

UNIVERSIDAD DE OVIEDO

ESCUELA POLITÉCNICA DE MIERES

**MÁSTER UNIVERSITARIO EN GEOTECNOLOGÍA Y DESARROLLO DE
PROYECTOS SIG**

**DEPARTAMENTO DE BIOLOGÍA DE ORGANISMOS Y SISTEMAS
ÁREA DE ECOLOGÍA**

TRABAJO FIN DE MÁSTER

MODELADO DE LA DISTRIBUCIÓN DE LOS ECOSISTEMAS FORESTALES DEL PARQUE NACIONAL PICOS DE EUROPA: UNA APROXIMACIÓN METODOLÓGICA BASADA EN TÉCNICAS DE TELEDETECCIÓN

AUTOR: Daniel Pfitzer López

TUTOR: Susana Suárez Seoane

COTUTORES: Francisco De Borja Jiménez-Alfaro

José Valentín Rocas Díaz

JULIO, 2023

Índice

Resumen	1
1. Introducción	2
2. Objetivos	4
3. Área de estudio	5
4. Material y métodos	8
4.1. Aproximación metodológica	8
4.2. Fuentes de datos	10
4.2.1. Variables predictoras	10
4.2.2. Variables respuesta	11
4.3. Procesado de las variables predictoras	13
4.4. Muestreo de puntos de entrenamiento	14
4.4.1. Muestreo	14
4.4.2. Filtros de calidad: altura mínima de la cobertura arbolada y NDVI mínimo	17
4.5. Selección de variables de entrada al modelo	18
4.6. Análisis de modelado espacial	19
4.7. Validación	20
4.7.1. Validación cruzada	20
4.7.2. Análisis de incertidumbre	20
4.7.3. Matriz de confusión	21
4.8. Complemento QGIS	22
5. Resultados	24
6. Discusión	28
7. Conclusiones	30
Referencias	31
Anexos	34
A. Complemento de QGIS	1
B. Salidas generadas por el complemento de QGIS	72

Resumen

En este estudio se desarrolló una herramienta metodológica de modelado espacialmente explícito, basada en técnicas de teledetección, cuya funcionalidad es identificar, de manera automatizada en un entorno SIG, los patrones de distribución de los ecosistemas forestales presentes en espacios geográficos con interés ecológico, considerando como estudio de caso el Parque Nacional Picos de Europa. Se combinaron datos espectrales obtenidos a partir del satélite Sentinel-2 con datos climáticos y topográficos mediante una aproximación de modelado jerárquica (modelo jerárquico) y se comparó el rendimiento del modelo con el de otro basado exclusivamente en datos espectrales (modelo no jerárquico) para predecir la distribución de los ocho ecosistemas forestales más relevantes en el área de estudio. El modelo jerárquico mostró una incertidumbre significativamente menor que el no jerárquico, resultando en una clasificación un 17.22% más precisa. El clima fue el factor más importante para explicar los patrones de distribución, seguido por los datos de satélite y los topográficos. El desarrollo en Python del complemento para QGIS permitió una mayor flexibilidad en el procesado y análisis de los datos. Este enfoque, basado en una aproximación metodológica jerárquica, presenta relevancia en aplicaciones relacionadas con la conservación, la planificación y la gestión de recursos y de espacios naturales.

Abstract

This study presents a GIS tool developed for spatially explicit modeling using Picos de Europa National Park as a case study. This software is based on remote sensing techniques which automatically identify the distribution patterns of forest ecosystems in geographic areas of ecological interest. A hierarchical classification model for eight forest habitats was generated by employing a combination of climatic, topographic, and spectral data, collected from the Sentinel-2 satellite, which was subsequently compared to a non-hierarchical model, solely based on spectral data. The hierarchical model showed significantly less uncertainty than the non-hierarchical one, resulting in a more accurate classification by 17.22%. The climate was found to be the preeminent factor in explaining distribution patterns, followed by satellite and topographic data. The development of the QGIS plugin in both Python allowed for greater flexibility both data processing and analysis. This approach holds potential applications in the conservation, planning, and management of natural spaces.

1. Introducción

En las últimas décadas, el estudio de la distribución espacial de las especies vegetales ha cobrado creciente importancia (Guisan y Thuiller, 2005; Geary y col. 2020), puesto que proporciona una eficaz herramienta para la toma de decisiones relacionadas con la gestión de los ecosistemas, incluyendo la identificación de áreas prioritarias para la conservación, así como la evaluación de los efectos del cambio climático sobre la biodiversidad (Margules y Pressey, 2000). A partir de datos geográficos de distribución de un ecosistema forestal, es posible determinar distintos indicadores de su estado de conservación que permiten valorar su potencial inclusión en la Lista Roja de Ecosistemas de la Unión Internacional para la Conservación de la Naturaleza (UICN). Es el caso de la extensión de ocurrencia (EOO) y el área de ocupación (AOO), parámetros definidos por Rodríguez y col. (2015). EOO se entiende como «el área contenida dentro del polígono convexo que incluye todas las ubicaciones (Figura 1a) donde un ecosistema está presente» (Figura 1b), mientras que el AOO se define como «el área interior de la extensión de ocurrencia que es ocupada en realidad por el ecosistema» (Figura 1c).

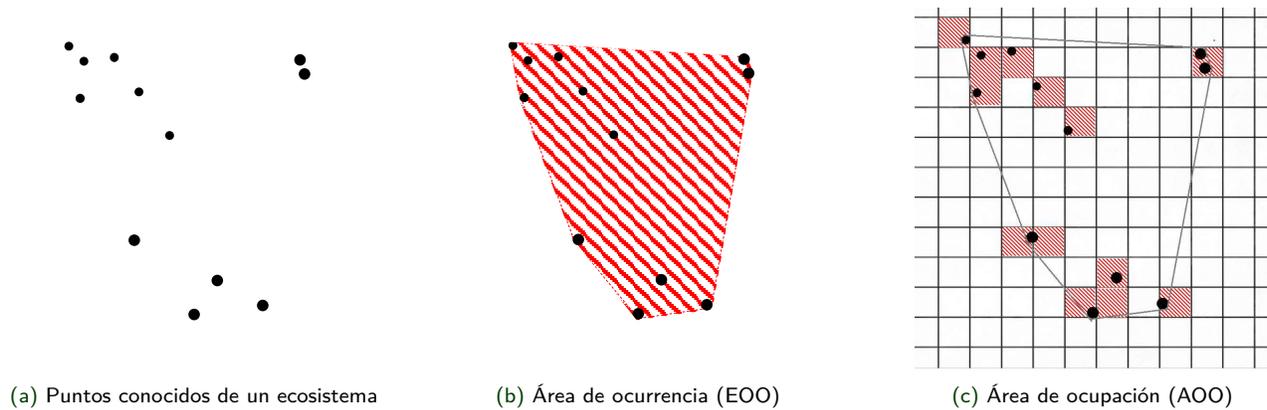


Figura 1: Esquema conceptual de la extensión de ocurrencia (EOO) y Área de Ocupación (AOO) a una escala determinada

