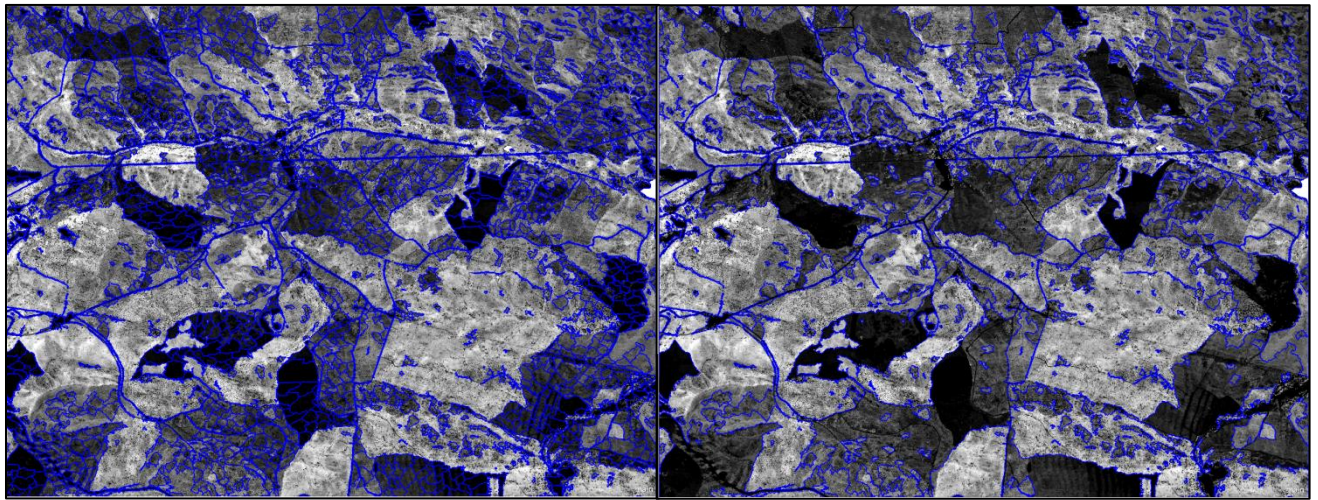


Figura 6. Efecto del algoritmo “*Merge region*” de eCognition: a la izquierda, fusión de la clase “Bosque”; a la derecha, fusión de la clase “No Bosque”

Este algoritmo, además de simplificar la trama de ambas clases, posibilita que, al establecer un segundo nivel de segmentación, los polígonos generados en el primero no condicionen la formación de los nuevos.

➤ Segmentación sobre segundo nivel: *Multiresolution segmentation*

A continuación, se procedió a generar un segundo nivel de segmentación mediante el algoritmo *multiresolution segmentation* pero, en esta ocasión, ponderando positivamente las capas de Sentinel en detrimento del MDAV. El reparto de los pesos sobre cada capa se estableció empíricamente, en base a la observación y análisis de la información disponible, buscando cual reflejaba en mayor medida aquellas características que se pretendían destacar, mediante numerosas pruebas ensayo-error. Puede resultar sorprendente el no haber incluido ninguno de los dos SWIR, cuyo rango de registro del espectro electromagnético detecta especialmente bien el suelo desnudo, entre las capas que participarán de esta segunda segmentación. La toma de esta decisión surgió del análisis empírico, en el que se comprobó que, tanto las capas del visible como el CP1 definían mejor las cortas a hecho, a través de un mayor contraste con los píxeles vecinos. Esto se debía a la mayor resolución espacial de las capas del visible y el CP1 -10m-, frente a los 20m de los SWIR. Por otro lado, si que se incluyó el NBR, que aporta información del SWIR2. Los parámetros de escala, forma y compacidad también sufrieron algún ligero cambio respecto a la primera segmentación en base a las mismas pruebas empíricas.

Tabla 7. Parámetros del algoritmo “*Multiresolution segmentation*” de eCognition.

Parámetro	<i>Multiresolution segmentation</i>
Domain	Pixel level
Level name	Second level
Image Layer weights	BLUE (5), CP1 (3), RED (5), NBR (5), resto (0)
Scale parameter	120
Shape	0,1
Compactness	0,5

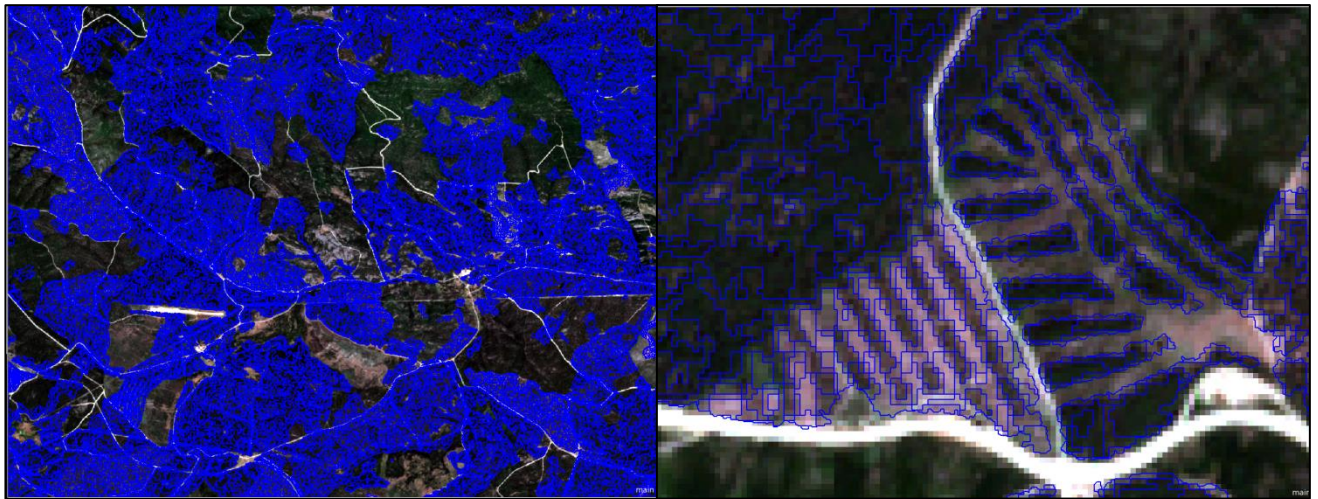


Figura 7. Segmentación generada por el algoritmo “*Multiresolution segmentation*” de eCognition sobre el primer nivel la derecha, ampliación sobre un sector.

Como se aprecia en la figura 6, esta segunda segmentación sí que logra representar las cortas a hecho posteriores al vuelo LiDAR, a través de múltiples objetos de pequeño tamaño que responden principalmente a la similitud espectral entre píxeles, tal y como se estableció en el parámetro forma.

Una vez se consideró positivo el resultado de la segmentación, se procedió a extender el arbol de reglas que recoge los algoritmos empleados en la segmentación al total del área de estudio, antes de exportarlo a formato shapefile (.shp) de ESRI para poder trasladar la segmentación al entorno de ArcGIS, desde donde se seleccionarían los puntos *verdad-terreno*.

- Exportar la segmentación a formato shape: *export vector layer*

Para exportar la segmentación, es preciso introducir un nuevo algoritmo al arbol de reglas de eCognition. En esta ocasión, el algoritmo empleado fue *export vector layer*, el cual permite exportar una segmentación vectorial, dotándola de una tabla de atributos configurable a partir de las capas de entrada, de las cuales el programa puede extraer una amplia gama de estadísticos, así como una serie de propiedades geométricas y contextuales propias del entorno eCognition. Este algoritmo también permite establecer umbrales que excluyan del proceso de la exportación aquellos objetos que no los cumplan. En este caso se estableció un umbral que excluía todo polígono con un valor medio de rojo inferior a 0, para así evitar la aparición de polígonos creados fuera del área de estudio con valores anómalos.

Tabla 8. Parámetros del algoritmo “*Export vector layer*” de eCognition.

Parámetro	<i>Export vector layer</i>
Domain	Image object level
Level	Second level
Threshold condition	Mean Red ≥ 0
Shape Type	Polygons
Attributes	Media y desviación estándar de las capas de entrada
	Propiedades de forma y medidas
	Diferencia con los vecinos (media y absoluta)

➤ Depurado de los datos

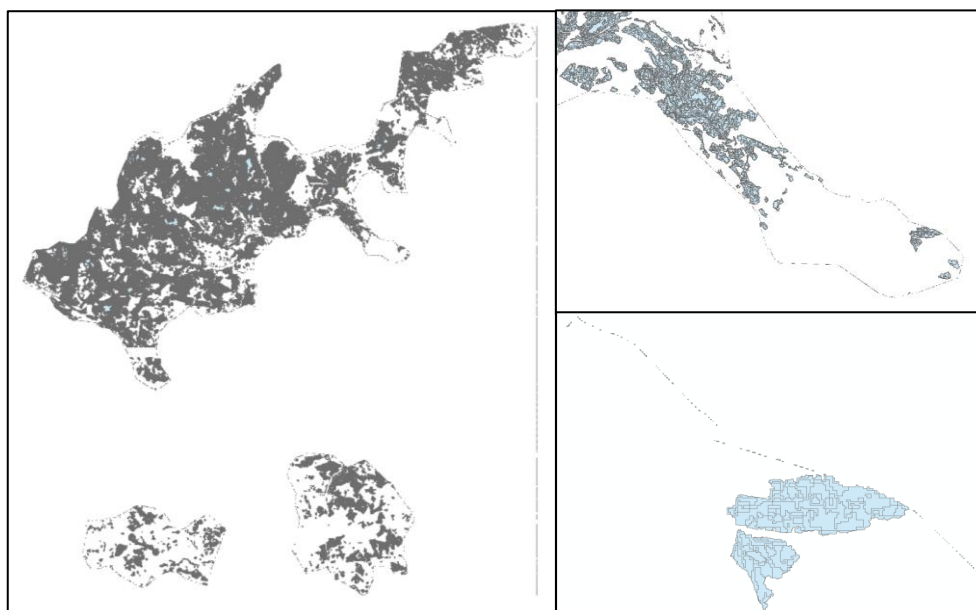


Figura 8. Vista de la capa vectorial extraída de eCognition. A la derecha, detalle de los polígonos residuales.

La capa vectorial generada en eCognition incluye una enorme cantidad de información, tanto espacial -segmentos- como temática, gracias a la tabla de atributos generada en con el algoritmo *export vector layer*. Esta información, tanto la espacial como la vectorial, es susceptible de presentar errores producto del complejo proceso de segmentación-exportación, por lo que es preciso analizarla en profundidad antes de proceder a la toma de datos *verdad-terreno*.

En primer lugar, se comprobaron los polígonos generados. Como se aprecia en la figura 7, se generó alrededor de todos los segmentos, a modo de perfil, una serie de polígonos de muy pequeño tamaño que no corresponden a ninguno de los segmentos creados sino, más bien, se trata de un residuo de la exportación. Igualmente aparece a la derecha de la imagen una línea conformada por otra serie de polígonos con no más de unos pocos píxeles de ancho. Se trata igualmente de un producto residual que se hubo de suprimir.

Para eliminar los polígonos que forman la línea a la derecha de la imagen, se seleccionaron manualmente y se suprimieron. Para corregir los polígonos residuales que perfilaban el área de estudio, se realizó una selección por atributos sobre el campo *Area_pxl* en el que se seleccionaron todos aquellos polígonos con un área inferior a 40 píxeles, tras haber realizado las comprobaciones empíricas necesarias para establecer ese umbral. En la figura 8, se puede apreciar la ubicación de los polígonos seleccionados, siempre coincidiendo con bordes, tanto interiores como exteriores, del área de estudio segmentada.

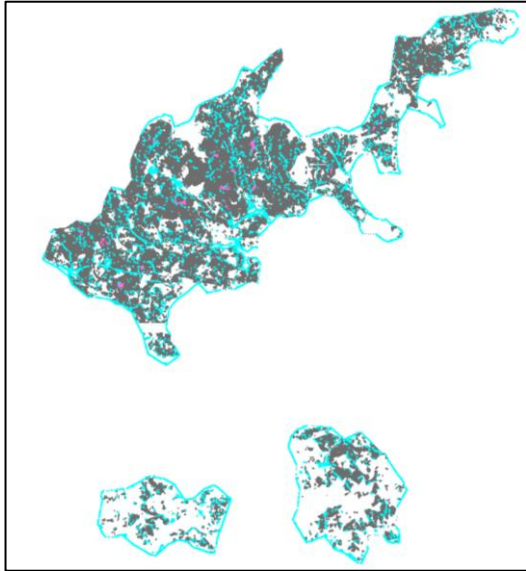


Figura 8. Selección de los polígonos residuales de eCognition.

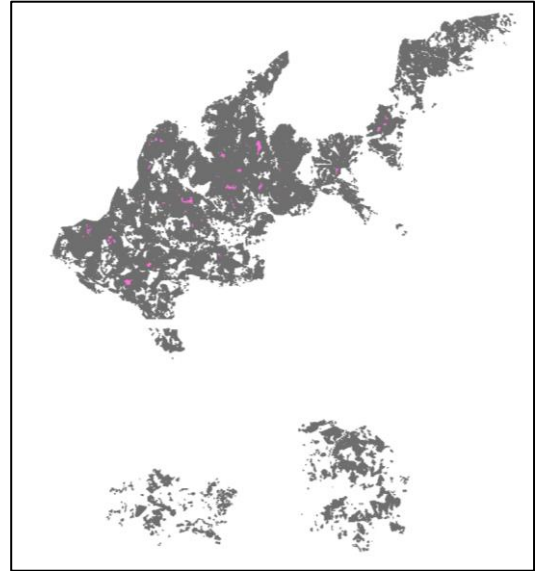


Figura 9. Capa vectorial de la segmentación después de eliminar los segmentos residuales.

En la figura 9, se puede ver como esos polígonos que perfilaban el área de estudio, así como la línea que aparecía al este, han desaparecido.

El siguiente paso consiste en comprobar los datos de la tabla de atributos ya que, en ocasiones, eCognition exporta datos que no se corresponden con lo esperable, debido a las transformaciones que sufren desde que parten a nivel de pixel hasta que se agregan en segmentos. Tras examinar los distintos campos de la tabla de atributos, se detectaron como corruptos los campos Canopy, NDVI, P95, Slope, Aspect, así como sus respectivas desviaciones estándar, por presentar valores anómalos de forma sistemática.

Para reparar estos datos se siguió el siguiente proceso. En primer lugar, se eliminaron los campos afectados de la tabla de atributos. A continuación, se implementaron al entorno ArcGIS 10.3 los respectivos rásters de los que surge la información. Después, mediante la herramienta *Zonal Statistics as Table*, del *Spatial Analyst*, empleando como *feature zone data* el shape de la segmentación, se fueron generando tablas que recogían la media y la desviación estándar de cada uno de los rásters. Finalmente se fue agregando esta información a la tabla de la segmentación mediante sucesivos *joins*.

➤ Errores en la clasificación de segmentos ocasionados por los datos LiDAR

Examinando detenidamente la segmentación obtenida, se identificaron una serie de segmentos que se correspondían sobre el terreno con roquedos y que aparecían dentro de la categoría de bosque. Se trataba, en concreto, de las zonas más escarpadas de los Picos de Urbión, en las que el relieve forma crestas y cortados muy agrestes. Esto ocasiona que, debido a los fuertes cambios de pendiente, el algoritmo empleado para generar el MDAV *-gridmetrics-*, confundiera estas zonas con árboles y las clasificara como tal.

Consecuentemente, cuando se estableció la clase de bosque en eCognition para aquellos segmentos con un MDAV superior a 8 m, estos segmentos pasaron el filtro. Dado que se trataba de pocos segmentos, el proceso para eliminarlos fue manual. En caso de tratarse de un número elevado de casos, se hubieran debido de identificar aplicando umbrales para tratar de aislar la respuesta espectral del roquedo.

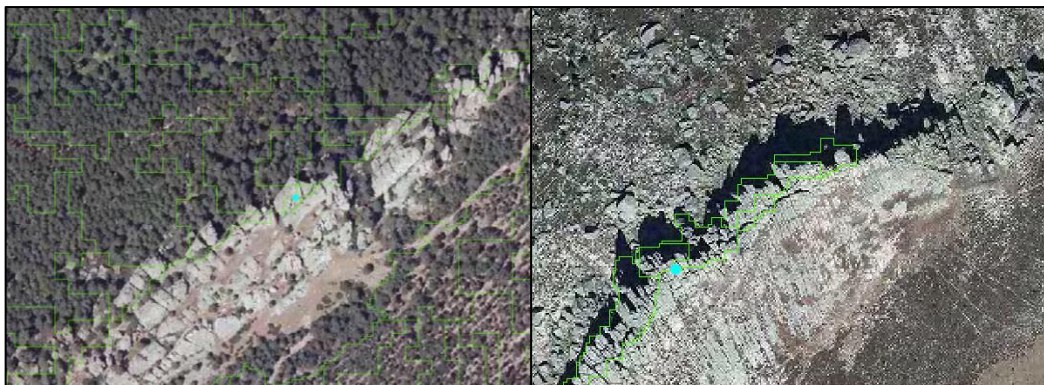


Figura 9. Vista de puntos incorrectamente clasificados por efecto de los datos LiDAR sobre el terreno abrupto.

3.3. Toma de datos *verdad-terreno*

Solucionado el problema de la depuración de datos, la siguiente fase consistía en seleccionar una serie de segmentos de naturaleza conocida -de ahí su denominación como *verdad-terreno*- para que sirvieran de subconjunto base sobre el que desarrollar el modelo clasificador con Random Forest. Estos datos debían estar acompañados de las mismas variables presentes en la tabla de atributos de la segmentación añadiendo un campo categórico que recogiera la categoría a la que pertenece ese segmento.