En este contexto, resulta fundamental disponer de herramientas que permitan estimar de forma precisa y eficiente la distribución espacial de los ecosistemas. Especial relevancia presentan en este sentido los modelos de distribución de especies (SDM – Guisan y col. 2013), los cuales se basan en el establecimiento de relaciones estadísticas entre observaciones de campo (variables respuesta) y datos ambientales (variables predictoras) a través de modelos empíricos (Guisan y Zimmermann, 2000). No obstante, la utilidad de estos modelos en aplicaciones ambientales está frecuentemente limitada por la disponibilidad de datos de calidad a la resolución espacial y temporal necesaria para reflejar adecuadamente los procesos ecológicos subyacentes (Keil y col. 2013).

En cualquier caso, y debido a la propia naturaleza de las variables predictoras, cuando se parametriza un SDM es habitual la necesidad de combinar capas de información geográfica con distinta resolución espacial (por ej. variables climáticas, topográficas, edáficas o socio-económicas) en un único modelo. En este caso, es posible organizarlas en grupos jerárquicos. Esto implica que los ecosistemas pueden ser modelados a partir de diferentes grupos de variables predictoras en pasos independientes y sucesivos, lo que se denomina SDM jerárquico (Hefley y Hooten, 2016). Una de las principales ventajas de este enfoque frente a un SDM no jerárquico es proporcionar una forma estructurada de organizar el modelo y permitir la reutilización de submodelos comunes (Kang y col. 2009).

Dentro de este marco, Álvarez-Martínez y col. (2018) desarrollaron un método para predecir la distribución espacial de ecosistemas en paisajes heterogéneos combinando datos climáticos a baja resolución espacial con datos topográficos y espectrales (satélite) de resolución más fina. Los datos climáticos permiten generar un primer modelo de área de ocupación (AOO) potencial, que predice el grado de idoneidad climática para un ecosistema. Este primer modelo climático, análogo al de área de ocurrencia (EOO), es determinante para el resultado final, ya que el clima es uno de los principales factores que explica y limita la distribución espacial, tanto de especies vegetales como animales (Guisan y Zimmermann, 2000). Para rectificar la falta de precisión del modelo AOO potencial causada por la falta de disponibilidad de información climática a alta resolución, éste se subdivide utilizando datos topográficos y espectrales de satélite con mayor resolución espacial, dando lugar a un modelo de AOO real, análogo al AOO de Rodríguez y col. (2015).

En grandes áreas de estudio, resulta común el empleo de imágenes espectrales derivadas de satélite para predecir la distribución espacial de ecosistemas, ya que permiten discriminar de forma rápida y eficiente entre distintos tipos de vegetación y presentan una amplia disponibilidad de de manera continuada en el espacio y en el tiempo (Lengyel y col. 2008). Sin embargo, en el caso de paisajes heterogéneos con topografía irregular, se pueden generar problemas de confusión en el modelo espectral que impiden una correcta clasificación de estos ecosistemas (Randin y col. 2020). Esto se debe a que las condiciones microclimáticas originadas por la orografía afectan a la variabilidad de la vegetación (Álvarez-Martínez y col. 2018), resultando más eficaz la combinación de datos espectrales con variables predictoras climáticas y topográficas. En relación con esta cuestión, es necesario el desarrollo de nuevas herramientas que permitan a los gestores del territorio evaluar los patrones de distribución de los ecosistemas presentes en áreas con especial interés ecológico de forma sencilla y con bajo esfuerzo.

2. Objetivos

El objetivo principal de este trabajo es desarrollar una herramienta metodológica de modelado espacial, basada en técnicas de teledetección, cuya funcionalidad es identificar, de manera automatizada en un entorno SIG, los patrones de distribución de los ecosistemas forestales más relevantes presentes en espacios geográficos con interés ecológico, considerando como estudio de caso el Parque Nacional Picos de Europa.

Los objetivos específicos del estudio son los siguientes:

- (i) Comparar el rendimiento de un modelo jerárquico basado en clima, topografía y datos espectrales capturados por el satélite Sentinel-2 con un modelo no jerárquico basado exclusivamente en los datos de satélite.
- (ii) Evaluar el papel de las familias de variables predictoras (clima, topografía y satélite) más relevantes a la hora de explicar la distribución de los ecosistemas forestales, identificando la aportación relativa de cada una de las variables a los modelos.
- (iii) Analizar espacialmente el nivel de incertidumbre de las clasificaciones realizadas.
- (iv) Desarrollar un complemento de QGIS capaz de replicar la metodología en cualquier área geográfica de estudio.

Se plantean las siguientes hipótesis para este estudio:

1. La integración de datos climáticos y topográficos con la información espectral de Sentinel-2 permitirá desarrollar un modelo jerárquico de distribución de los ecosistemas forestales más preciso que el basado exclusivamente en datos espectrales, dado que la topografía y el clima juegan un papel determinante a la hora de modelar especies y ecosistemas forestales en áreas de montaña.
2. La información climática tendrá una aportación relativa al modelo jerárquico superior a las aportaciones de topografía y satélite, ya que el clima es el factor que más limita la presencia de las especies vegetales.
3. Las aportaciones relativas de las bandas e índices espectrales de Sentinel-2 al modelo jerárquico resultarán distintas en cada ecosistema forestal, ya que las especies vegetales presentan distinta respuesta espectral.

3. Área de estudio

Se ha elegido el Parque Nacional de los Picos de Europa (PNPE) como estudio de caso por su adecuación a los objetivos del trabajo. De las 64660 ha de extensión total, el 23.62 % corresponde a una variedad de ecosistemas forestales (15271 ha) distribuidos a lo largo de un fuerte gradiente climático y topográfico (Figura 2), donde se dispone de cartografía detallada de las diferentes unidades ecológicas¹ para generar modelos de distribución de ecosistemas forestales fiables.