Como categorías a recoger, se definieron dos: “Cortas a hecho” y “Bosque”. Todos los puntos tomados se ubicaron sobre segmentos clasificados como “Bosque” en eCognition, descartando así caminos, superficies agrícolas, pastos, roquedos u otras superficies no susceptibles de albergar una corta a hecho. A pesar de esto, las zonas limítrofes -especialmente las que perfilaban la red de pistas forestales-, habiendo sido clasificadas como bosque, presentaban rasgos espectrales propios de las dos superficies, debido a la superficie del pixel Sentinel, pudiendo dar lugar a errores en la posterior clasificación. Para contrarrestar este efecto, cuando se tomaron los puntos *verdad-terreno* de la categoría “Bosque”, que agrupaba bosques de coníferas y frondosas, se incluyeron puntos de estos segmentos de bosques anexos a las pistas.

En total se seleccionaron 200 puntos *verdad terreno*, 100 de cada una de las categorías. De los 100 pertenecientes a la categoría “Bosque”, 20 correspondían a segmentos limítrofes a las pistas y los demás se repartían entre bosques de coníferas -*Pinus sylvestris* y *P. pinaster*- y de frondosas -*Quercus pyrenaica*, fundamentalmente-. En la categoría “Cortas a hecho” se seleccionaron puntos correspondientes a cortas a hecho de menos de dos años y de entre dos y seis años de antigüedad, valorando la posibilidad de operar con ellas por separado o de manera conjunta, como finalmente se hizo.

Tabla 9. Número de registros *verdad-terreno* por cobertura, clase (siendo 1, Cortas a hecho, y 2, Bosque) y sub-clase (siendo 1, cortas a hecho posteriores a 2014; 2, cortas a hecho de entre 2010 y 2014; 3, bosques de coníferas; 4, bosques de frondosas; y 5, segmentos contiguos a pistas forestales).

Cobertura	Clase	Sub-clase	Nº registros
Cortas a hecho posteriores a 2014	1	1	50
Cortas a hecho entre 2010 y 2014		2	50
Bosque de coníferas	2	3	40
Bosque de frondosas		4	40
Segmentos anexos a las pistas forestales		5	20

La muestra de puntos *verdad-terreno* se materializó sobre una capa vectorial de tipo poligonal en la que se incluyeron los registros en dos campos creados a tal efecto denominados “Clase”, que recogía la pertenencia a la categoría de “Cortas a hecho”, con el valor 1, y “Bosque” con el valor 2; y “Sub-clase” que distinguía entre cortas a hecho posteriores a 2014 -valor 1-, cortas a hecho de entre 2010 y 2014 -valor 2-, bosques de coníferas -valor 3-, bosques de frondosas -valor 4- y segmentos contiguos a las pistas forestales -con valor 5-. Debían seleccionarse segmentos cuya pertenencia a una u otra categoría no ofreciera dudas, tanto por la fiabilidad de la fuente, como por lo homogéneo del segmento. Para la posterior introducción en el entorno Random Forest, esta capa debía de contener, por un lado, un campo identificador compartido con la capa de la segmentación, para posteriormente poder aplicar el modelo sobre esta. Por otro lado, los campos “Clase” y “Sub-clase” con sus respectivas categorías contempladas. Por último, todos aquellos campos correspondientes a los estadísticos LiDAR, productos Sentinel y derivados y propiedades de forma y contexto extraídas del entorno eCognition. Preparada la capa, se hubo de exportar a (.txt), para su inclusión en randomForest.

Para seleccionar los segmentos se emplearon dos fuentes: el mosaico de las ortofotos y la imagen multibanda Sentinel 2. De manera que, si se observaba una corta a hecho en la ortofoto y no en Sentinel, se clasificaba en el campo “Sub-clase” como 2, ya que la ortofoto data de 2014, mientras que la imagen Sentinel es del 3 de julio de 2016. Lo mismo ocurría, a la inversa, con aquellas que solo aparecían en Sentinel, que eran clasificadas como sub-clase 1. Para clasificar los bosque se empleaban ambas fuentes de manera complementaria, valorando la mayor resolución de las ortofotos y la mayor actualidad de la multibanda Sentinel junto con sus propiedades espectrales, que permitieron discriminar entre coníferas y frondosas utilizando una composición de color 4-5-3 (NIR, SWIR1, RED).

Una vez se dispuso de los 200 puntos verdad terreno, se procedió a preparar dos subconjuntos destinados, el primero de ellos, con el 70% de los registros, a la generación del modelo, y el segundo de ellos, con el 30% restante, a la posterior validación de los resultados obtenidos. Para ello se utilizó la herramienta *Subset Features* del módulo *Geostatistical Analyst* de ArcGIS 10.3. Por lo tanto, para la generación del modelo se emplearon 140 puntos verdad terreno -70 de cada categoría-, y para la validación se reservaron 60.

3.4. Clasificación mediante Random Forest

Como señala Bourel (2012) el algoritmo Random Forest propuesto por Breiman, *combina las técnicas de CART y Bagging con el propósito de incorporar la aleatoriedad en las distintas etapas de la construcción de un árbol obtenido por CART.*

A efectos de este proyecto, se empleó Random Forest, dentro del entorno de R_Studio, para clasificar de manera automática los polígonos resultantes de la segmentación a partir de un modelo estadístico que el algoritmo desarrolla en base a un subconjunto, en este caso, la tabla de atributos *verdad-terreno* exportada a formato de texto (.txt). Para determinar a qué clase pertenece cada segmento, el algoritmo genera múltiples árboles de decisión definiendo umbrales de manera aleatoria sobre cada uno de los campos de la tabla y seleccionando los resultados más repetidos.

En el entorno de R_Studio, Random Forest se articula mediante un *script* de R. Para ello el primer paso es instalar e invocar las bibliotecas necesarias:

<pre>install.packages("randomForest", dep=TRUE) install.packages("party", dep=TRUE) install.packages("spdep", dep=TRUE) install.packages("ape", dep=TRUE) install.packages("caret", dep=TRUE) install.packages("dplyr", dep=TRUE) install.packages("sampling", dep=TRUE) install.packages("tree", dep=TRUE) install.packages("rfUtilities", dep=TRUE) install.packages("rgdal", dep=TRUE) install.packages("gmodels", dep=TRUE)</pre>	<pre>library(party) library(randomForest) library(spdep) library(ape) library(caret) library(dplyr) library(sampling) library(tree) library(rfUtilities) library(rgdal) library(gmodels)</pre>
---	--

Cuadro 1. Instalación e invocación de bibliotecas para Random Forest.

A continuación, debe de establecerse el directorio de trabajo y seleccionarse la información que servirá para generar el modelo, la tabla *verdad-terreno*.

<pre>setwd("D:/alberto_2016/Random Forest/mdav") datafile="VT_bosque2_70.txt" datos=read.table(datafile, header = TRUE, sep=",", dec=".") colnames(datos)</pre>
--

Cuadro 2. Definición de directorio principal, creación de objeto con los datos y visualización de nombres de las columnas.

En este caso, la tabla que contiene la información de la verdad terreno es “VT_bosque2_70.txt”, que se guarda en un objeto llamado “datafile” para, a continuación, guardarlo en otro de lectura llamado “datos”, con encabezado, que utiliza la coma como separador y el punto para delimitar los decimales. Con *colnames* se le ordenó que mostrase los nombres de las columnas:

[1]	"TARGET_FID"	"Area_Pxl"	"Asymmetry"	"Compactnes"	"Length_Pxl"	"Max_diff"	"BLUE"	"CP1"	"CP2"
[10]	"CP3"	"DNA_CP3"	"DNA_CP2"	"DN_BLUE"	"DN_CP1"	"DN_CP2"	"DNA_CP1"	"DNA_SWIR2_"	"DN_NBR"
[19]	"DN_NDVI"	"DN_NIR"	"DN_RED"	"DN_SWIR1"	"DN_SWIR2"	"DNA_BLUE"	"DNA_SWIR1"	"DNA_RED"	"DNA_NIR"
[28]	"DNA_NDVI"	"DNA_NBR"	"DN_GREEN"	"DNA_GREEN"	"DN_CP3"	"GREEN"	"LFCC"	"MDAV"	"NBR"
[37]	"NIR"	"RED"	"SWIR1"	"SWIR2"	"SD_CP1"	"SD_CP2"	"SD_BLUE"	"SD_LFCC"	"SD_MDAV"
[46]	"SD_NBR"	"SD_NIR"	"SD_RED"	"SD_SWIR1"	"SD_SWIR2"	"SD_GREEN"	"SD_CP3"	"Width_Pxl"	"Aspect"
[55]	"SD_Aspect"	"Slope"	"SD_Slope"	"CANOPY"	"SD_CANOPY"	"P95"	"SD_P95"	"NDVI"	"SD_NDVI"
[64]	"CLASE"	"SUB-CLASE"							

Cuadro 3. Lista completa de las variables.

En el cuadro 3, se muestran los nombres de las columnas de la tabla: el identificador “TARGET_FID”; el campo “CLASE”, que ejercería de variable dependiente; los campos con propiedades geométricas tales como “Area_Pxl” o “Asymmetry”; los extraídos directamente de Sentinel, como “BLUE” o “CP1”; los estadísticos LiDAR, como “LFCC” o “MDAV”; las respectivas desviaciones estándar de unos y otros, precedidas por el prefijo “SD_”; y, finalmente, las propiedades contextuales aportadas por eCognition, precedidas por “DN_” cuando hacen referencia a la diferencia entre vecinos -difference neighbour-, y por “DNA_” cuando hacen referencia a la diferencia absoluta entre vecinos.

Después de realizar varias pruebas, se llegó a la conclusión de que el modelo funcionaba mejor con los datos binarios del campo “Clase” que con las 5 sub-clases. Por tanto, la siguiente sentencia definía el campo “CLASE” de la tabla como un factor, esto es, le daba carácter categórico para poder emplearlo como variable dependiente en el modelo de clasificación. Con el comando *summary* se le indicaba que resumiera el campo recién creado:

```
datos$CLASE<-as.factor(datos$CLASE)
summary(datos$CLASE)
```

Cuadro 4. Definición del campo “CLASE” como factor y resumen de este.

El resultado que ofreció indica que contabilizó un total de 70 registros para cada una de las categorías. Este dato es sumamente relevante para definir el tamaño de la muestra a emplear por el modelo.

```
> summary(datos$CLASE)
 1  2
70 70
```

Cuadro 5. Resumen del campo “CLASE”.

A continuación, se introdujo la fórmula del algoritmo Random Forest, de la que resultaría el modelo clasificador, que se registró en el objeto “rf_Segmentacion”. El primer parámetro de la fórmula establecía como variable dependiente el campo “CLASE” de la tabla guardada en el objeto “datos”. El siguiente parámetro, recogido en el objeto “data” determinaba cuáles de las variables independientes o explicativas participaron del modelo clasificador. La selección de estas variables no es ni automática ni aleatoria, sino que depende de un proceso que se explicará más adelante. El siguiente parámetro, *ntree*, determina el número de árboles aleatorios que genera el algoritmo para definir el modelo. Teóricamente, a mayor número de árboles más contrastado estará el modelo, si bien llega un momento en el que el aumento resulta improductivo, conllevando además un aumento considerable del tiempo de procesado. Los siguientes parámetros, *importance* y *replace*, determinaban que el modelo registrara la importancia de las variables y que se pudiera sobrescribir muestras, respectivamente. El parámetro *nodesize*, por su parte, estableció el tamaño mínimo de los nodos terminales. El siguiente parámetro,

```
rf_Segmentacion<-randomForest(datos$CLASE~ ., data=datos[,c(7,10,33,36,38,39,40)],
ntree=10000, importance=T, replace=TRUE, nodesize=4, strata=datos$CLASE,
sampsiz=c(70,70))
```

Cuadro 6. Fórmula de Random Forest.

strata, concretó el campo empleado para estratificar la muestra. Finalmente, *sampsiz*, decretaba el número de registros de cada clase en la variable dependiente que se utilizarían para generar el modelo. Es recomendable que todas las categorías tengan similar número de registros a este fin. En este caso se establecieron 70 registros para cada clase gracias a que, previamente, se realizó un resumen de los datos del campo “CLASE” que informó del número de registros de cada categoría presente en el campo.

Acompañando a la fórmula del algoritmo, se dispusieron una serie de sentencias encaminadas a facilitar la selección de las variables sobre las que se generaría el modelo:

```
varImpPlot(rf_Segmentacion) imp<-importance(rf_Segmentacion)
colnames(imp) I = importance(rf_Segmentacion)[,3]
MIR = I/max(I)
sort(MIR)
colnames(datos)
```

Cuadro 7. Sentencias para clasificar las variables explicativas según su importancia.

La primera, creó un gráfico con la importancia de cada una de las variables incluidas en la fórmula para generar el modelo (Figura 11). La segunda, creó a su vez un objeto que recogía la importancia de las variables y determinaba que método de clasificación se emplearía de los dos que ofrece el algoritmo: "MeanDecreaseAccuracy" y "MeanDecreaseGini". La tercera, crea un objeto que estandariza esta importancia respecto al valor máximo, es decir, la transforma en un índice. A continuación, se indica que ordene de menor a mayor la importancia de las variables introducidas. Este paso es determinante a la hora de seleccionar las variables puesto que, el proceso de selección-descarte funciona de la siguiente manera: La primera vez que se ejecuta el algoritmo, se incluyen en el objeto "data" todas las variables de las que se disponen. El algoritmo desarrollará un modelo con un error determinado, producto de un proceso de validación cruzada que lleva implementado. Cuando se obtienen los datos de importancia ordenados, se han de ir suprimiendo aquellas variables que queden en la cola de la lista, es decir, las que menos aporten, bien sea por irrelevantes o por redundantes. Este proceso de descarte continua, ejecutando el algoritmo sucesivas veces, mientras que el error dado disminuya o se mantenga igual, ya que se parte de la premisa de que, a igual error, cuantas menos variables influyan en el modelo, más estable será este.

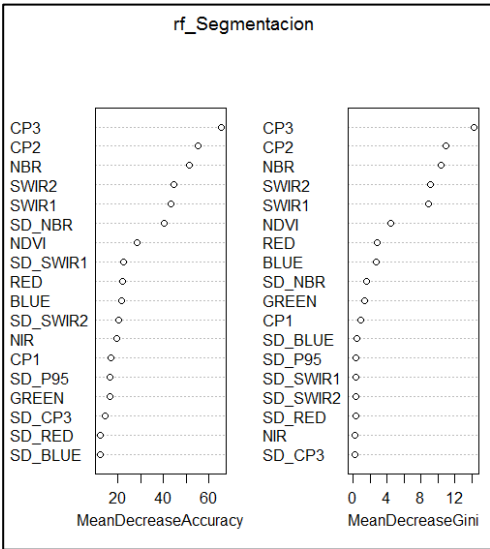


Figura 10. Gráfico de importancia de las variables

En el cuadro 8, aparecen parte de las variables ordenadas de menor importancia a mayor. Se trata de una de las fases iniciales, en las que apenas se habían descartado variables todavía. En esta fase del proceso, las variables a descartar serían "DN_SWIR2" primero, "DNA_SWIR1" segundo, "SD_NIR" después, y así sucesivamente, en un proceso empírico de ensayo error hasta ajustar el modelo. Para quitar una variable de la fórmula, se debe de identificar su orden en la tabla "datos" y quitarlo del objeto "data" en la ecuación.

```
> sort(MIR) #ordena las variable segun su importancia, de menos a mas.
DN_SWIR2 DNA_SWIR1 SD_NIR SD_CP1 Slope SD_Slope LFCC SD_CP2 DNA_NBR SD_GREEN
0.08684735 0.08734635 0.09882167 0.10035002 0.10639195 0.11130403 0.11311790 0.11580558 0.12446858 0.13957208
DNA_CP1 DNA_GREEN NDVI DN_NBR DN_NDVI Area_Pxl DNA_SWIR2 NIR SD_BLUE DNA_BLUE
0.14098649 0.14762923 0.16143319 0.16708879 0.17422098 0.17425178 0.18667735 0.19419483 0.21164245 0.21649005
SD_SWIR1 Max_diff Asymmetry Width_Pxl SD_SWIR2 DNA_RED SD_RED SD_CP3 SD_NBR CP1
0.22717012 0.22731994 0.22957893 0.24365128 0.27043339 0.27761531 0.29642811 0.33893397 0.34981650 0.44508968
CP2 CP3 NBR SWIR1 GREEN BLUE RED SWIR2
0.47801992 0.65055606 0.76441407 0.76743022 0.83797116 0.89806451 0.90088948 1.00000000
```

Cuadro 8. Variables explicativas en fase inicial del proceso de ajuste-descarte.

Para facilitar este proceso, se introdujo, justo después de la sentencia que ordenaba las variables según su importancia, el comando *colnames* para la tabla "datos", de manera que se pudiera ver la posición de cada variable en la tabla.

Las siguientes sentencias suponen la creación de una nueva tabla, a partir de la tabla contenida en el objeto “datos”, con un campo llamado “pred_rf_Segmentacion”, de tipo factor, en el que figuraría la categoría en la que el modelo clasificaría cada segmento de la tabla original. De esta manera, exportando esta tabla al entorno ArcGIS y vinculándola a la *verdad-terreno*, a través de una selección por atributos, es posible comparar en qué medida coinciden las predicciones del modelo con la verdad conocida y, así, obtener una visión más fiable del error logrado en el modelo.

```
rf_Segmentacion
datos$pred_rf_Segmentacion<-
as.factor(predict(rf_Segmentacion,data="datos",type="response"))
write.table(datos, file="tabla_VT_predicciones.txt", row.names=FALSE, col.names=TRUE,
sep="\t", dec=".")
```

Cuadro 9. Creación del campo "pred_rf_Segmentacion" y creación de la tabla con las predicciones.