El Parque Nacional de Picos de Europa se localiza en la zona central de la Cordillera Cantábrica, en las provincias de Cantabria, Asturias y León, a 20km del mar Cantábrico (Figura 3). Presenta un clima eminentemente Atlántico en la vertiente norte con influencia Mediterránea en el sur, categorizado como Macrobioclima Templado (Rivas, 2007). Las montañas de los Picos de Europa son los tres macizos más altos y extensos de la Cordillera Cantábrica, caracterizados por una topografía muy accidentada donde las altas cumbres se combinan con valles profundos, gargantas y cañones, alcanzando desniveles de más de 2000 metros (García y González, 2000). El amplio rango de variación altitudinal determina que un fuerte gradiente de temperatura que influye en la distribución de la vegetación, favoreciendo una gran diversidad de ecosistemas. Si se agrupan cada uno de los espacios que se suceden altitudinalmente con las consiguientes variaciones de temperatura, lo que se conoce como termopisos o pisos bioclimáticos (Rivas, 2007), se identifican cuatro clases donde los ecosistemas forestales están claramente diferenciados (Felpete, 2011):

- Termotipo Termotemplado (piso Termocolino): por debajo de 250 m de cota.
- Termotipo Mesotemplado (piso Colino): entre 250 y 850 m de cota.
- Termotipo Supratemplado (piso Montano): entre 900 y 1500 m de cota.
- Termotipo Orotemplado (pisos Subalpino y Alpino): de 1600 m de cota en adelante. El límite superior está en torno a 2000 m de altura.



Figura 2: Superficies (ha) de los principales ecosistemas forestales presentes en el PNPE. La barra superior indica la superficie forestal y no forestal en el área de estudio. Las 3 barras inferiores representan la superficie ocupada por cada tipo de bosque. Datos: Prieto y Bueno (2013)

¹Mapa de Vegetación 1:10000 del PNPE (Prieto y Bueno, 2013)

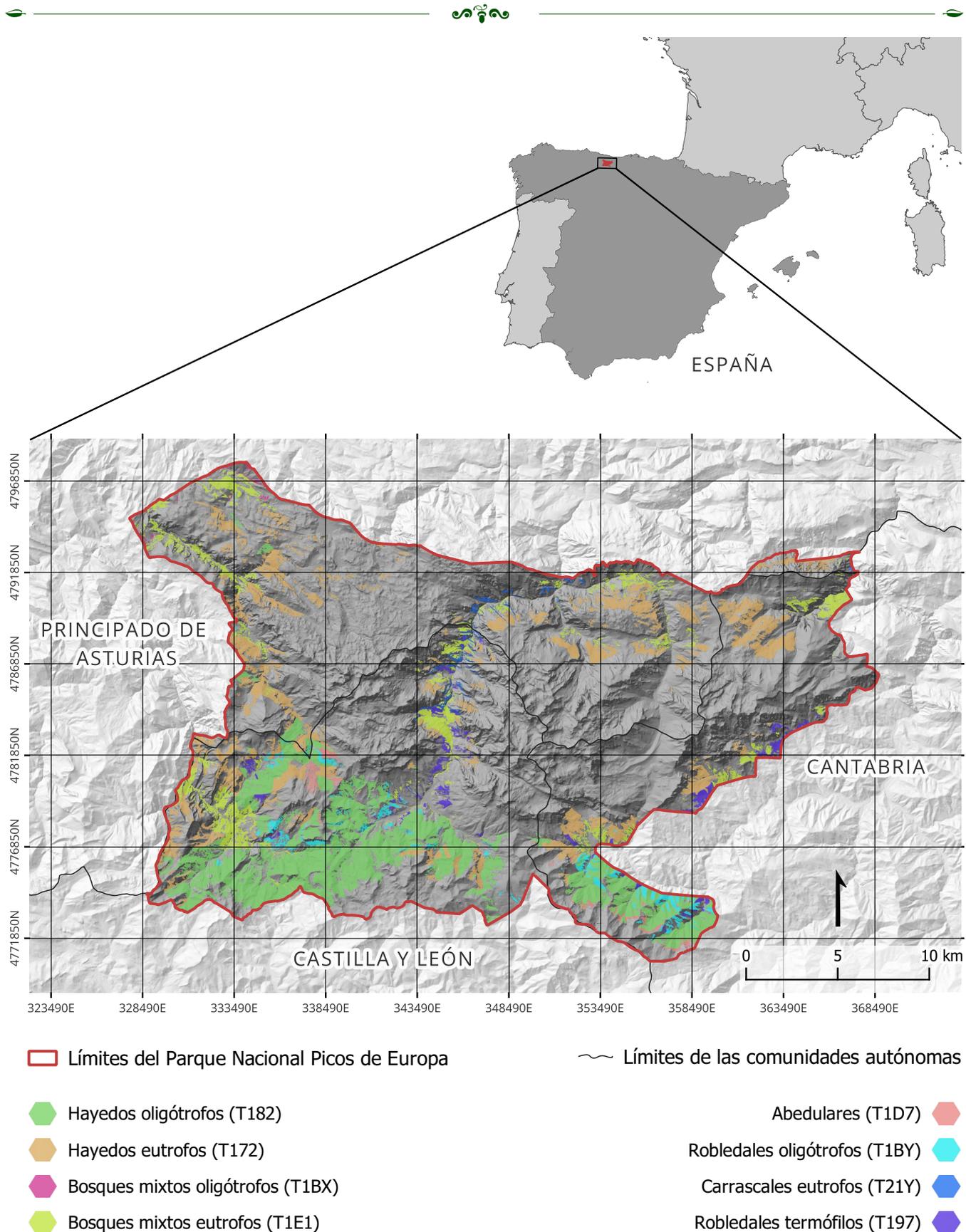
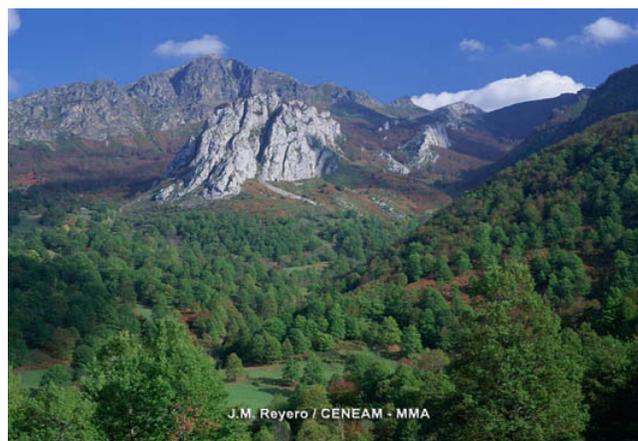


Figura 3: Mapa de situación del Parque Nacional Picos de Europa. Coordenadas UTM Huso 30, expresadas en metros. Las unidades de vegetación siguen la denominación EUNIS (sección 4.2) y están tomadas del Mapa de Vegetación 1:10.000 de Picos de Europa (Prieto y Bueno, 2013)