Para generar un diagrama cruzado (figura 12), similar a una matriz de confusión, en el que comprobar que categorías tienden a confundirse con otras, se incluyó esta sentencia:

```
Diagrama<-CrossTable(datos$CLASE,datos$pred_rf_Segmentacion)
```

Cuadro 10. Generación del diagrama cruzado

Tras realizar todo el proceso explicado, ir descartando variables una a una hasta que el error del modelo subió, lo que supone que se eliminó una variable explicativa relevante que debía ser considerada, se consiguió ajustar el modelo logrando un error OOB del 1,43%.

En la tabla 19 se muestran las variables que integraron el modelo una vez quedó ajustado, junto con su índice MIR de importancia relativa. La mayor relevancia la obtuvo el CP3, seguido del CP2. Estas dos variables sintetizan buena parte de la información espectral de Sentinel, pudiendo relacionarse con la humedad -CP3- y el verdor -CP2-, y ofrecen la máxima resolución espacial de Sentinel. A continuación, aparece el índice NBR, ambos SWIR y la desviación estándar del NBR. Estas cuatro variables poseen la capacidad de detectar el suelo desnudo, que presenta sus mayores valores de reflectividad en la región del espectro electromagnético propia del infrarrojo de onda corta -SWIR-. La explicación al buen funcionamiento del NBR responde a que es un índice diseñado para detectar zonas afectadas por incendios forestales en gran medida similares a las cortas a hecho. Además, la fórmula que lo define incluye el SWIR. Finalmente, ya por debajo del umbral del 0.5 de importancia relativa, aparecen el NDVI, las desviaciones estándar de los SWIR, las bandas del visible, el NIR, el CP1 y las desviaciones estándar de las bandas del visible y del CP3.

Tabla 10. Diagrama cruzado entre observaciones y predicciones.

Total Observations in Table: 140			
datos\$CLASE	datos\$pred_rf_Segmentacion		Row Total
	1	2	
1	68	2	70
	34.000	32.111	
	0.971	0.029	0.500
	1.000	0.028	
	0.486	0.014	
2	0	70	70
	34.000	32.111	
	0.000	1.000	0.500
	0.000	0.972	
	0.000	0.500	
Column Total	68	72	140
	0.486	0.514	

Por el contrario, las variables que menos determinantes resultaron y que, por ende, fueron descartadas primero por el modelo, fueron las propias de eCognition, relativas a las diferencias entre vecinos, normal (DN) y absoluta (DNA) y a la forma y tamaño -Asymmetry, Lenght, Area_pxl...-. Posteriormente el modelo también descartó las propias de los estadísticos LiDAR, comenzando por sus desviaciones estándar y siguiendo con sus valores promedio por segmento. El motivo de la escasa relevancia

de los productos LiDAR responde a su obsolescencia, patente al haber recabado la verdad-terreno sobre la información actualizada de Sentinel.

Tabla 11. Variables explicativas del modelo con su índice MIR ordenadas de menor a mayor relevancia.

SD_BLUE	SD_RED	SD_CP3	GREEN	SD_P95	CP1	NIR	SD_SWIR2	BLUE
0.1888189	0.1921018	0.2223860	0.2539184	0.2557144	0.2590684	0.2990612	0.3146711	0.3309965
RED	SD_SWIR1	NDVI	SD_NBR	SWIR1	SWIR2	NBR	CP2	CP3
0.3412591	0.3470795	0.4343869	0.6175872	0.6643653	0.6813230	0.7837518	0.8474385	1.0000000

Una vez generado el modelo, se había de extender al total de la muestra, es decir, a la segmentación completa. Igual que anteriormente, el primer paso consistía en crear un nuevo objeto con la tabla de la segmentación y asignarlo a su vez a un objeto de lectura “tabla_segmentacion”.

```
datafile2="Seg_bosque2.txt"
tabla_segmentacion=read.table(datafile2, header = TRUE, sep=",", dec=".")
colnames(tabla_segmentacion)
```

Cuadro 11. Creación de un objeto de lectura a partir de la tabla de la segmentación

Una vez que se dispuso de la tabla de la segmentación en un objeto de lectura, se procedió a ejecutar el modelo. Para ello, se creó un nuevo campo en el objeto “tabla_segmentacion” llamado “rf_Segmentacion”, que se rellenaría con la predicción categórica del modelo en base a las dos clases antes previstas:

```
tabla_segmentacion$rf_Segmentacion<-
predict(rf_Segmentacion,tabla_segmentacion,type="response")
summary(tabla_segmentacion$rf_Segmentacion)
```

Cuadro 12. Creación del campo rf_Segmentacion y resumen.

A continuación, se definió un objeto en el que se guardó una tabla con la predicción en forma de probabilidad de que cada registro pertenezca a una categoría, para crear en él dos nuevos campos con la probabilidad de que un registro pertenezca a la categoría 1 o la 2. Finalmente se unificaron las dos tablas mediante el comando *cbind*, y se generó un nuevo archivo de texto, a través del comando *write.table*, con el resultado final para poder vincularlo, desde ArcGIS, a la capa vectorial de la segmentación y así visualizar el resultado de la clasificación.

```
tabla_segmentacion_prob<-as.data.frame(predict(rf_Segmentacion,tabla_segmentacion,type="prob"))
colnames(tabla_segmentacion_prob) <- c("pred_rf_Segmentacion_prob0","pred_rf_Segmentacion_prob1")
tabla_segmentacion_def<-cbind(tabla_segmentacion,tabla_segmentacion_prob)
write.table(tabla_segmentacion_def, file="clasificacion_pendiente2.txt", row.names=FALSE,
col.names=TRUE, sep="\t", dec
```

Cuadro 13. Generación del resultado final fusionando "tabla_segmentacion" con "tabla_segmentacion_prob"

4. RESULTADOS

4.1. Análisis del resultado final

Una vez se obtuvo la clasificación para toda la segmentación a partir del modelo de Random Forest, se vinculó mediante un *join* a la capa vectorial de la segmentación en ArcGIS para poder visualizar el resultado.

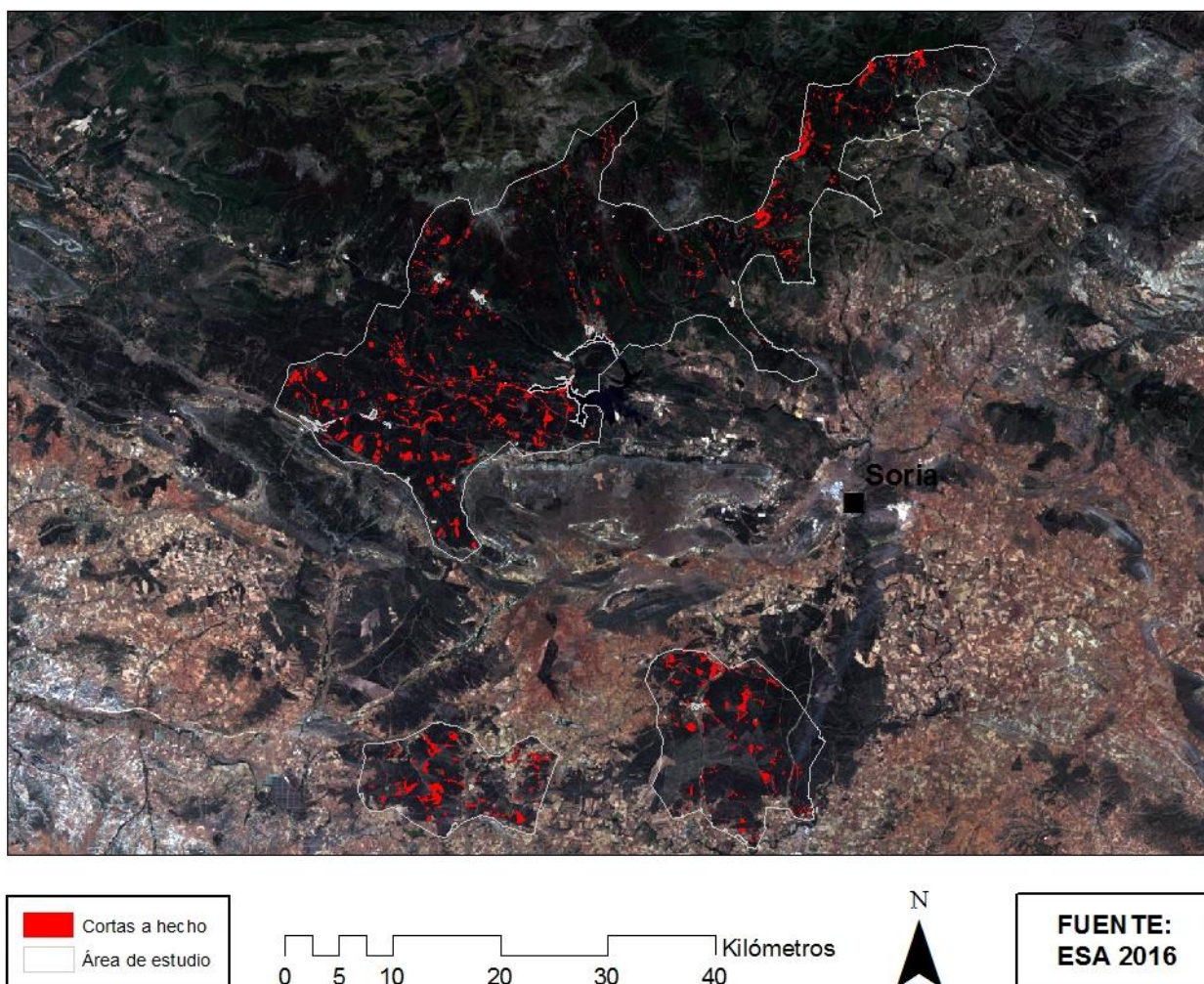


Figura 11. Cartografía de cortas a hecho.

Como puede apreciarse en la figura 11, el modelo permite cartografiar las cortas a hecho con un resultado razonablemente satisfactorio, aunque aparecen una serie de polígonos representados como cortas a hecho que no lo son en la realidad, generando “ruido”. Este ruido se compone, por una parte, de una serie de polígonos residuales tanto por su tamaño como por su distribución, fácilmente eliminables aplicando un filtro por superficie -que no se aplicará aquí para no retocar el resultado-; y por otra, de segmentos anexos a los caminos que, por encontrarse en la zona de transición entre el bosque y el camino, presenta una respuesta espectral con características similares a la de las cortas a hecho. Estos segmentos aparecen a pesar de haber sido incluidos entre los puntos de verdad-terreno correspondientes a la clase “Bosque”, si bien se trata de casos puntuales.

➤ Validación del resultado

Para validar cuantitativamente el resultado obtenido se generaron dos matrices de confusión a partir del subconjunto de la *verdad-terreno* reservado a este fin. Una sobre píxeles, desde ArcGIS y la otra sobre segmentos, en R.

La primera matriz se generó en el entorno de ArcGIS 10.3, empleando las capas de la *verdad-terreno* y de la segmentación, una vez se vinculó al resultado de la clasificación del modelo. Ambas capas se pasaron a ráster mediante la herramienta *Features to Ráster* del módulo *Conversion Tools*. A continuación, se realizó un *Combine* desde el *Spatial Analyst* y el ráster resultante se exportó al entorno de Excel para calcular la exactitud del usuario y del productor y sus respectivos errores de comisión y de omisión.

Tabla 12. Matriz de confusión a nivel de pixel.

	MATRIZ DE CONFUSIÓN (PÍXEL)	Referencias			Exactitud del usuario	Error de Comisión
		1: Corta	2: Bosque	Total		
Clasificación	1: Corta	1326	224	1550	85,55	14,45
	2: Bosque	182	5882	6064	97,00	3,00
	Total	1508	6106	7614	Fiabilidad Global: 94.66%	
	Exactitud del productor	87,93	96,33			
	Error de Omisión	12,07	3,67			

El resultado obtenido es razonablemente positivo. La fiabilidad global del modelo, es decir, la relación entre el número de valores bien clasificados y el total, asciende hasta el 94,66%. La exactitud del usuario para la clase “Cortas a hecho” alcanza el 85,5% por lo que su complementario error de comisión se sitúa en el 14,45%. Esto supone que un 14,45% de los píxeles pertenecientes a la clase “Cortas a hecho” han sido clasificados como bosque. Para la clase “Bosque”, por su parte, el error de comisión es del 3%, es decir, un reducido porcentaje de píxeles correspondientes a esta categoría han sido clasificados como cortas. La exactitud del productor, complementaria al error de omisión, alcanza para la categoría de “Cortas” el 87,9% de exactitud, lo que significa que el 12% de los píxeles de las referencias de la categoría “Cortas” no están clasificados como tales. Finalmente, para la clase “Bosque”, se sitúa en el 96,3%, por lo que el 3,6% de las referencias de “Bosque” aparecen clasificadas como cortas.

Para obtener el estadístico *Kappa*, que estima la aleatoriedad y la corrige de la fiabilidad global del modelo, se recurrió al software estadístico IBM SPSS 22. Previamente hubo de convertir la información de los segmentos a nivel de ráster y organizarla en una tabla con dos campos: referencias y clasificación. Para ello se generó una malla poligonal en la que cada celda tuviera la misma superficie que el pixel Sentinel. Después se calculó el centroide de cada polígono y finalmente se extrajo el valor de los rásters con los valores observados -referencias- y clasificados. IBM SPSS 22 devolvió el siguiente resultado:

Tabla 13. Medidas simétricas obtenidas de IBM SPSS 22

	Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Medida de acuerdo Kappa	,834	,009	70,415	,000
N de casos válidos	7091			

Donde el valor del estadístico *Kappa* es de 0,834

La segunda matriz se realizó a nivel de segmentos en el entorno de R, por lo que el número de registros que considera es el mismo que el del subconjunto de segmentos verdad-terreno destinados a la validación, es decir 60.

Tabla 14. Matriz de confusión a nivel de segmentos.

	MATRIZ DE CONFUSIÓN (SEGMENTO)	Referencias			Exactitud del usuario	Error de Comisión
		1: Corta	2: Bosque	Total		
Clasificación	1: Corta	24	2	26	92,31	7,69
	2: Bosque	6	28	34	82,35	17,65
	Total	30	30	60	Fiabilidad global: 86.66%	
	Exactitud del productor	80	93,33			
	Error de Omisión	20	6,66			

Como puede apreciarse, la fiabilidad global del modelo ha descendido desde el 94,66% hasta el 86,66%. Por su parte, los errores, tanto de comisión como de omisión, han variado respecto a la matriz orientada al pixel. Así, el error de comisión para las cortas ha disminuido desde el 14,4% hasta el 7,6%, mientras que para la clase “Bosque” se ha incrementado del 3% al 17,64%. El error de omisión para la clase “Cortas” ha aumentado desde el 12% hasta el 20%, y el de la clase “Bosque” ha pasado del 3,6% al 6,6%.

Igual que con la matriz con datos al pixel, se calculó el estadístico *Kappa* para comprobar la fiabilidad del modelo atendiendo a la concordancia atribuible al azar.

Tabla 15. Medidas simétricas obtenidas de IBM SPSS 22

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Medida de acuerdo	Kappa	,733	,087	5,732	,000
N de casos válidos		60			

Donde el valor del estadístico *Kappa* es de 0,733

4.2. Discusión de la utilidad del modelo

En el punto anterior, se ha cuantificado la fiabilidad del modelo, así como también la dimensión de los errores de comisión y omisión o el estadístico *Kappa*. Los resultados obtenidos son relativamente satisfactorios, especialmente al comprobar sobre el terreno, bien sea *in situ*, bien sea a través de un SIG, la capacidad del modelo para cartografiar cortas a hecho. Si la validación no arroja mejores resultados se debe principalmente al ruido que genera el modelo, a todos esos segmentos clasificados como cortas que no lo son en realidad. Dentro de estos segmentos se puede distinguir entre dos tipologías: por un lado, segmentos aleatoriamente distribuidos por el área de estudio, cuya respuesta espectral se acerca a la de las cortas a hecho, de muy pequeño tamaño, por lo que son fácilmente descartables aplicando un filtro por superficie. Por el otro lado, están los segmentos anexos a los caminos -ya que los caminos se suprimieron al segmentar definiendo un umbral sobre el MDAV- cuya respuesta espectral aún la de los bosques y la de los caminos. Estos segmentos aparecen clasificados como cortas aun después de haber sido incluidos en la *verdad-terreno* de la clase “Bosque”, si bien solo ocurre en casos

puntuales. En ambos casos, el motivo es el mismo, la similitud entre la respuesta espectral del suelo desnudo visible en las cortas a hecho y la del suelo desnudo de caminos u otras superficies sin cobertura arbórea.

Ante esta situación surge una limitación técnica de Sentinel. Distinguir entre dos superficies cuyo comportamiento espectral es análogo, como son las cortas a hecho -claros artificiales en una masa boscosa con presencia de suelo desnudo y, en muchos casos, de vegetación incipiente-, y las zonas de transición entre caminos y bosques, que igualmente presentan estos dos elementos, es complicado, quizás con mayor resolución espectral fuera posible, y probablemente una mayor resolución espacial reduciría la superficie de los píxeles con mezcla espectral dentro de un objeto segmentado y, por lo tanto, su peso específico dentro de éste. De esta forma, Random Forest tendría menos dificultad a la hora de discernir entre dichas superficies.