Tal como ilustra la Figura 2, el 82 % de la superficie de ecosistemas forestales del PNPE corresponde a hayedos (*Fagus sylvatica* L.). Estos bosques crecen sobre cualquier tipo de sustrato en los pisos Montano y Subalpino (hasta los 1700 m), preferentemente en laderas umbrías (Ruiz de la Torre, 2001). Las mayores masas de hayedo corresponden al tipo de hayedo oligótrofo, desarrollado sobre suelos pobres de naturaleza silíceo (Arquer y Valderrábano Luque, 2007). Los bosques desarrollados sobre suelos ricos en nutriente son los hayedos eutrofos. Los bosques mixtos atlánticos, formados por robles (*Quercus robur* L.) entremezclados con tilos (*Tilia* L.), fresnos (*Fraxinus* L.), castaños (*Castanea sativa* Mill.), hayas y nogales (*Juglans regia* L.) se distribuyen en las zonas húmedas del piso Colino, junto a tilares y carrascales (*Quercus ilex* L.) en suelos calcáreos. En el piso Montano, además de hayedos, se desarrollan bosques de robles (*Quercus robur* L. y *Quercus petraea* Matt.) y abedules (*Betula pubescens* Ehrh.). Por encima de estas cotas, en el piso Subalpino y Alpino, la vegetación arbórea desaparece y se instalan comunidades arbustivas (Felpete, 2011), las cuales escapan del objetivo de este trabajo centrado en hábitats puramente forestales. La Figura 4 muestra algunos ejemplos de los ecosistemas descritos en el PNPE: hayedos eutrofos, oligótrofos y un bosque mixto.



(a) Hayedo oligótrofo (*Fagus sylvatica*).
Fotografiado por C. Valdecantos
(Galería fotográfica de Picos de Europa)



(b) Bosque mixto de hayas (*Fagus sylvatica*) y robles albares (*Quercus petraea*). Fotografiado por J.M. Reyero
(Galería fotográfica de Picos de Europa)



(c) Hayedo eutrofo en las proximidades de Sotres – Fuente:
Arquer y Valderrábano Luque (2007)



(d) Hayedo oligótrofo en Camaleño, Cantabria – Fuente:
Arquer y Valderrábano Luque (2007)

Figura 4: Hayedos y bosque mixto en el Parque Nacional de los Picos de Europa

4. Material y métodos

4.1. Aproximación metodológica

El marco conceptual en el que se basa la aproximación metodológica desarrollada en este estudio está reflejado en Álvarez-Martínez y col. (2018), donde se calculan de forma jerárquica el AOO potencial (AOO-P), local (AOO-L) y real (AOO-R). El AOO-P, estimado a partir de datos climáticos, se define como el área con idoneidad climática para un tipo de ecosistema forestal y resulta útil para estimar distribuciones en áreas de estudio extensas, pero está limitado por su baja resolución espacial. El AOO-L se calcula como un subconjunto del AOO-P considerando rasgos estructurales del paisaje (topografía) a mayor resolución. Mientras que Álvarez-Martínez y col. (2018) incluyen datos espectrales en el modelo AOO-L, en este estudio se opta por generar un modelo adicional, denominado AOO real parcial (AOO-RP), basado en la información de Sentinel-2. Aislar la información de satélite permite comprobar su aportación al modelo y medir qué método de modelado presenta un mejor rendimiento (sección 4.6). El AOO-RP representa, para cada ecosistema forestal, un subconjunto del AOO-L donde se considera la respuesta espectral como predictor a la escala de paisaje. Finalmente, el AOO-R corresponde al modelo donde un tipo de ecosistema forestal tiene la probabilidad más alta de presencia entre todos los modelos AOO-RP. La Figura 5 sintetiza el marco conceptual.

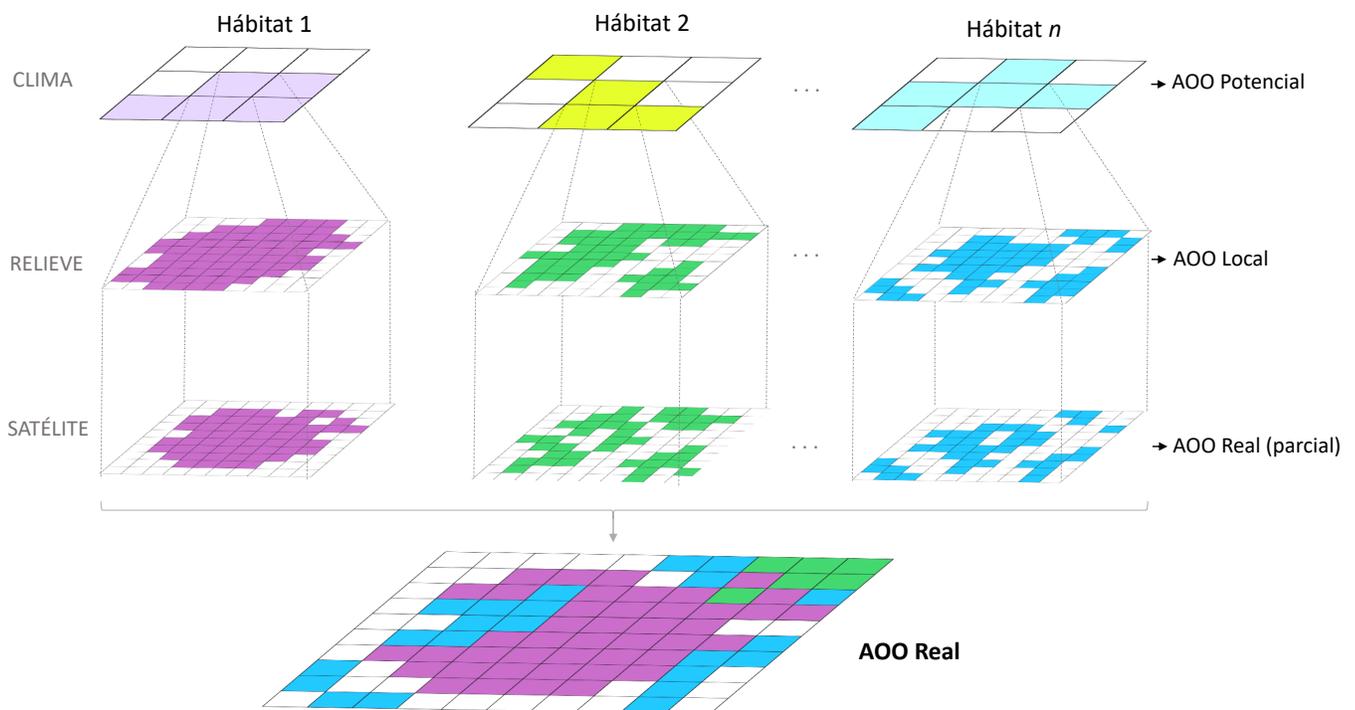


Figura 5: Etapas de modelado de las distribuciones de las áreas de ocupación (AOO). El AOO-P representa las zonas con idoneidad climática para un tipo de ecosistema forestal. EL AOO-L representa un subconjunto del AOO-P donde es más probable que ocurra un tipo de ecosistema forestal determinado considerando rasgos estructurales de la matriz del paisaje. El AOO-R es un subconjunto del AOO-L donde un tipo de ecosistema forestal tiene la mayor probabilidad de presencia considerando datos de Satélite