En este sentido, contar con datos LiDAR actualizados facilitaría en gran medida la detección de las cortas y la diferenciación entre superficies similares espectralmente pero no así en su caracterización tridimensional sobre la nube de puntos. De la misma manera que contar con una secuencia de imágenes de distintas fechas con la suficiente diferencia temporal entre ellas permitiría realizar un estudio de cambios en la cobertura del suelo en base, como se ha hecho con anterioridad en múltiples estudios, al NDVI o, para el caso que nos ocupa, tal vez al NBR. Cualquiera de estas alternativas resultaría eficaz y de fácil aplicación. Sin embargo, actualmente ambas son inviables mediante información geográfica de libre disposición, ya que los datos del PNOA datan de 2010, y Sentinel 2A lleva registrando desde junio de 2015, por lo que no permite realizar estudios multitemporales.

Ante esta coyuntura, la metodología aquí propuesta permite solucionar este problema de manera eficiente, aun con la limitación especificada, resultando además sencillo mejorar el resultado, como se explicó con anterioridad, mediante un sencillo filtrado por superficie.

4.3. Vías descartadas

Hasta el momento, se ha descrito el proceso seguido hasta llegar al resultado final, sin incluir todas aquellas vías que, por no ofrecer los resultados deseados o, por no conseguir solucionar el problema para el que se planteaban, quedaban descartadas. Sin embargo, son precisamente estas, las que a través de la praxis empírica del ensayo-error, particularmente a través del error, iban señalando el camino a seguir. Por esta razón, en este apartado se detallarán alguna de estas *vías muertas* que, indirectamente, han contribuido a la resolución del objetivo principal.

Hasta la conclusión de la fase de “Procesado de datos” no hubo necesidad de plantearse alternativas puesto que no surgieron problemas que las requirieran. Sin embargo, una vez se realizó el primer modelo clasificador y se vinculó el resultado a la segmentación para su visualización cartográfica, se comprobó que, a pesar de la alta fiabilidad que ofrecía el modelo, confundía sistemáticamente las cortas a hecho con otras superficies, desde superficies agrícolas, prados, roquedos y caminos hasta bosques con baja densidad de pies y alta presencia de suelo desnudo en su respuesta espectral.

Para solucionar esto, se optó, en un primer momento, por descartar la idea inicial de recoger 5 categorías -sub-clases- en la *verdad-terreno*, agrupándolas en solo dos, “Cortas a hecho” y “Bosque”, considerando que, probablemente, un modelo dicotómico más simple mejoraría el resultado. Esta medida resultó moderadamente efectiva, por lo que la idea inicial quedó desechada.

Aunque el error del modelo mejoró, al extenderlo al total de la muestra seguía presentando errores de clasificación sistemáticos por lo que la siguiente opción pasó por realizar un filtrado en la fase de la “Segmentación” a partir del percentil 95. De manera que se estableció un umbral que dejara fuera de la capa vectorial resultante de la segmentación todos aquellos segmentos cuyo valor promedio del P95

para la fecha de los datos LiDAR -2010- fuera inferior a 10 metros, por considerarse que ninguna masa forestal con un P95 inferior sería susceptible de ser cortada en los siguientes 6 años. Con esta medida se pretendía descartar, además de masas jóvenes, superficies agrícolas, prados y, especialmente, caminos.

Este filtro, aunque bien planteado, no resolvió el problema, entre otras superficies, el modelo clasificaba como cortas bosques con baja densidad de pies, por su elevado contenido de suelo desnudo, lo que provocó que se empleara como umbral el MDAV junto con el LFCC, para garantizar una densidad mínima de pies, además de una altura media.

Otra fórmula que se barajó consistía en realizar una primera clasificación a partir de la capa vectorial producto de la segmentación filtrada con el P95. Del resultado obtenido se extrajeron aquellos segmentos clasificados como cortas a hecho y, sobre ellos, se tomó una nueva muestra de puntos *verdad-terreno* recogiendo dos categorías, aquellos puntos que realmente se correspondían con cortas a hecho sobre el terreno, por un lado, y por el otro todos los demás, que representaban el error en la clasificación. Sobre esta *verdad-terreno* y sobre esta capa que incluía los segmentos bien y mal clasificados como cortas se corrió un nuevo Random Forest con la esperanza de que, en esa ocasión, posibilitara la diferenciación adecuada entre categorías. Sin embargo, el resultado no fue el esperado ya que el error se disparó a valores por encima del 20%.

Otra de las vías que se exploró fue la de realizar una regresión logística binomial en lugar de una clasificación con Random Forest, pero se prescindió de esta posibilidad ya que tendía a sobreestimar el resultado.

Aunque ninguna de estas alternativas participó en el proceso metodológico final, todas han permitido mejorar la comprensión del problema a resolver, a la par que han supuesto una dedicación de tiempo y esfuerzo, por lo que se consideró conveniente reflejarlas en el documento final.

5. CONCLUSIONES

Como resultado del presente proceso metodológico, es posible concluir que la cartografía de cortas a hecho a través de una sola imagen Sentinel 2, contando con datos LiDAR PNOA como información auxiliar, es factible con un resultado moderadamente positivo.

Al realizar una valoración cuantitativa a nivel de píxel, mediante una matriz de confusión, se obtuvo una fiabilidad global del modelo del 94,66%, un error de comisión para la clase “Corta” del 14,45% y para la clase “Bosque”, del 3%; y un error de omisión para las cortas del 12,07% y para los bosques, del 3,67%. Por su parte, el estadístico *Kappa* se situó en 0,834 sobre 1.

La validación a nivel de segmentos arrojó unos resultados algo peores, en términos generales. La fiabilidad global del modelo alcanzó el 86,66%, el error de comisión para la clase “Corta” el 7,69% y para la clase “Bosque” el 17,65%. El error de omisión se situó, para la primera, en el 20%, y para la segunda en el 6,66%. El estadístico *Kappa* alcanzó el 0,733 sobre 1.

Las mayores dificultades que encuentra el modelo a la hora de clasificar, responden a la similitud espectral entre segmentos de cortas a hecho y segmentos cuya respuesta espectral quede constituida por varias superficies, con un peso relativo considerable de suelo desnudo.

Para solucionar esta cuestión, se realizó, en un primer momento, una criba sobre las superficies que no cumplían con las características necesarias para soportar una corta a hecho. Para ello se clasificaron como “No Bosque”, en el entorno de eCognition, aquellos segmentos resultantes de la segmentación inicial, en base al MDAV obtenido de los datos LiDAR PNOA de 2010, cuya fracción de cubierta fuera inferior al 60%, para así descartar masas boscosas de baja densidad. Junto a esta primera

condición, se estableció otra que dejaba fuera del siguiente nivel de segmentación aquellos segmentos cuyo valor medio del MDAV fuera inferior a 8 en 2010.

A pesar de acotar los segmentos susceptibles de albergar una corta a hecho, seguían apareciendo segmentos que, pese a cumplir con las condiciones establecidas, eran incorrectamente clasificados por el modelo. Principalmente se trataba de segmentos anexos a las pistas forestales cuya respuesta espectral incluía suelo desnudo y vegetación. Para tratar de favorecer la correcta clasificación de los mismos, se incluyeron segmentos con estas condiciones en la *verdad-terreno* de la clase “Bosque”, buscando incluir su variabilidad espectral en esta. Resultó relativamente efectivo, ya que la mayoría de estos segmentos quedaron correctamente clasificados, si bien aún persistían una serie de segmentos de transición entre bosque y pista clasificados como cortas a hecho.

Una alternativa metodológica, tal vez una vía futurible, hubiera sido, en la segmentación, favorecer las formas elongadas en detrimento de las compactas mediante el parámetro *Compactness* que quedó desactivado al darle un valor neutro (0,5). Si en vez de optar por esta parametrización, que fue la que dio mejores resultados tras múltiples pruebas realizadas, se hubieran potenciado formas alargadas tratando de ajustarse mejor a las fajas de las cortas a hecho, es posible que las propiedades de forma-contexto que aportó eCognition a la tabla de atributos de la segmentación hubieran dado mejor resultado en el posterior proceso de ajuste y descarte en Random Forest.

Respecto al cumplimiento de los objetivos del proyecto, puede afirmarse que el objetivo principal de proponer una metodología que permita cartografiar cortas a hecho en masas de coníferas, en base a una imagen Sentinel con el apoyo de datos LiDAR PNOA, ha sido cumplimentado con un grado de satisfacción considerable, a pesar de dejar alguna puerta abierta a explorar otras vías en futuras investigaciones.

Los objetivos específicos, que giraban en torno a la integración y adaptación del proyecto y del alumno del entorno académico al profesional, se pueden dar por correctamente cumplimentados. En concreto, el amoldarse al ámbito forestal e integrar sus conceptos y técnicas con los propios de las tecnologías de la información geográfica. Otro objetivo específico satisfecho fue el recorrer las diferentes fases propias de un proyecto técnico de teledetección, como son la adquisición, corrección y composición, de imágenes o datos brutos, su transformación para generar productos derivados, como los estadísticos LiDAR o los índices y neocanales de Sentinel 2, y su ulterior clasificación. Finalmente, se hizo uso de competencias y conocimientos adquiridos a lo largo del Máster relativos a teledetección, cartografía, programación, clasificación orientada a objetos o el procesado de datos LiDAR.

6. BIBLIOGRAFÍA

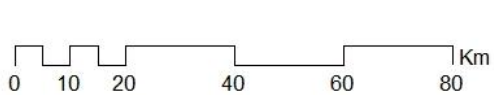
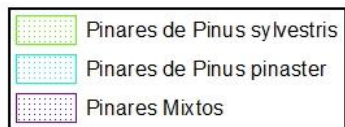
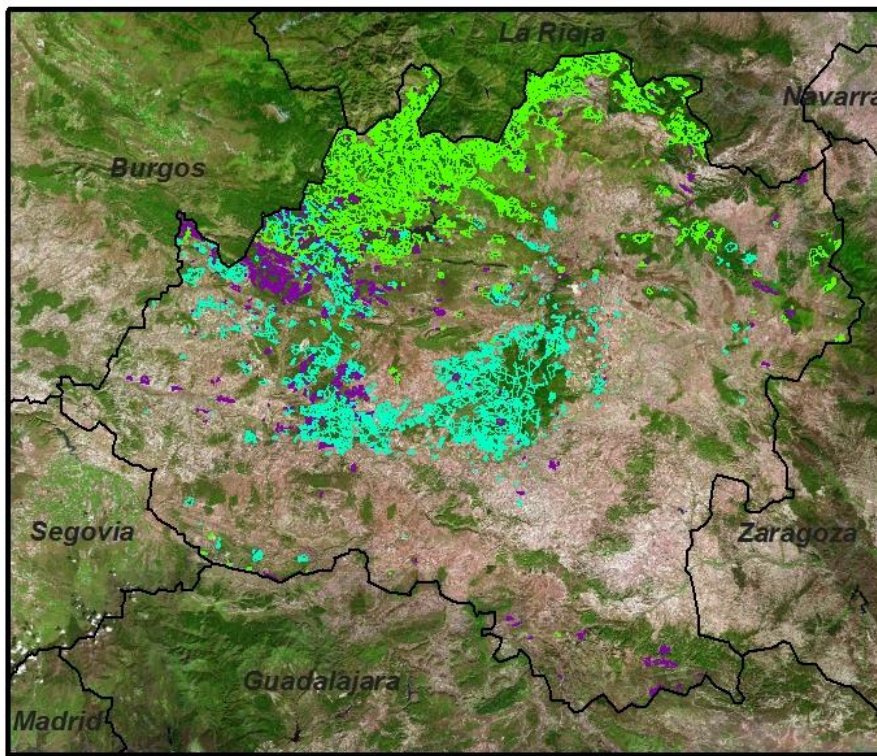
- AIRBUS (2016): AIRBUS DEFENCE &SPACE. Accesible en línea desde <<https://airbusdefenceandspace.com/newsroom/news-and-features/esa-selects-airbus-defence-and-space-for-two-new-sentinel-2-satellites/>> [Último acceso el 31/08/2016]
- Alvarez, F. (2005): Inventariación de las masas de *Pinus radiata* D. Don en El Bierzo (León) empleando teledetección con imágenes multispectrales y clasificadores orientados a objetos. Trabajo final de Postgrado y Master Internacional la distancia en SIG (Programa UNIGIS).
- Álvarez-Bermúdez, X; Valero Gutiérrez del Olmo, E; Picos-Martin, J y Ortiz-Torres, L. (2015): Study of land cover of Monte Forgoselo using Landsat Thematic Mapper 5 images (Galicia, nw Spain), *DYNA82*, pp.173-181.
- Andersen, H-E; McGaughey, R.J; Reutebuch, S.E. (2005): Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment* vol.94 (2005) 441 – 449.
- Astrium,(2012): "SPOT 6 / SPOT 7 Technical Sheet," Acceso en línea <http://www.astrium-geo.com/files/pmedia/edited/r18072_9_spot_6_technical_sheet.pdf> [Último acceso el 23/06/2016]
- Bourel, M. (2012): "Métodos de agregación de modelos y aplicaciones". *Memoria de Trabajos de Difusión Científica y Técnica*, núm. 10.
- Chavez, J. (1988): An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment* 24:459-479
- Chuvieco, E. (1996): *Fundamentos de Teledetección espacial*. Ediciones RIALP. Madrid. 568p
- Chuvieco, E. (2008): *Teledetección ambiental. La observación de la Tierra desde el Espacio*. Editorial Ariel, S.A., Barcelona. 613 pp.
- CNIG. (2016): Centro de Descargas del Centro Nacional de Información Geográfica. Acceso en línea <<http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do#selectedSerie>> [Último acceso el 27/08/2016]
- Cohen, W.B.; Fiorella, M.; Gray, J.; Helmer, E.; Anderson, K. (1998): An Efficient and Accurate Method for Mapping Forest Clearcuts in the Pacific Northwest Using Landsat Imagery. *Photogrammetric Engineering & Remote Sensing*, Vol. 64, No. 4, pp. 293-300.
- Doyle, F.J. (1978): "Digital terrain models: an overview". *Photogrammetric Engineering and Remote Sensing*, 44(12): 1481- 1485.
- European Space Agency. (2015): SENTINEL-2 Technical Guide. Disponible en: <<https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi>> [Último acceso el 23/06/2016]
- García-Martín, A.; J.; De la Riva, J.; Pérez-Cabello, F. y Montorio, R. (2009): "Estimación de la biomasa residual forestal en el contexto de la gestión de los incendios forestales en áreas mediterráneas". *Teledetección: Agua y desarrollo sostenible. XIII Congreso de la Asociación Española de Teledetección*. Calatayud, 23-26 de septiembre de 2009. pp. 345-348.
- García-Martín, A.; García Galindo, D.; Pascual, J.; De la Riva, J.; Pérez-Cabello, F. y Montorio, R. (2011): Determinación de zonas adecuadas para la extracción de biomasa residual forestal en la provincia de Teruel mediante SIG y teledetección. *GeoFocus (Artículos)*, nº 11, p. 19-50.
- Goncalves, G., Seco, L., Reyes, F., Miranda, D. y Coimbra, P. (2008): Land cover classification of rural areas using LiDAR data: a comparative study in the context of fire risk. *Silvi-Laser 2008*, Edimburgo. Pp. 427-436.
- González Ferreiro, E; Barreiro Fernández, L; Miranda, D. (2013): "Clasificación digital de zonas forestales usando datos lidar, ortofoto y análisis orientado a objetos". *Ambienta*, nº 105, pp.66-75.
- Huang, L; Ni, L, (2008): "Object-Oriented Classification of High Resolution Satellite Image for Better Accuracy". *8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Shanghai, P. R. China, June 25-27, 2008, pp. 211-218.
- Koch, B. (2011): La teledetección como apoyo a los inventarios forestales nacionales EFN, *Antología de conocimiento para la evaluación de los recursos forestales nacionales*, FAO-SLU, pp 18.
- Lambert, J; Denux, J-P; Verbesselt, J; Balent, G; Cheret, V. (2015): "Detecting Clear-Cuts and Decreases in Forest Vitality Using MODIS NDVI Time Series", *Remote Sensing*, 7, 3588-3612.