Para adaptar el marco conceptual de modelización jerárquica a los objetivos del presente trabajo, se propone la metodología ilustrada en la Figura 6. En primer lugar, las variables predictoras (sección 4.2) se procesaron para adecuarlas a las necesidades del modelo (sección 4.3). A continuación, se generaron varios

muestreos aleatorios en las zonas de bosque variando la distancia mínima entre puntos de muestreo, y se seleccionó el que cumplía con un criterio basado en un umbral de autocorrelación espacial de las variables predictoras (sección 4.4.1). Asimismo, se eliminaron posibles puntos erróneos del muestreo mediante un filtrado de altura mínima de la cobertura arbolada (MCH) y un NDVI mínimo, asegurando que zonas de no bosque no fueron incluidas como datos de entrenamiento de bosque en los modelos (sección 4.4.2). Siguiendo la metodología de Álvarez-Martínez y col. (2018), se realizaron tres modelos de distribución para cada tipo de bosque basados en datos de presencia y ausencia del ecosistema: el AOO-P, que predice las zonas con condiciones climáticas idóneas para el tipo de bosque; el AOO-L como un subconjunto del AOO-P donde la distribución del tipo de bosque es más probable considerando variables estructurales del paisaje; y el AOO-PR como un subconjunto del AOO-L considerando datos espectrales provenientes del satélite Sentinel-2 (sección 4.6). Se calculó el modelo final, el AOO-R, asignando a cada píxel una clasificación del ecosistema forestal según la máxima probabilidad entre todos los AOO-RP. También se determinó la diferencia de probabilidad entre el primer y segundo tipo de bosque más probable (incertidumbre, sección 4.7.2) y la sensibilidad del modelo (4.7.3) como medidas de validación.

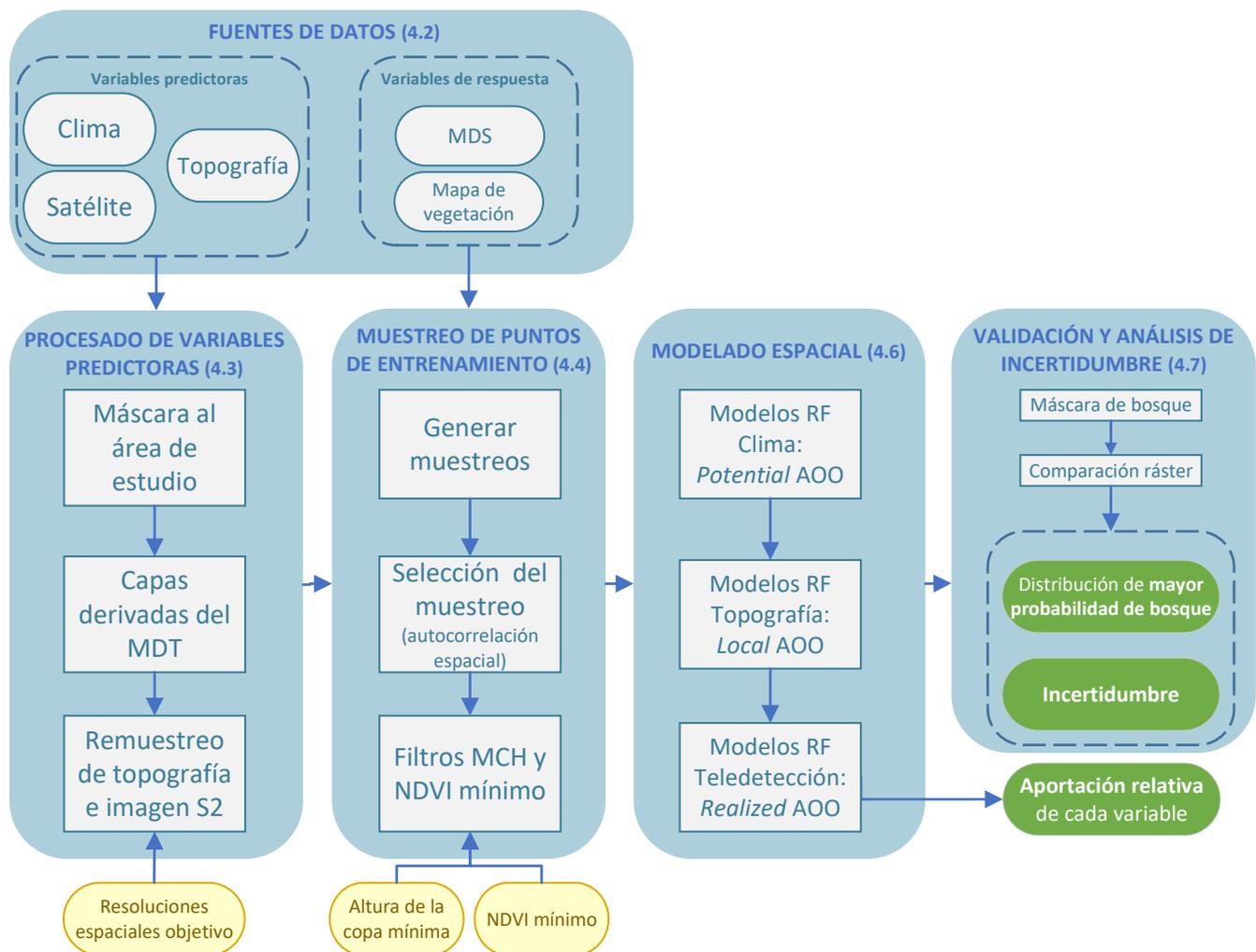


Figura 6: Diagrama del flujo de procesos de la metodología desarrollada. Los datos resaltados en amarillo representan parámetros de entrada ajustables en la interfaz del *plugin* de QGIS, mientras que los datos en verde indican los resultados finales obtenidos

4.2. Fuentes de datos

4.2.1. Variables predictoras

Los datos climáticos provienen del Atlas Climático Digital de la Península Ibérica de Ninyerola y col. (2007). Este atlas contiene 48 capas ráster con una resolución espacial de 200 m para la temperatura media (mínima, media y máxima) y la precipitación de cada mes para la Península Ibérica. Dichos datos fueron generados mediante interpolación espacial a partir de los datos de 2285 estaciones meteorológicas. Los datos topográficos provienen del mapa digital del terreno del Centro Nacional de Información Geográfica (2018) con una resolución de 2 metros (MDT02). Para cubrir el área de estudio, se descargó un mosaico de 14 imágenes de 13.84 x 9.67 km cada una, a partir de las cuales se derivaron el resto de capas topográficas (ver sección 4.3). Tanto el atlas climático como los MDT02 fueron seleccionadas por contar con la mayor resolución espacial disponible.