- Lucas Santolaya, J.A. (Sin fecha): “La ordenación casi centenaria de Pinar Grande” Documentación interna del Servicio Territorial de Medio Ambiente de Soria. Junta de Castilla y León. Disponible en línea < http://www.casadelatierra.com/orden_pinargrande.pdf > [Último acceso 07-07-2016]
- Malila, William A., (1980): "Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat". *LARS Symposia*. pp. 385.
- MAGRAMA (2016): *Mapa forestal de España a escala 1:200000. Soria. Hoja 6-4*. Ministerio de Agricultura, Ganadería y Medio Ambiente, Madrid. Accesible en línea < <http://www.magrama.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/mfe200.aspx> > [último acceso 25-06-2016]
- Martínez Cañadas, A; Gallud Lázaro, J.A; Abad Cerdá, F.J. (1998): “Distorsiones en la adquisición de imágenes de teledetección y su corrección”, *Ensayos: Revista de la Facultad de Educación de Albacete nº 13*, pp.349-360
- Pinillos, F.; Moro, O. (2004): “Pinares y el Valle (Burgos, Soria), Certificado de Gestión Forestal Sostenible: Un sello de garantía” *Actualidad Leader: Revista de desarrollo rural*, nº 25, pp.10-11.
- PNOA (2016): Plan Nacional de Ortofotografía Aérea. Ministerio de Fomento. Acceso en línea < <http://pnoa.ign.es/coberturalidar> > [Último acceso el 27/08/2016]
- Riaño, D.; Valladares, F.; Condés, S. y Chuvieco, E. (2004): “Estimation of leaf área index and covered ground from airborne laser scanner (Lidar) in two contrasting forest”, *Agricultural and Forest Meteorology*, nº 124, pp.269-275.
- Ruiz, L.A;del Rey, A; Estornell, J; Ruiz, R. (2007): “La teledetección como herramienta para el análisis del crecimiento urbano y su representación en 3D” *ACE: Arquitectura, Ciudad y Entorno*, Vol. 2, nº 4.
- Skole, D.L. y Tucker, C.J. (1993): “Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988”. *Science* Vol.260, pp.1905-1910.
- Sociedad Española de Ciencias Forestales (2005): “*Diccionario Forestal*”. Ediciones Mundi-Prensa. Madrid. pp. 1314.
- U.S. Geological Survey.(2015): Landsat—Earth observation satellites: U.S. Geological Survey Fact Sheet 2015–3081. Acceso en línea < <http://dx.doi.org/10.3133/fs20153081> > [Último acceso el 23/06/2016]
- Wulder, M.A. y Franklin, S.E. (2003): *Remote sensing of forest environments: concepts and case studies*. Editorial Kluwer Academic Publishers, Boston. 514 pp. (Acceso a través de <http://www.springer.com/us/book/9781402074059> el 22/06/2016)

ANEXO CARTOGRÁFICO

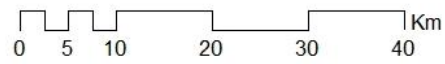
- Principales masas de coníferas en la provincia de Soria

PINARES DE SYLVESTRIS, PINASTER Y MIXTOS EN SORIA

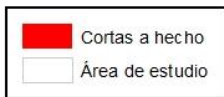
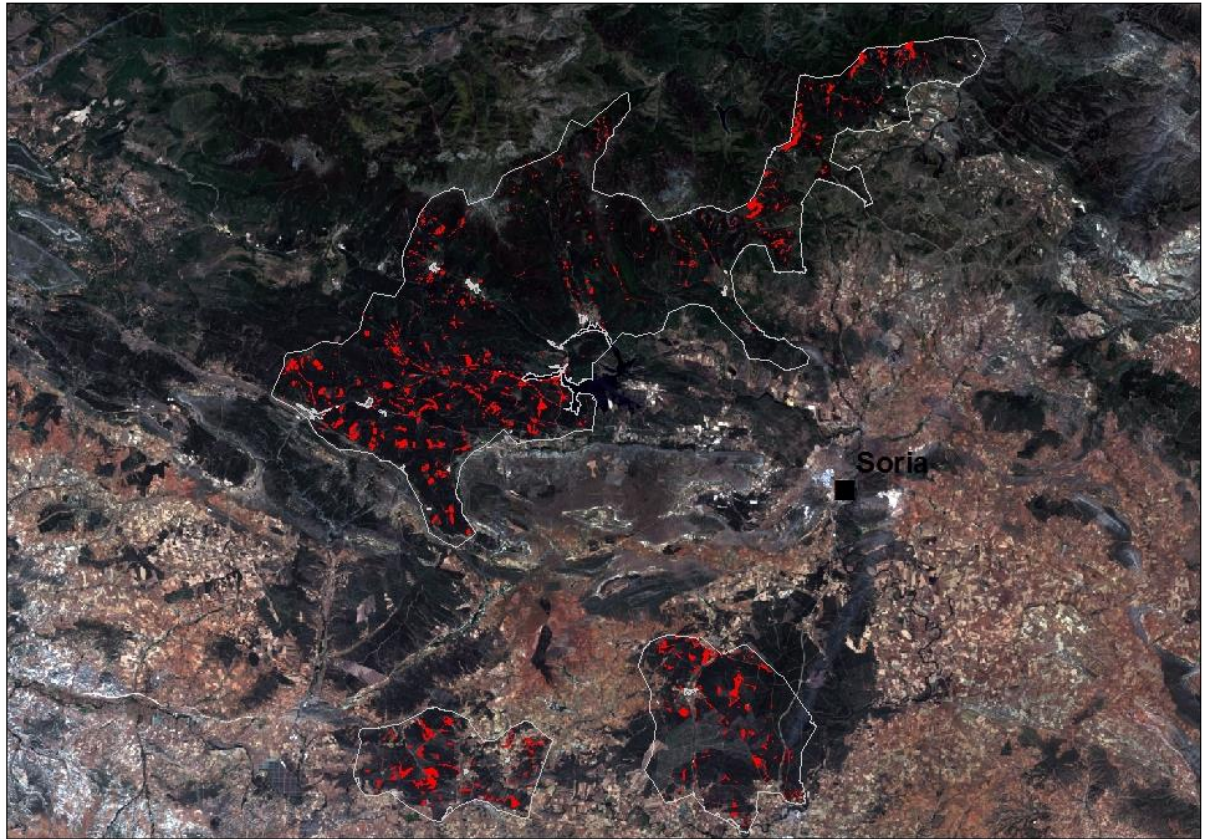


FUENTE:
Geoportal IDEE 2016,
IGN-SIANE 2016,
MAGRAMA-MFE 2016

➤ Área de estudio

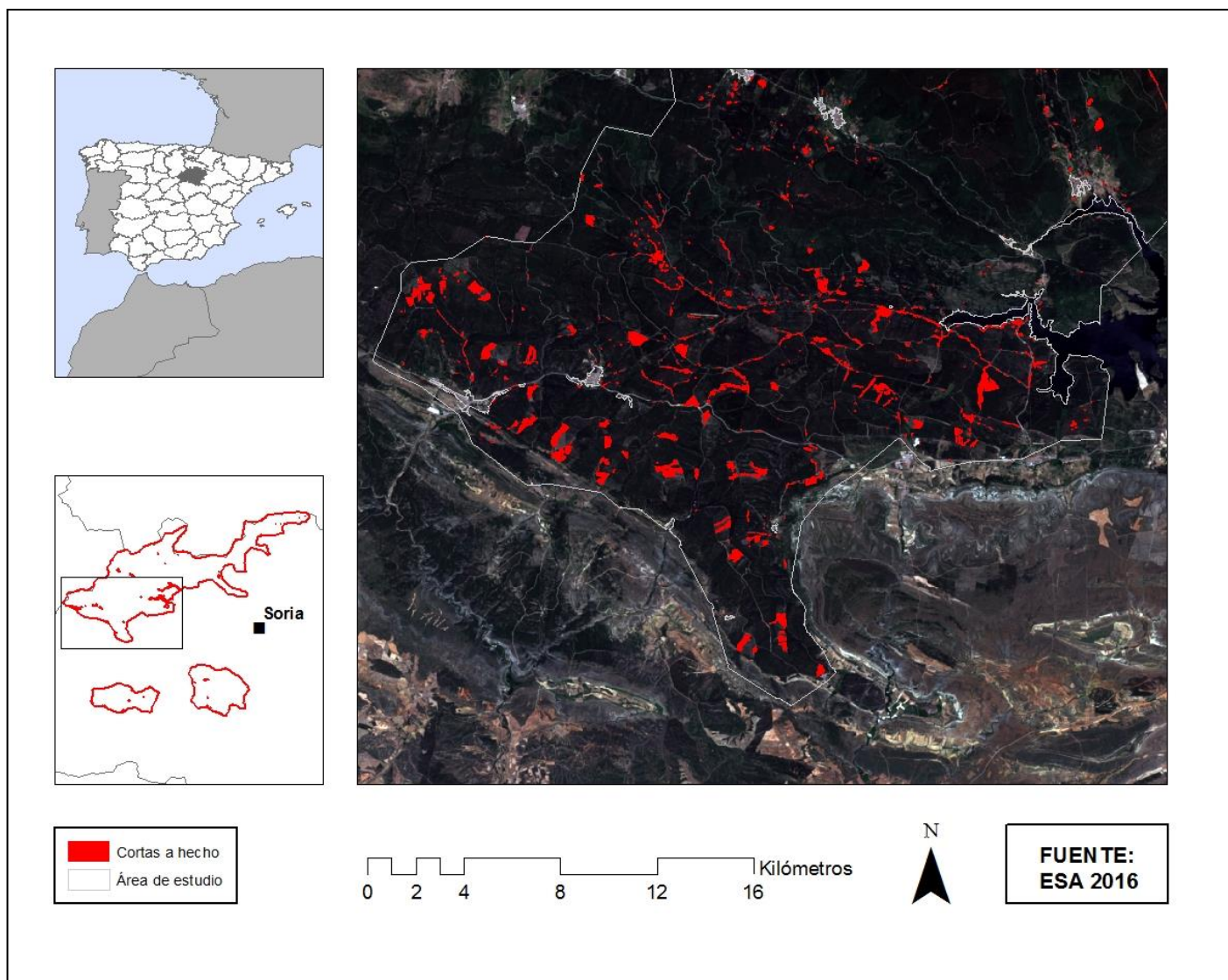


➤ Cartografía de cortas a hecho según la metodología propuesta

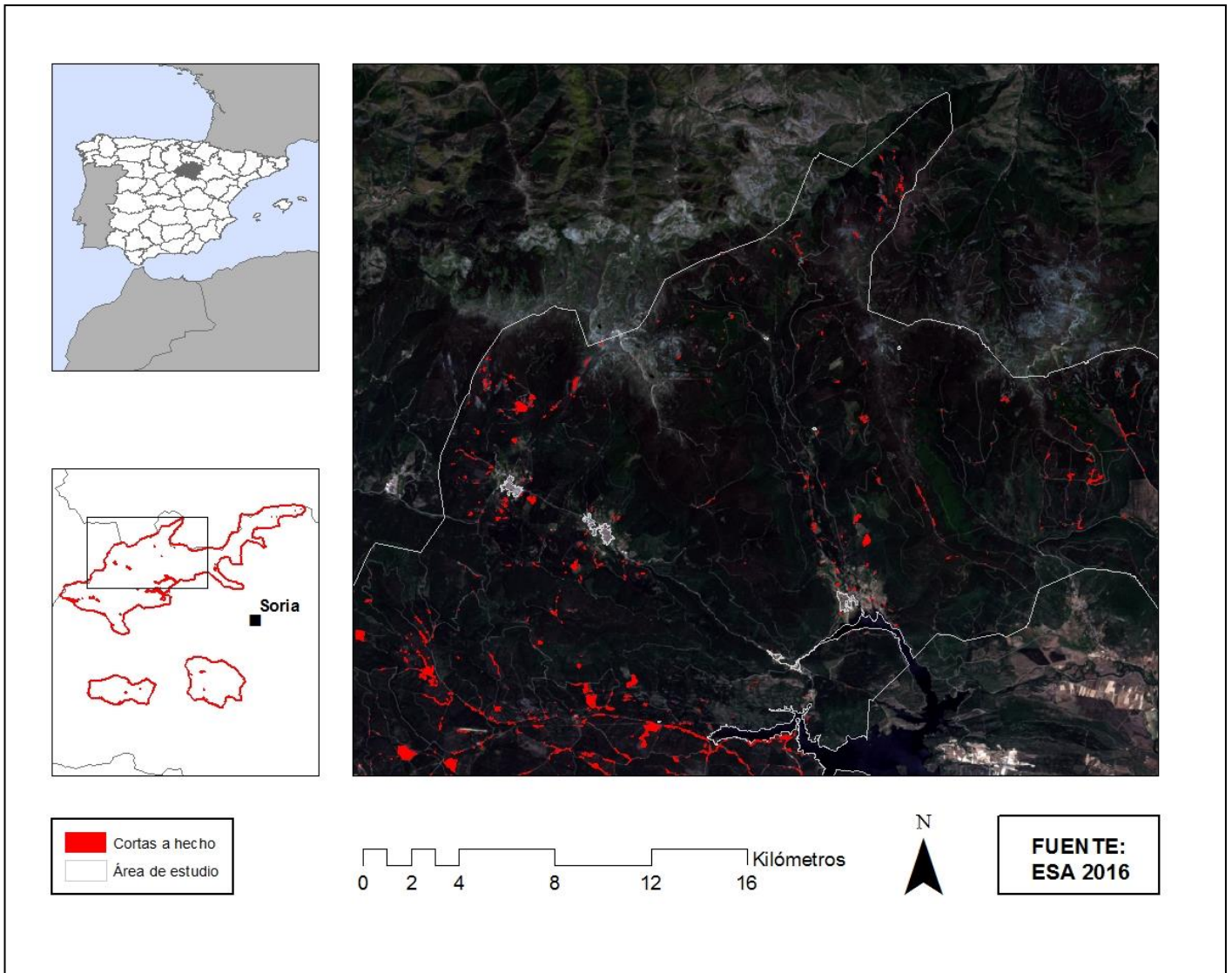


**FUENTE:
ESA 2016**

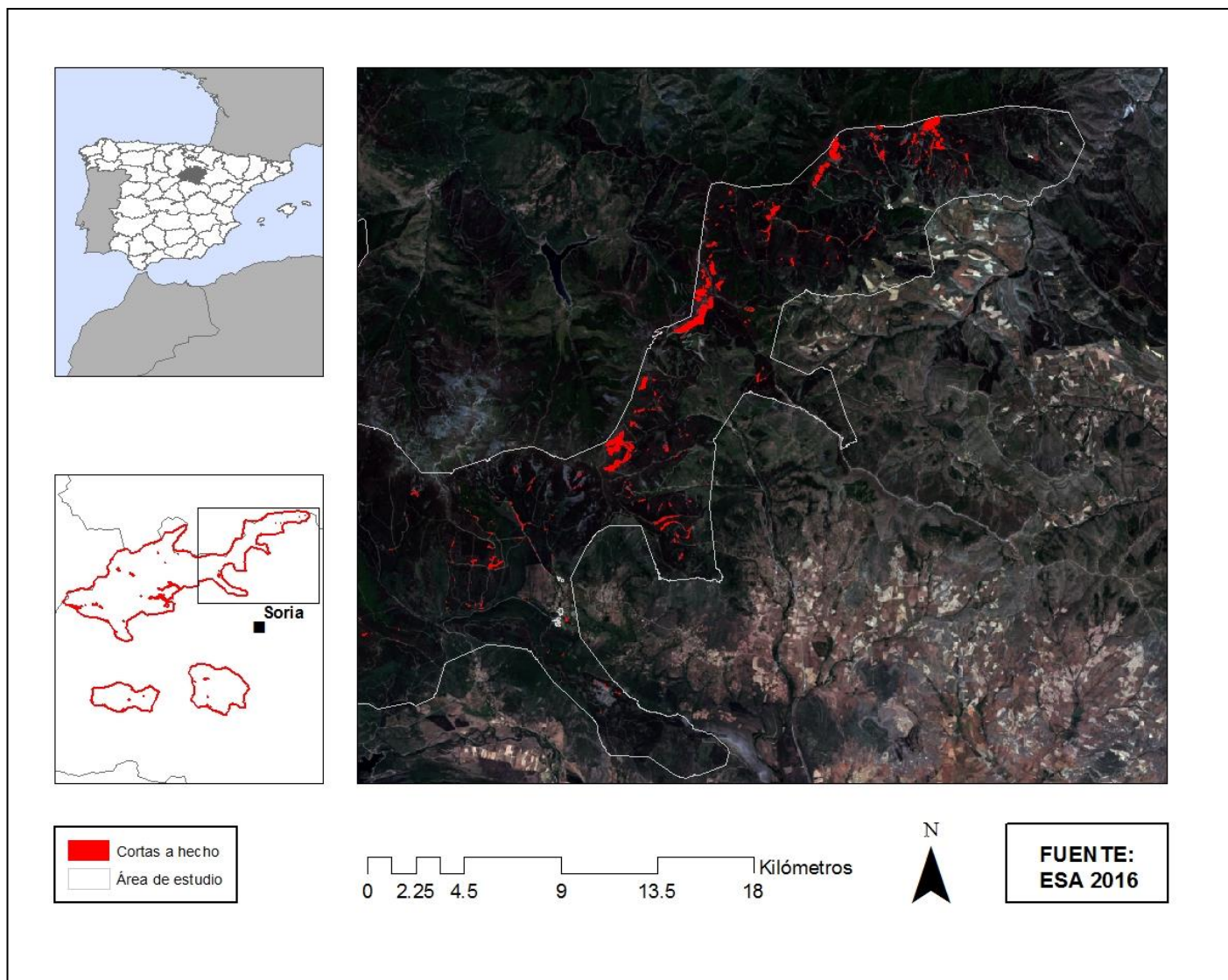
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 1