En cuanto a los datos espectrales, se descargó una imagen del satélite Sentinel-2 nivel L2A (USGS Copernicus Sentinel-2, 2017) del servidor de la Agencia Espacial Europea con fecha del 20 de agosto de 2017. Sentinel 2 cuenta con un generador de imágenes multispectrales (MSI) que ofrece 13 bandas espectrales, de las cuales se han seleccionado las correspondientes al espectro de la luz visible (B2-B4), infrarrojo cercano (B5-B8A) e infrarrojo de onda corta (B11 y B12). La imagen, que cubre la zona de estudio, posee unas dimensiones de 109.8 x 109.8 km y cuenta con una resolución espacial que varía entre 10 y 20 m dependiendo de la banda espectral particular: 10 m para las tres bandas del visible y 20 m para el resto de las bandas². La elección de esta fecha se justifica por ser la imagen disponible más cercana en el tiempo a las fechas fenológicas de los tipos de bosques seleccionados en su momento de mayor actividad fotosintética, según las métricas espectrofenológicas de datos Sentinel-2 de Trullén y col. (2022) en el mismo ámbito geográfico de estudio. Esto significa que la fecha seleccionada corresponde con el momento del año en el que los bosques objeto de estudio presentan mayor cantidad de clorofila, de tal forma que el NDVI, índice sensible a la actividad fotosintética de la vegetación (Pettorelli y col. 2005), permitirá discernir entre coberturas vegetales y no vegetales de forma clara (sección 4.4.2). Otra justificación importante es que en verano se consigue la mayor calidad radiométrica de la imagen debido a la menor presencia de nubes y de proyección de sombra (Chuvieco, 2016). La tabla 1 sintetiza las variables predictoras que alimentan el programa.

Los predictores topográficos permiten explicar las características edafológicas y topográficas, mientras que los predictores espectrales son capaces de detectar variaciones de reflectividad en las copas de los árboles.

Tabla 1: Datos de entrada para la modelización. Los niveles 1, 2 y 3 corresponden a los tres niveles de AOO: potencial, local y real parcial. Las resoluciones espaciales indicadas son las resoluciones originales de cada conjunto de datos, no las resoluciones remuestreadas con las que se llevaron a cabo los modelos

Nivel	Datos	Resolución (m)	Fuente
1	Climáticos	200	Ninyerola y col. 2007
2	Topográficos (MDT02)	2	Centro Nacional de Información Geográfica, 2018
3	Espectrales	10 - 20	“Datos de Copernicus Sentinel-2” (2017)

² A excepción de las bandas 1 (aerosoles), 9 (vapor de agua) y 10 (cirros), que cuentan con una resolución espacial de 60m

4.2.2. Variables respuesta

Las variables respuesta se generaron a partir del mapa de vegetación a escala 1:10.000 del Parque Nacional Picos de Europa (Prieto y Bueno, 2013), que representa la distribución espacial de las principales unidades de vegetación cartografiadas en el área de estudio. Dado que la clasificación de los ecosistemas forestales incluida en este mapa es demasiado detallada para poder ser discriminada por los predictores considerados, se realizó una transcripción a la nomenclatura regional EUNIS (European Nature Information System), un sistema pan-europeo completo y jerárquico para la identificación de ecosistemas forestales. De esta manera, se agruparon las subcategorías de ecosistemas forestales consideradas originalmente en el mapa de vegetación en ocho grupos de bosques bajo la nomenclatura EUNIS (última columna de la Figura 7). Estas agrupaciones de bosques se caracterizan por tener similitudes en cuanto a sus propiedades edafológicas, topográficas y de composición de copas. La Figura 7 ilustra la transcripción desde las unidades ecológicas del mapa de vegetación a los tipos de bosque en nomenclatura EUNIS seleccionados.

Para filtrar puntos de muestreo clasificados erróneamente como bosque y generar una máscara binaria de zonas de bosque y no bosque, se empleó el Modelo Digital de Superficies (CNIG, 2020) con una resolución espacial de 5 metros, que es la de mayor resolución disponible. Se descargó un mosaico compuesto por 4 imágenes de 27.61 km de ancho y 19.27 km de alto que cubren el área de estudio. El cálculo de un modelo de alturas (MDA) de la cobertura de copas³ se realizó a partir del MDS05 (sección 4.4.2). La exclusión del MDA del conjunto de variables predictoras se debe a la morfología altamente irregular y abrupta del PNPE. Al calcular el MDA a partir del MDS y MDE, se introduce una variabilidad en la altura irreal, lo que resulta en la aplicación de ruido al modelo. La Figura 8 presenta un modelo de entidad-relación que resume toda la información de entrada necesaria y cómo se almacena.

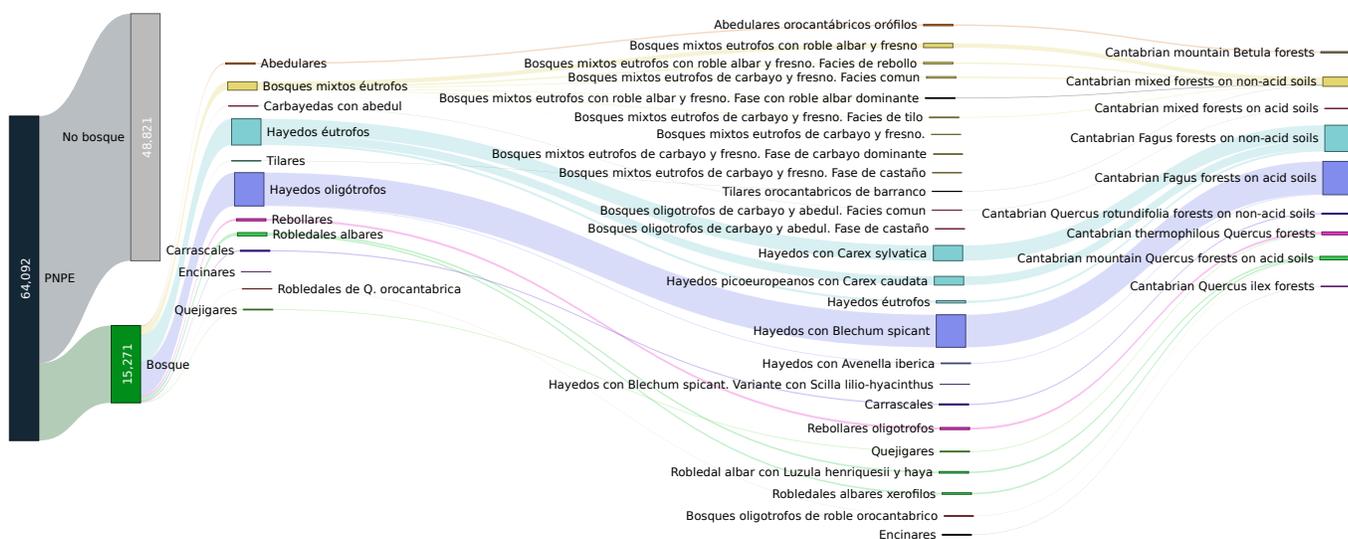


Figura 7: Resumen de la simplificación de bosques del Mapa de vegetación del PNPE a bosques con denominación EUNIS. El tamaño de cada nodo es proporcional a la superficie (ha) correspondiente. De izquierda a derecha, las divisiones siguen el siguiente orden: 1) Superficie total del PNPE → Superficie ocupada por bosques del PNPE → Denominación en castellano de las agrupaciones → Categorías de bosque originales del Mapa de Vegetación → **Categorías agrupadas con denominación EUNIS** (selección de los 9 tipos de bosques a modelar)

³ Canopy model

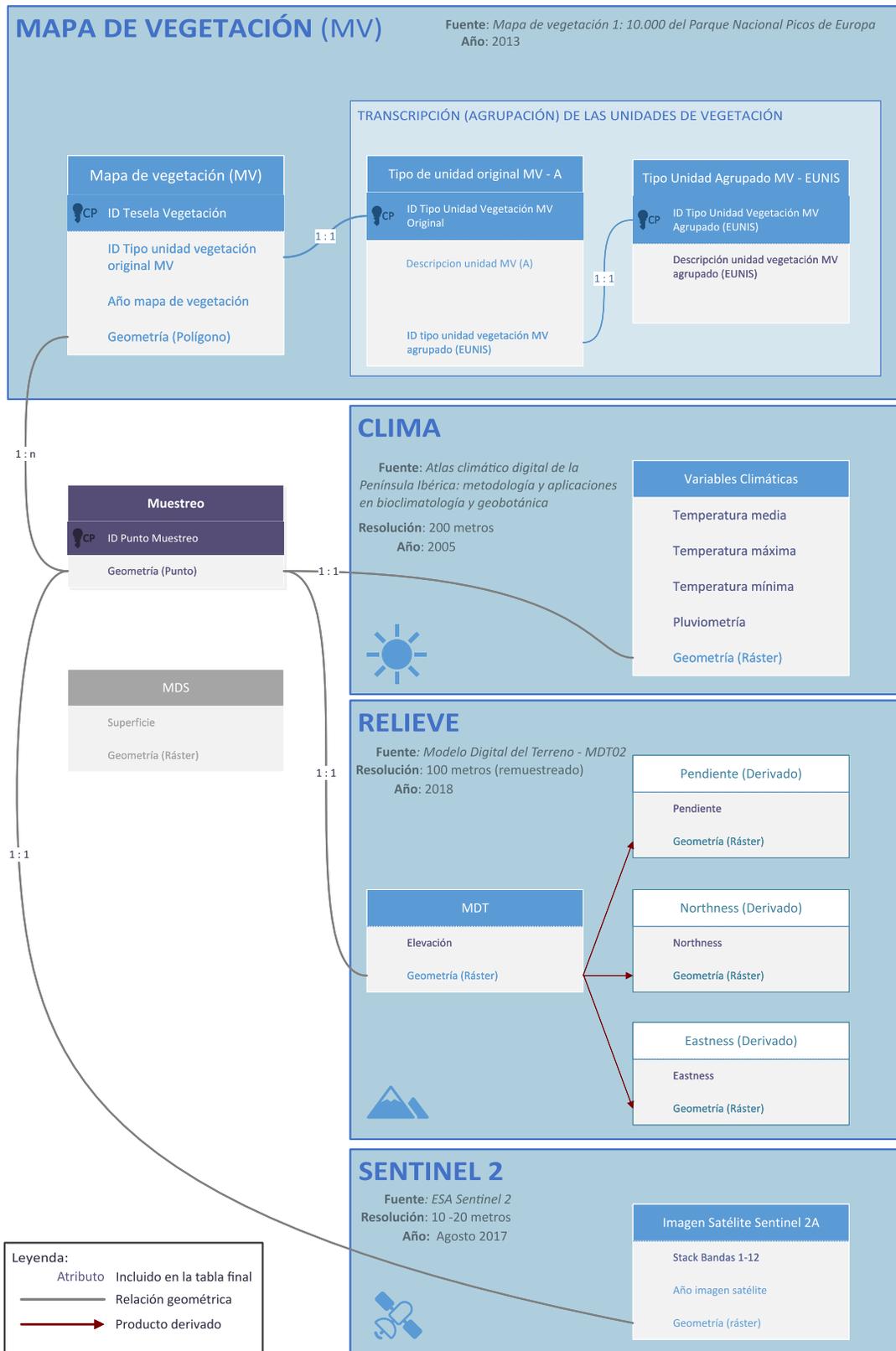


Figura 8: Resumen del esquema del modelo de almacenamiento de las capas de entrada y su interrelación a través de la capa de muestreo. Los atributos destacados en morado indican aquellos que extraídos mediante el muestreo (tabla final). Como se ha explicado anteriormente, el MDS solo intervino en el control de calidad del muestreo combinado con el NDVI

4.3. Procesado de las variables predictoras

La Tabla 2 sintetiza las características finales de las capas de entrada tras el procesamiento. La primera operación que se realizó es una máscara a la zona de estudio, para homogeneizar la extensión de todas las capas, optimizando su representación y reduciendo el tiempo de ejecución total del algoritmo al manejar menos datos. La segunda operación fueron diversos remuestreos, operación que consiste en cambiar la resolución de una capa de información, es decir, el tamaño de cada píxel que la compone. El remuestreo se llevó a cabo ya que cada nivel de modelo se procesa en una resolución específica. En este caso de estudio, la información de satélite posee menor resolución que los datos topográficos. Para cuantificar la aportación de las bandas de satélite mediante un método jerárquico, estas deben ser integradas con el resto de modelos (clima y topografía: AOO-L) en el último nivel (AOO-RP). Esto implica que la resolución de la topografía no puede ser inferior a la de satélite y, por lo tanto, condiciona la resolución de la información topográfica. El tamaño de píxel de la información topográfica debe ser remuestreado a, como mínimo, la resolución de Sentinel-2 (20 m).