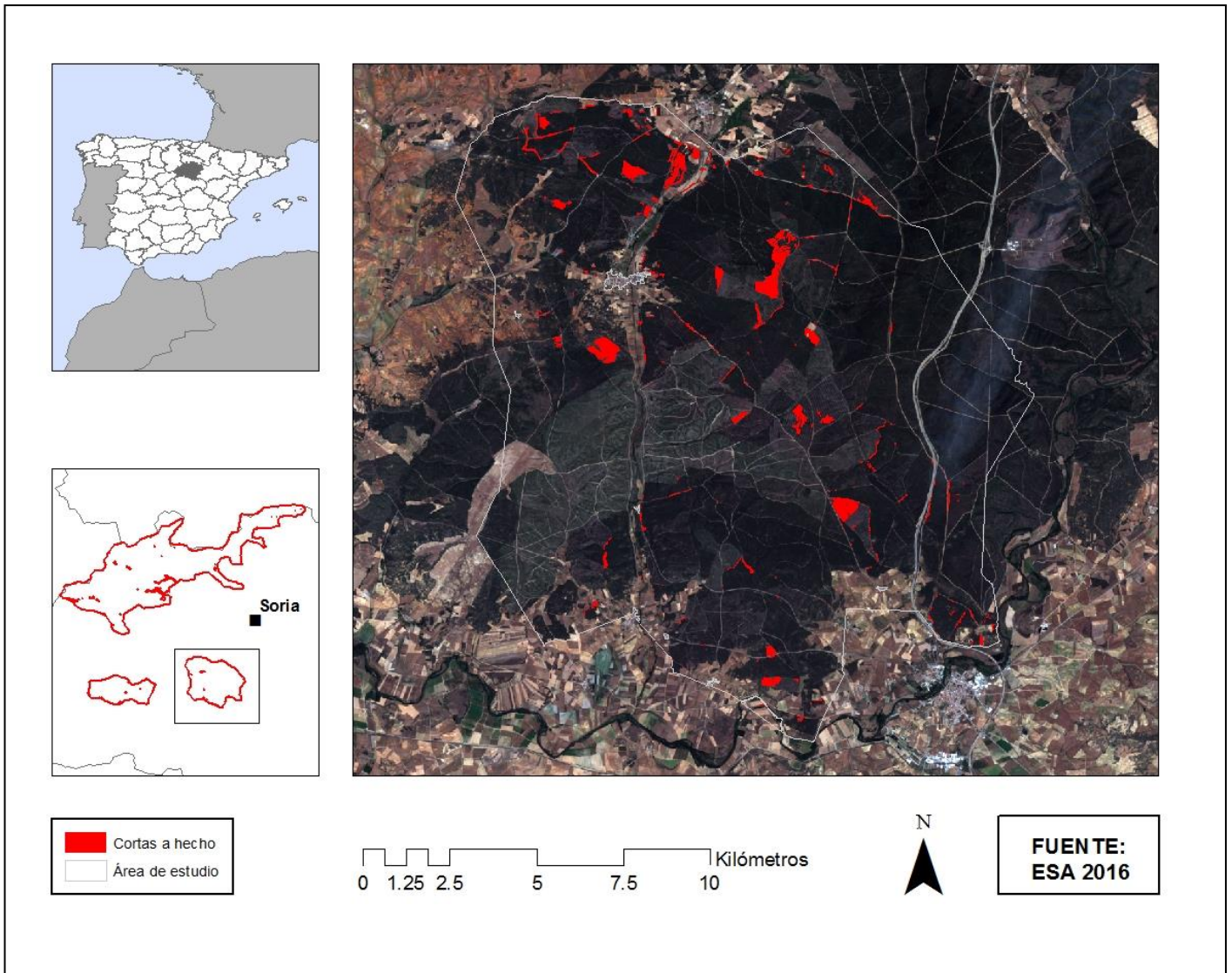
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 2



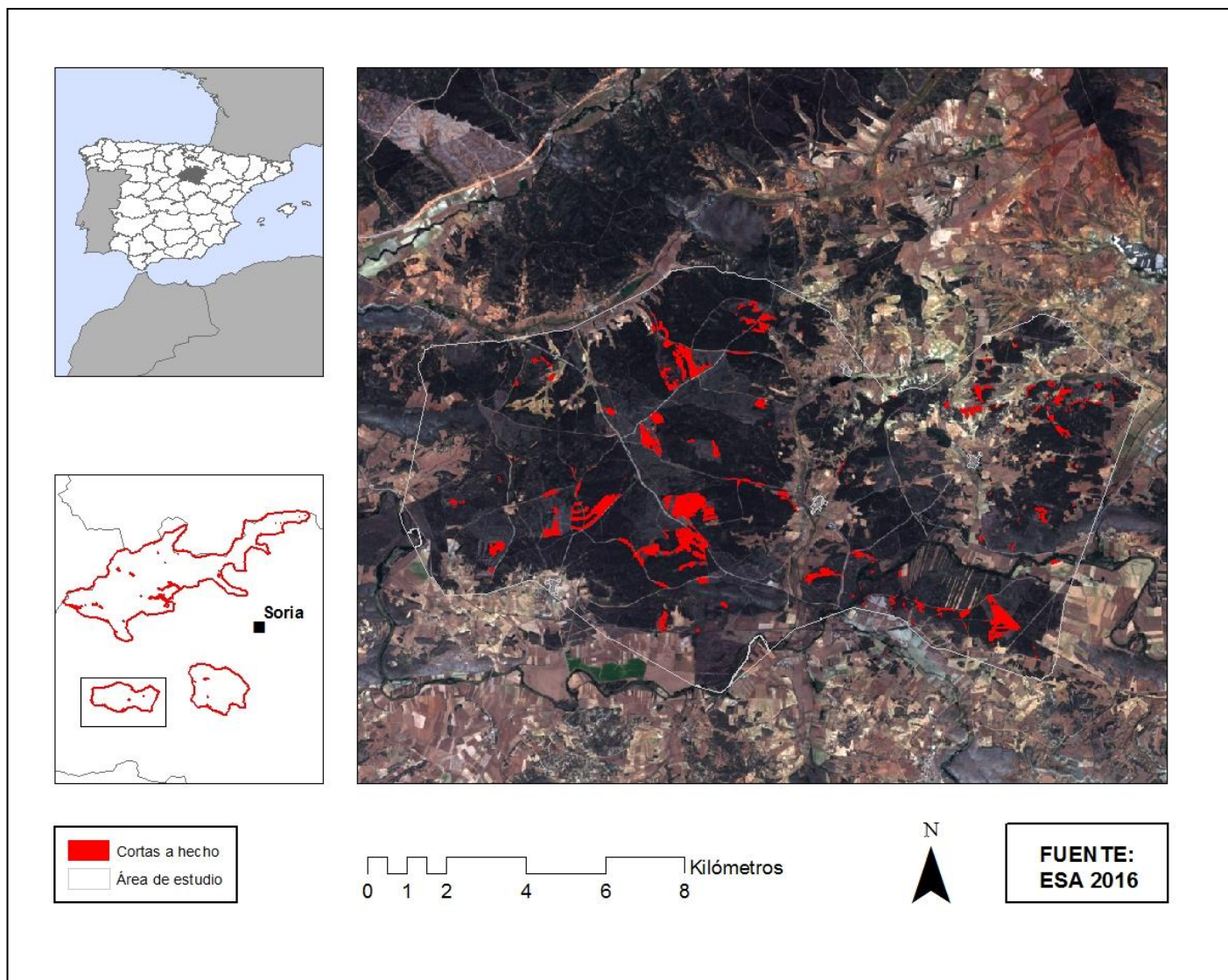
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 3



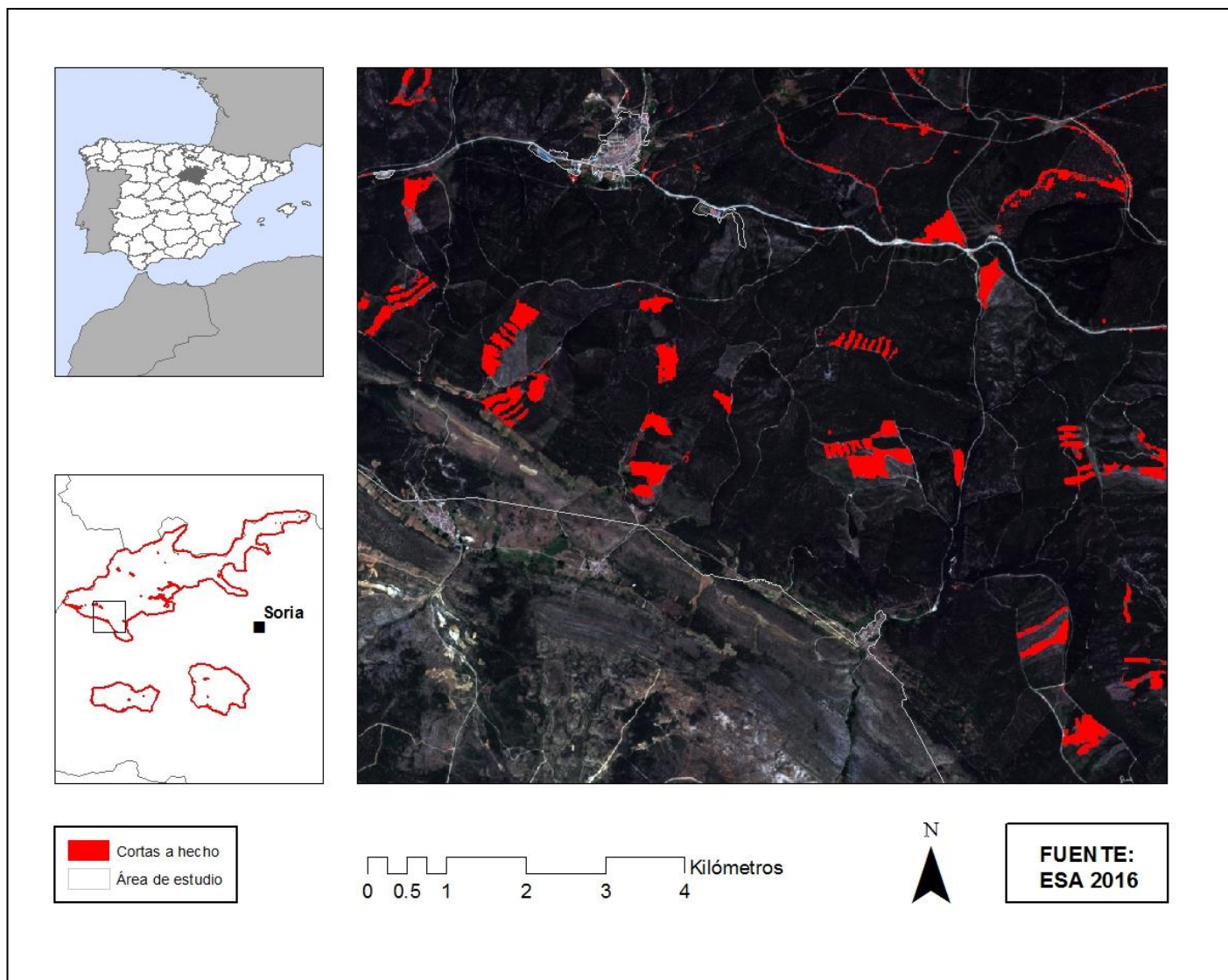
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 4



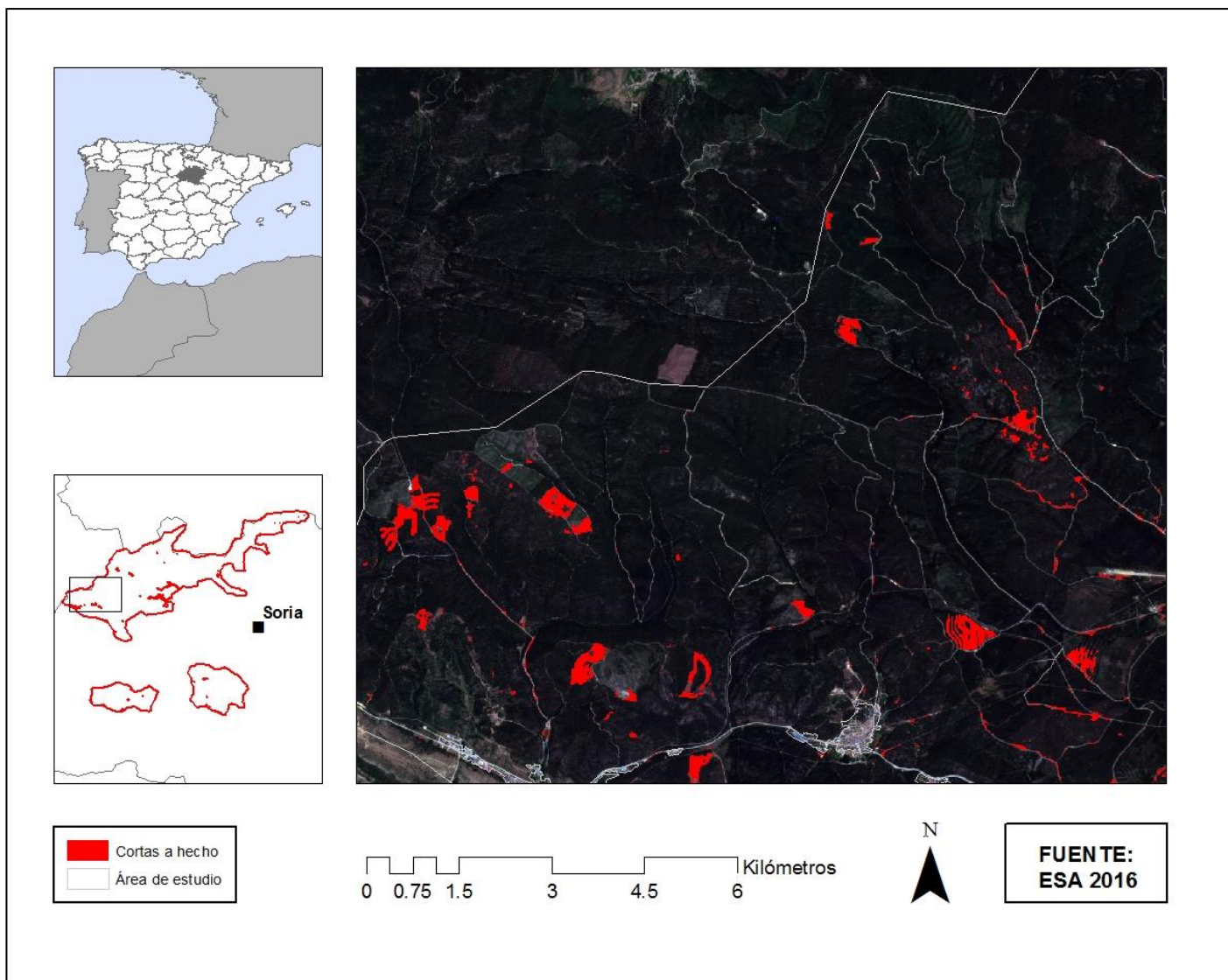
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 5



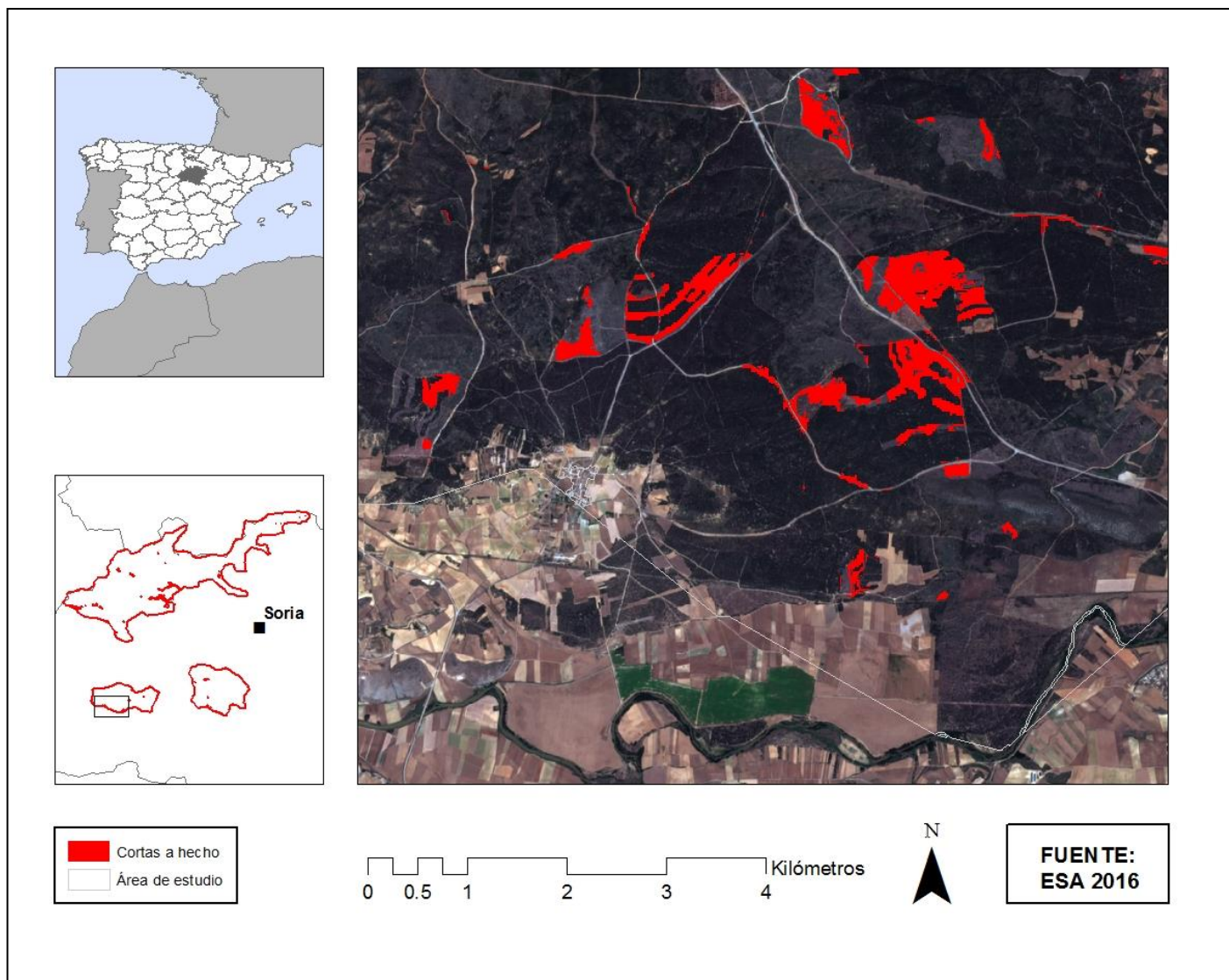
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 6



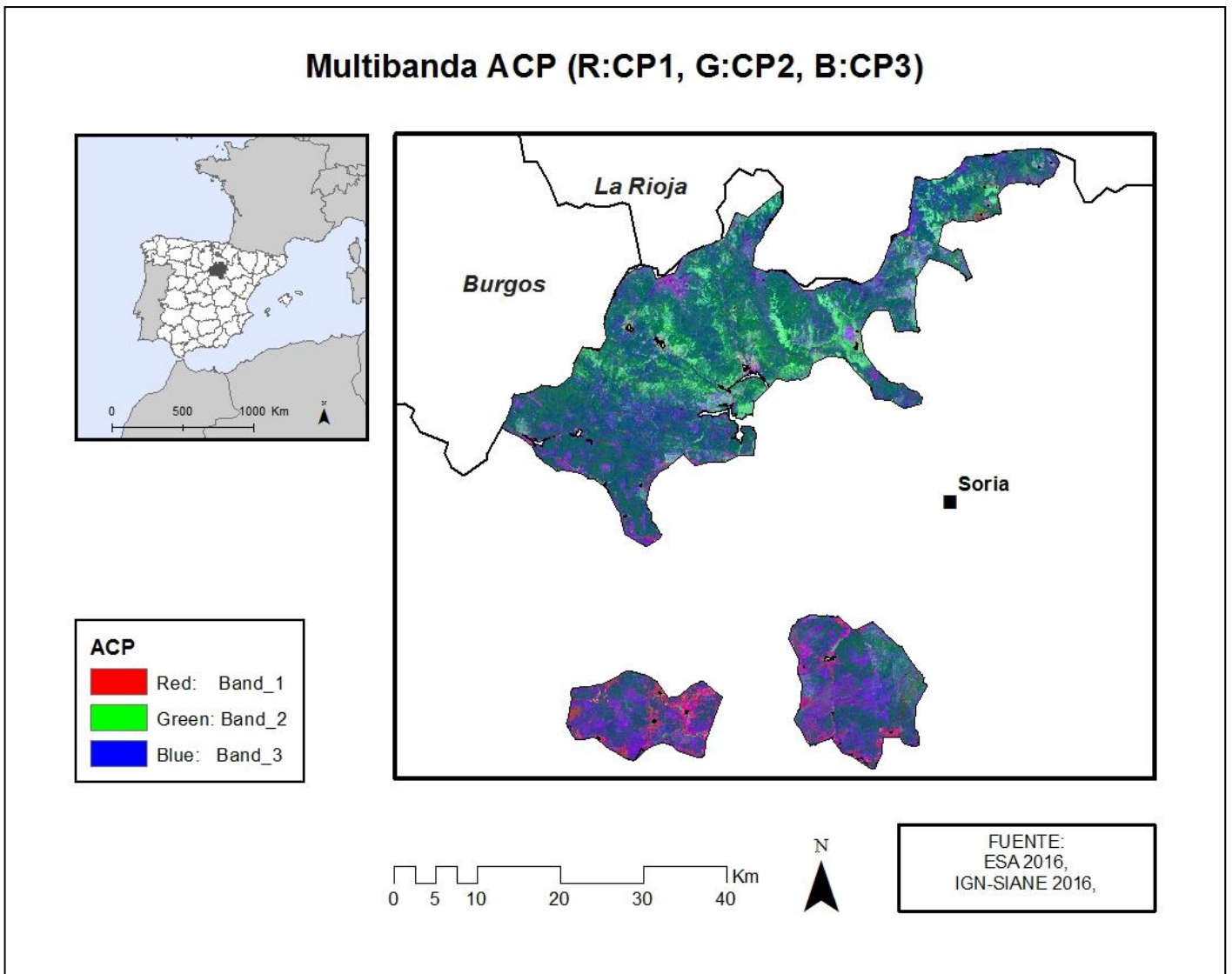
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 7



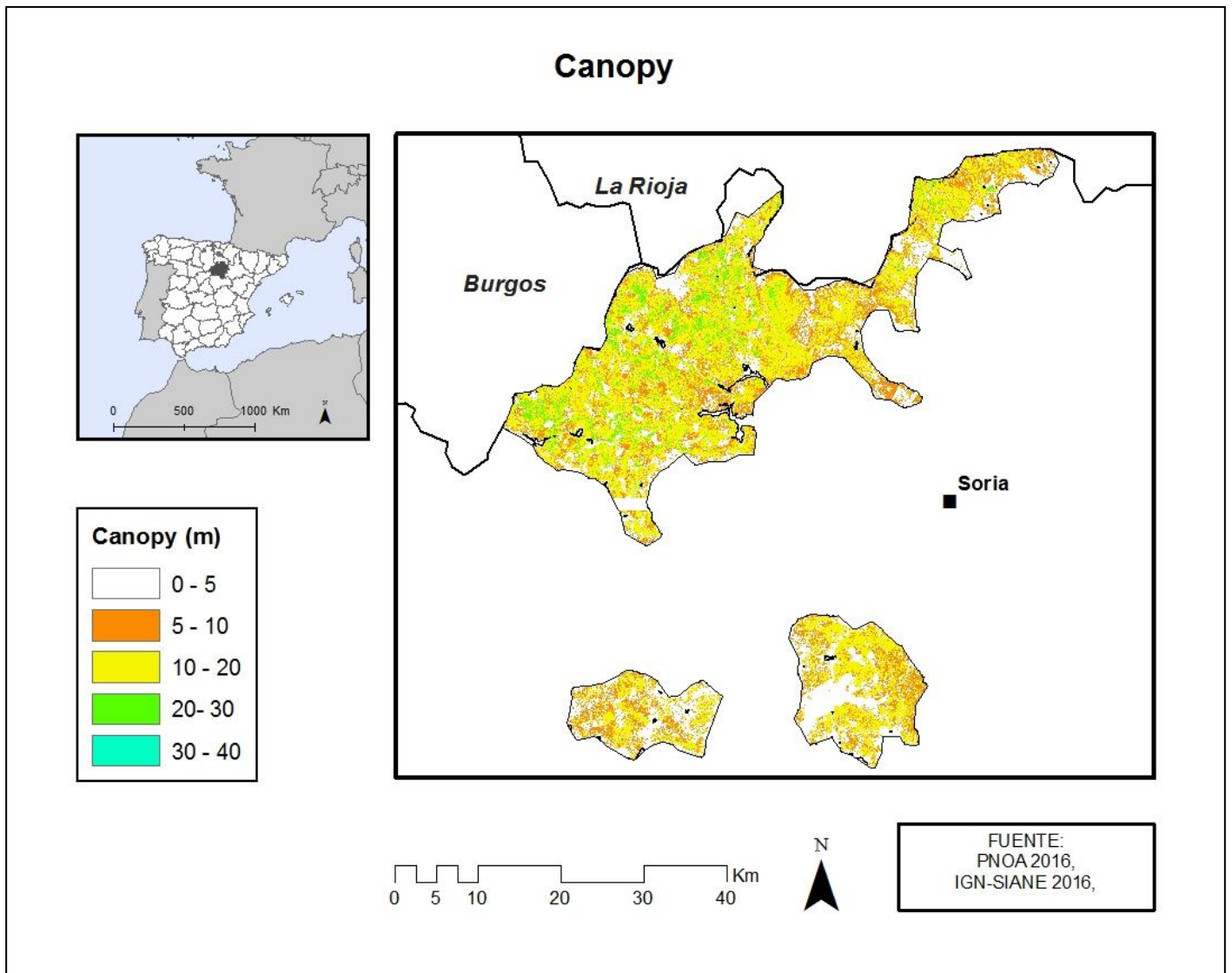
➤ Cartografía de cortas a hecho según la metodología propuesta: Vista en detalle 8



➤ Capas integrantes de la segmentación y la clasificación: Análisis de Componentes Principales

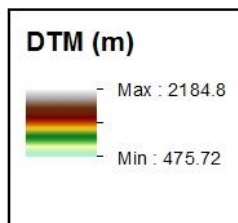
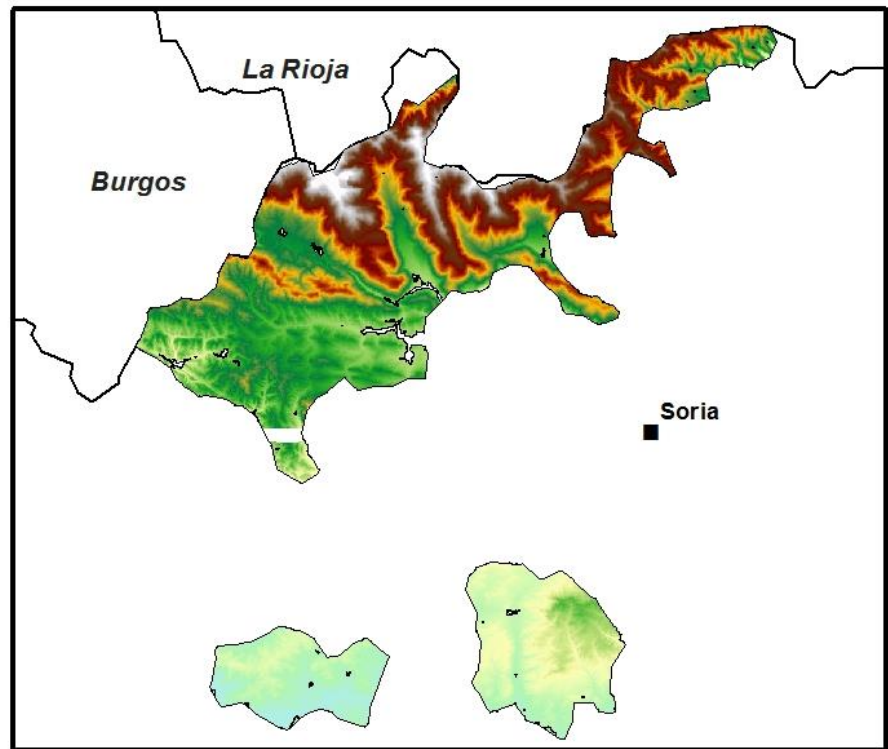


➤ Capas integrantes de la segmentación y la clasificación: Canopy



➤ Capas integrantes de la segmentación y la clasificación: Modelo Digital del Terreno

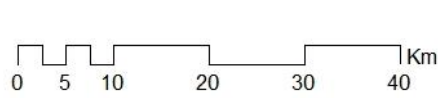
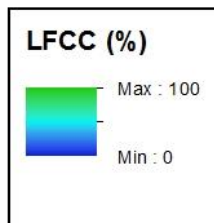
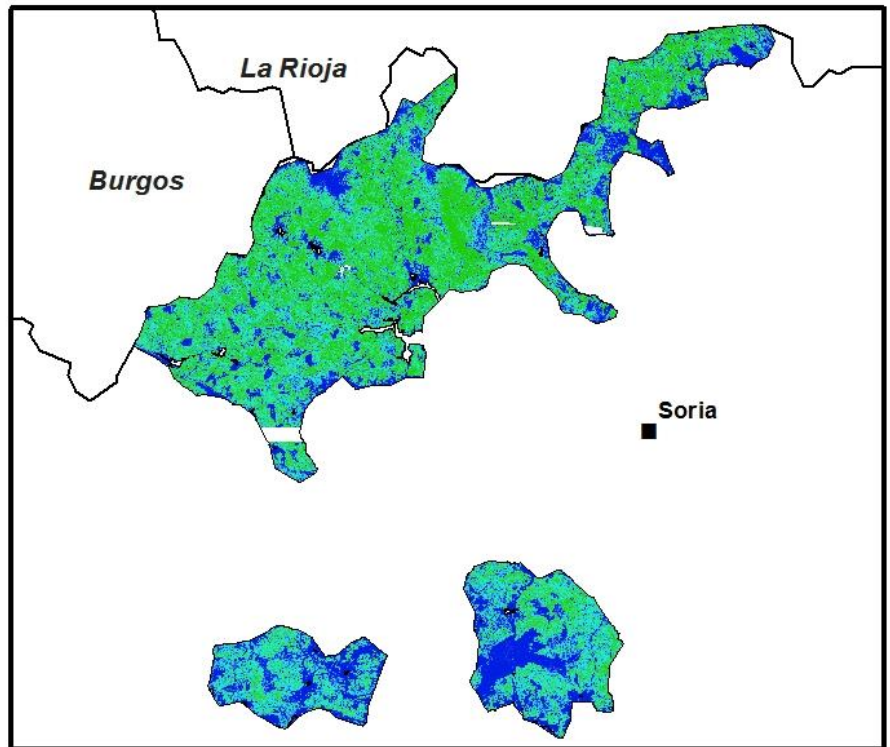
Modelo Digital del Terreno (DTM)



FUENTE:
PNOA 2016,
IGN-SIANE 2016,

- Capas integrantes de la segmentación y la clasificación: Fracción de Cobertura Cubierta

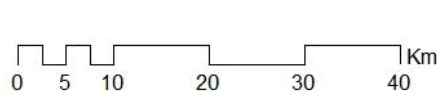
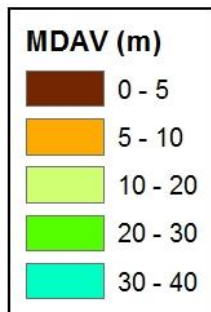
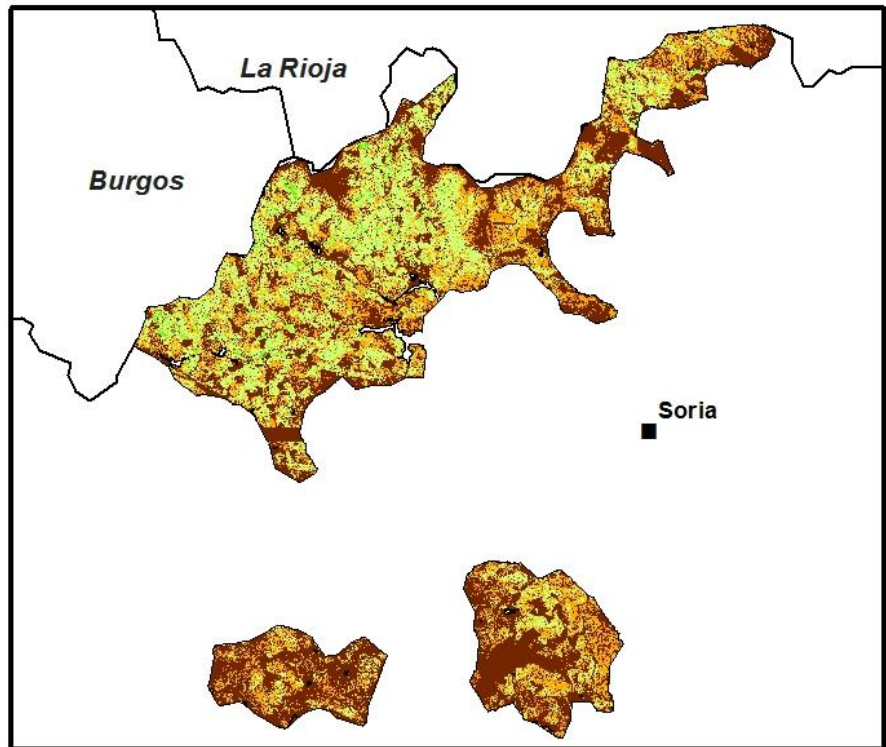
Fracción de Cobertura Cubierta (LFCC)



FUENTE:
PNOA 2016,
IGN-SIANE 2016,

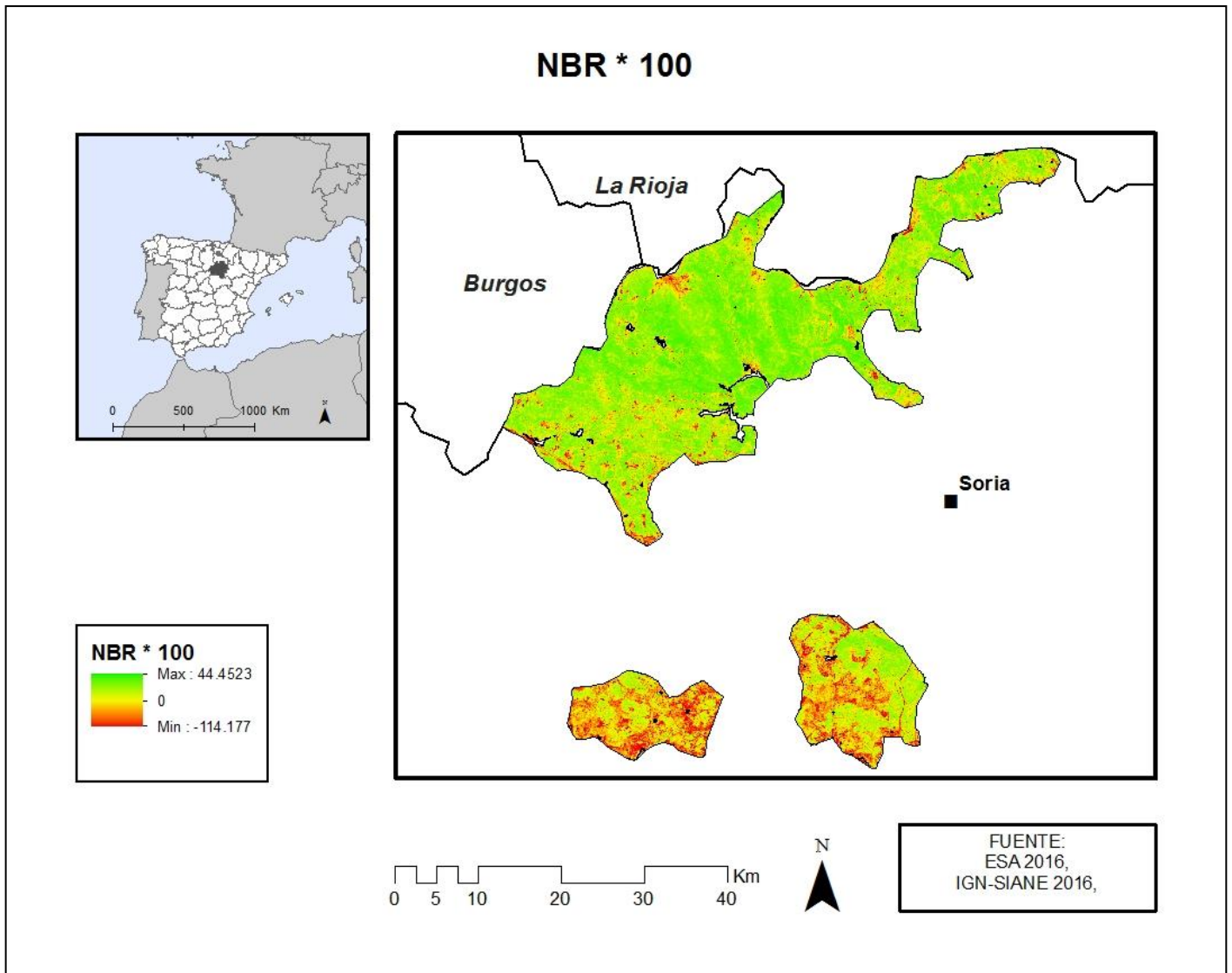
- Capas integrantes de la segmentación y la clasificación: Modelo Digital de Altura de la Vegetación

Modelo Digital de Altura de la Vegetación (MDAV)

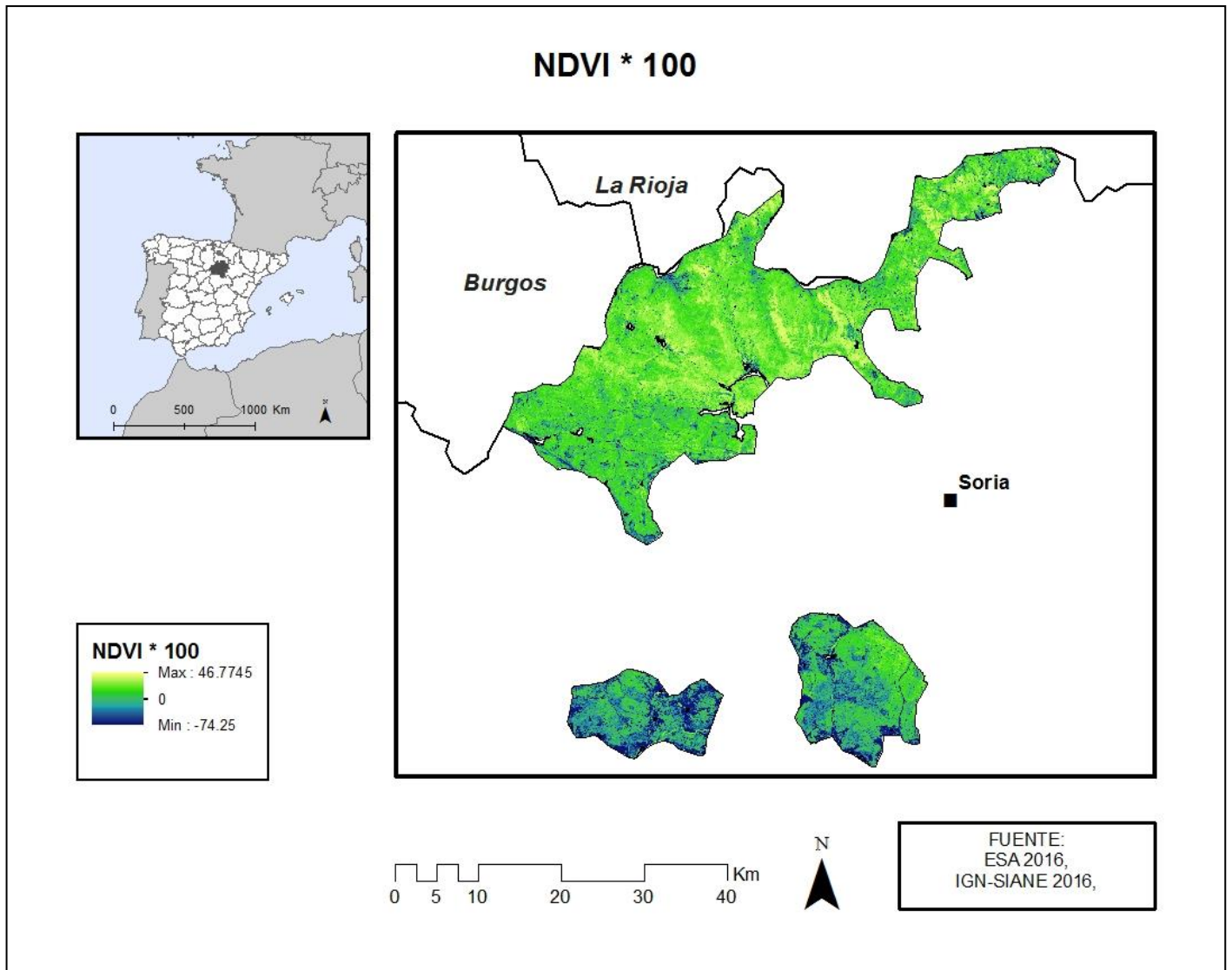


FUENTE:
PNOA 2016,
IGN-SIANE 2016,

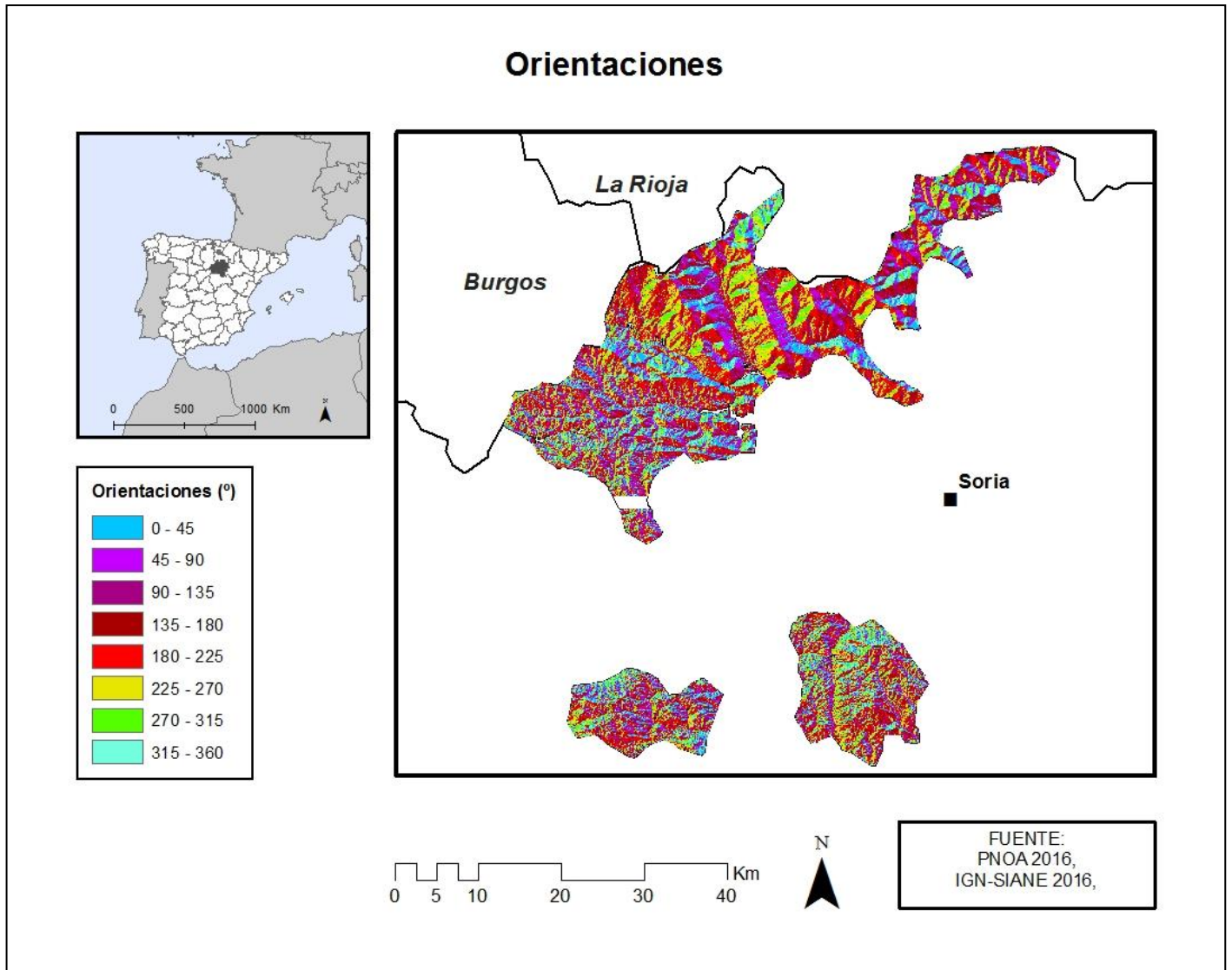
➤ Capas integrantes de la segmentación y la clasificación: Normalized Burn Ratio



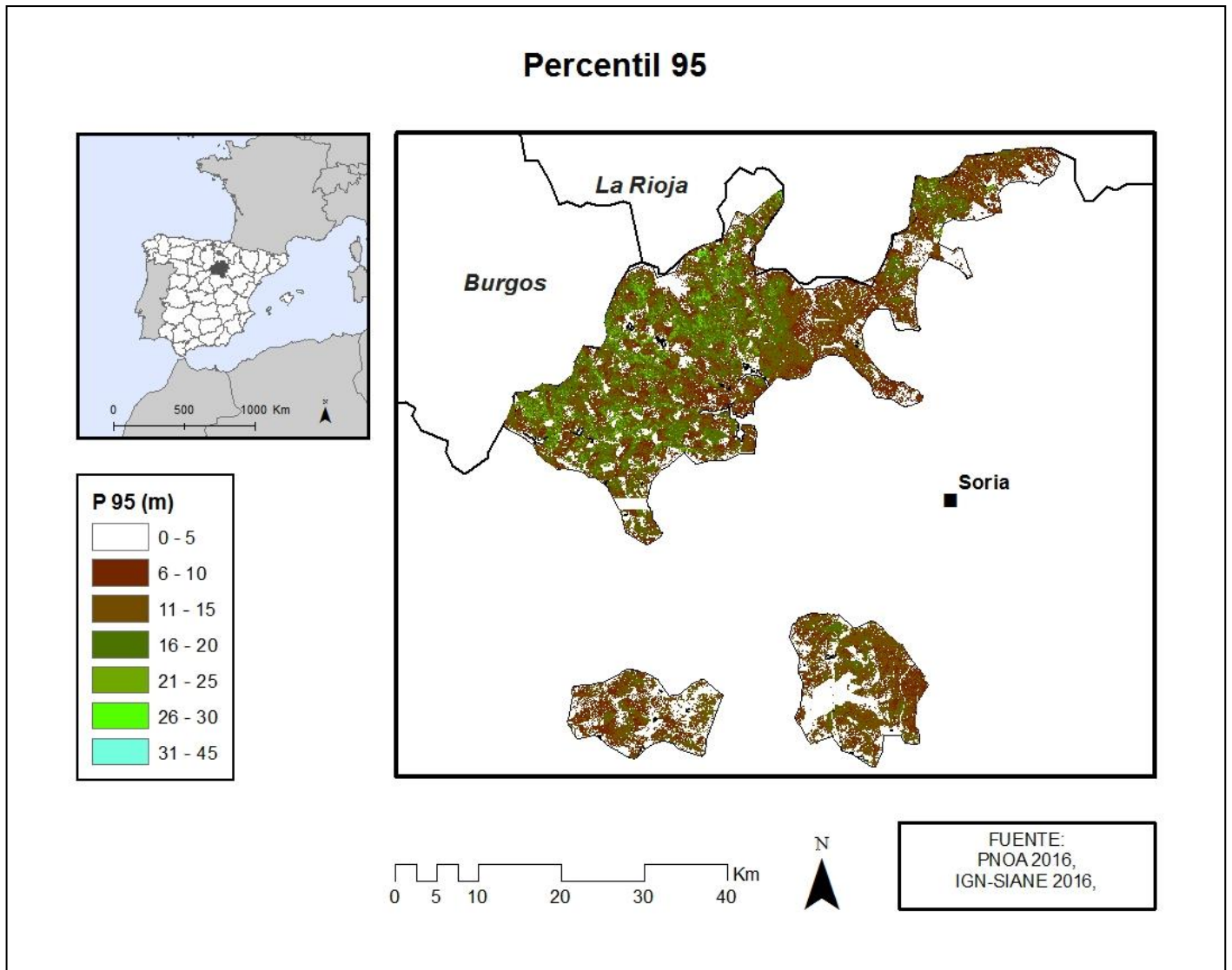
- Capas integrantes de la segmentación y la clasificación: Normalized Differenced Vegetation Index



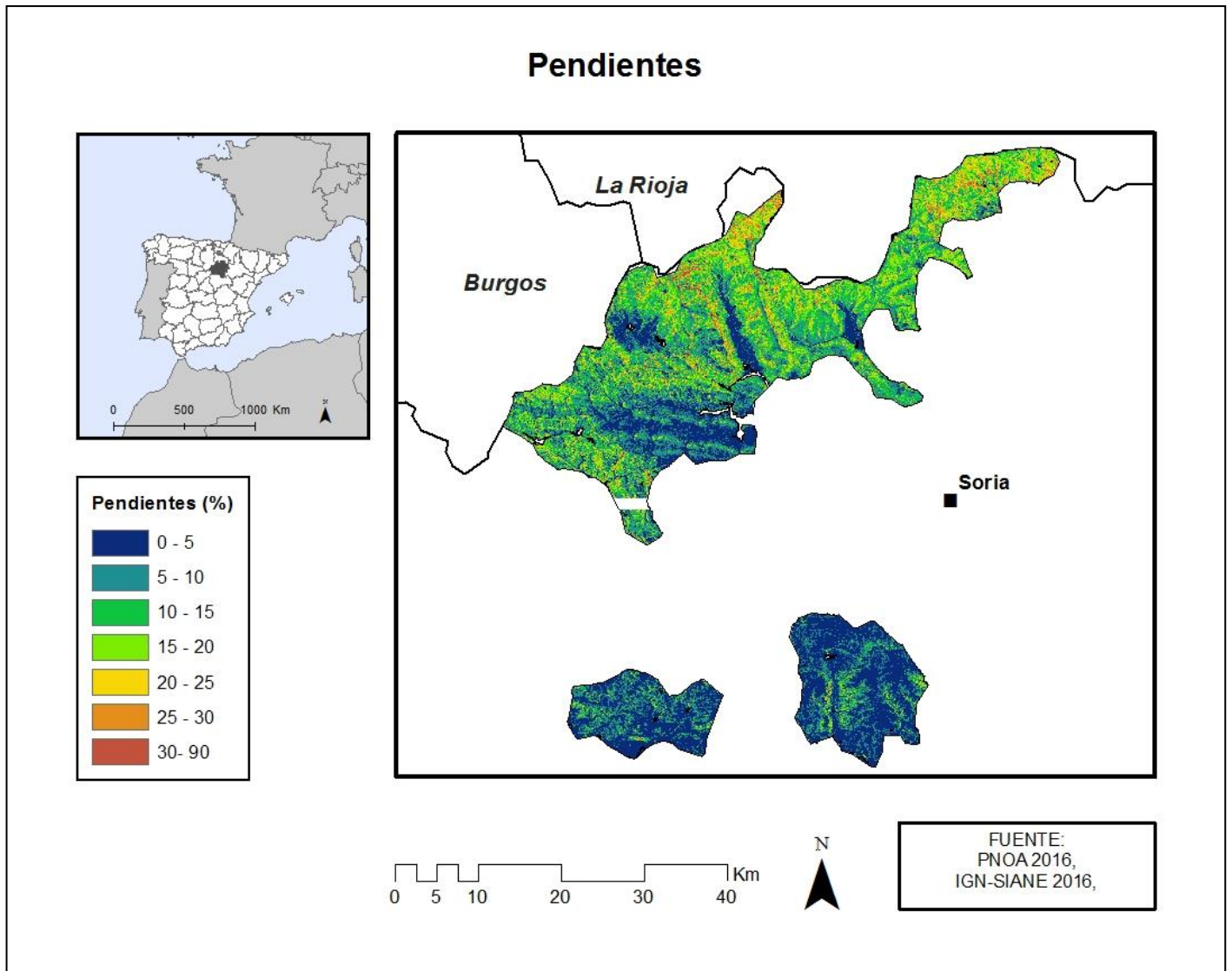
➤ Capas integrantes de la segmentación y la clasificación: Orientaciones



➤ Capas integrantes de la segmentación y la clasificación: Percentil 95



➤ Capas integrantes de la segmentación y la clasificación: Pendientes



➤ Capas integrantes de la segmentación y la clasificación: Imagen multibanda Sentinel 2A