Tabla 2: Resumen de las características de las capas de entrada a procesar

Nivel	Tipo	Capas iniciales	Capas finales	Resolución modelado (m)	Resolución original (m)
1	Clima	48	48	200	200
2	Topografía	1	8	20	2
3	Satélite	10	12	20	10-20

$$\Sigma = 59$$

$$\Sigma = 68$$

Las capas de entrada fueron suministradas al complemento de QGIS en el mismo sistema de referencia⁴, ya que han de coincidir con el que el usuario indica al programa para evitar errores. En primer lugar, el complemento comprobó si los datos de topografía eran un solo archivo ráster o una colección de varios ficheros (en caso de tratarse de varios, se llevaría a cabo una combinación automática en un solo fichero). Seguidamente, se realizó la máscara a las 59 capas ráster de entrada. Esta máscara (área de estudio) debe ser suministrada al programa en formato *shape* y en este caso corresponde con los límites del Mapa de Vegetación del PNPE. A continuación, manteniendo su resolución original, se derivaron tres capas a partir del MDT: componente norte (orientación a lo largo del eje Sur-Norte), componente este (orientación a lo largo del eje Oeste-Este) y la pendiente. Después, se remuestreó el MDT y sus 3 capas generadas siguiendo dos métodos de agregación: promedio y varianza ponderado según el área. El método de promedio calcula el valor medio de todas las celdas de entrada cuyos centros se encuentran dentro de la celda de salida. De la misma forma, se calculó la varianza. En ambos casos, intervienen los valores de todas las celdas de entrada que intersectan la celda de salida, ponderados según la proporción de la celda de origen que se encuentra dentro de la celda de salida. Además del promedio, se ha calculado la varianza ya que proporciona información sobre la heterogeneidad espacial local del paisaje. La heterogeneidad de la vegetación y topográfica muestra asociaciones particularmente fuertes con la riqueza de especies (Stein y col. 2014), lo que sugiere que la varianza puede ser un buen predictor de la distribución espacial del ecosistema forestal a modelar. La función de remuestreo por promedio ponderado según área también se empleó para procesar las bandas del visible de Sentinel-2 (B2, B3, B4), ya que estas se encuentran a 10 metros de resolución. Se derivó la capa para el índice NDVI (Rouse y col. 1974) y el ratio SWIR (relación entre bandas 12 y 11). El complemento está programado para que el usuario tenga la capacidad de seleccionar la resolución objetivo en cada nivel de modelo independientemente. Si se considera apropiado, el usuario puede optar por emplear una resolución objetivo menor para la topografía (por ej.: 100 m).

⁴EPSG:25830 en este estudio

4.4. Muestreo de puntos de entrenamiento

El siguiente paso implicó la generación de datos de entrenamiento para los modelos a partir de los predictores procesados. Para lograrlo, se llevó a cabo un muestreo aleatorio cuyas coordenadas se utilizaron para extraer los datos. Se seleccionaron únicamente zonas de bosque como regiones donde muestrear debido a la coexistencia de diferentes tipos de ecosistemas forestales a modelar, lo que permite usar puntos de presencia de un tipo de bosque como ausencias de otro. Esto resultó más eficiente ya que evitó la introducción de posibles puntos de no bosque que podrían ser bosque, permitiendo una mejor gestión del desequilibrio y la superposición de clases, mejorando de este modo el rendimiento predictivo del modelo (Zhang y col. 2019; Valavi y col. 2021). Sin embargo, esto presentó dos problemas a resolver:

- (a) Sin una distancia mínima entre puntos de muestreo, la información extraída podría estar sometida a auto-correlación espacial, lo cual podría llevar a una sobreestimación de la precisión del modelo y a una subestimación de la incertidumbre si no se conoce la naturaleza de la variable (Cruz-Cárdenas y col. 2014). Por otro lado, separar los puntos excesivamente puede llevar a tener insuficientes puntos de muestreo. La distancia mínima entre puntos de muestreo óptima es desconocida.
- (b) Para situar los puntos de muestreo en zonas de bosque únicamente, se seleccionaron las entidades del mapa de vegetación cuya descripción coincidía con la transcripción descrita en la sección 4.2 (Figura 7). Sin embargo, según la información de Prieto y Bueno, 2013, las superficies de estas entidades no corresponden a masas de vegetación continuas, lo cual se explica por la presencia de vías forestales, cortafuegos, claros, etc. que se omiten en el mapa de vegetación. Esto provoca posibilidad de muestrear puntos de no bosque como bosque.

En los siguientes apartados se explica el procedimiento (Figura 9) para abordar estas dificultades. Se seleccionó el mejor muestreo posible cuantificando la correlación espacial de las variables implicadas («muestreo» en la Figura 9), combinándolo con un método de control de calidad («filtros» en la Figura 9) basado en teledetección activa, denominado «altura mínima de la cobertura arbolada» (MCH), y teledetección pasiva (NDVI mínimo), asegurando que todos los puntos de muestreo eran de los tipos de bosque preseleccionados para el estudio (Figura 7).

4.4.1. Muestreo

A partir del mapa de vegetación, se seleccionaron aquellos bosques que coincidían con alguno de los ocho tipos de ecosistemas forestales seleccionados. A continuación, partiendo de una distancia mínima de separación de puntos de 150 m, se generaron muestreos aleatorios dentro de las entidades seleccionadas (bosques) hasta alcanzar una distancia máxima de 1200 m, aumentando en cada interacción cinco metros la distancia mínima, lo cual generó 210 muestreos. En cada muestreo, se generó el máximo número de puntos aleatorios posibles. La menor resolución de las capas de entrada corresponde a la de clima, con 200 m, con lo cual la distancia mínima de muestreo debería ser igual o superior a esa distancia con el fin de evitar que dos puntos se sitúen en un mismo píxel, introduciendo autocorrelación espacial. Sin embargo, se partió de una distancia mínima inferior con el fin de constatar tal efecto en un rango mayor. Se realizó la extracción de datos empleando las coordenadas de los puntos de muestreo en las 68 capas procesadas y se almacenaron en distintos *dataframes*, donde cada fila correspondía a un punto de muestreo con las 68 columnas de atributos para cada variable. Se generó un nuevo *dataframe*, donde cada fila representaba un muestreo y cada columna la autocorrelación espacial calculada mediante la I de Moran (Moran, 1950). A partir de este *dataframe* (Figura 10) se seleccionó la fila (muestreo) cuyas variables presentaban un valor de I de Moran inferior a 0.3. Este umbral se seleccionó mediante criterio experto para este estudio concreto,

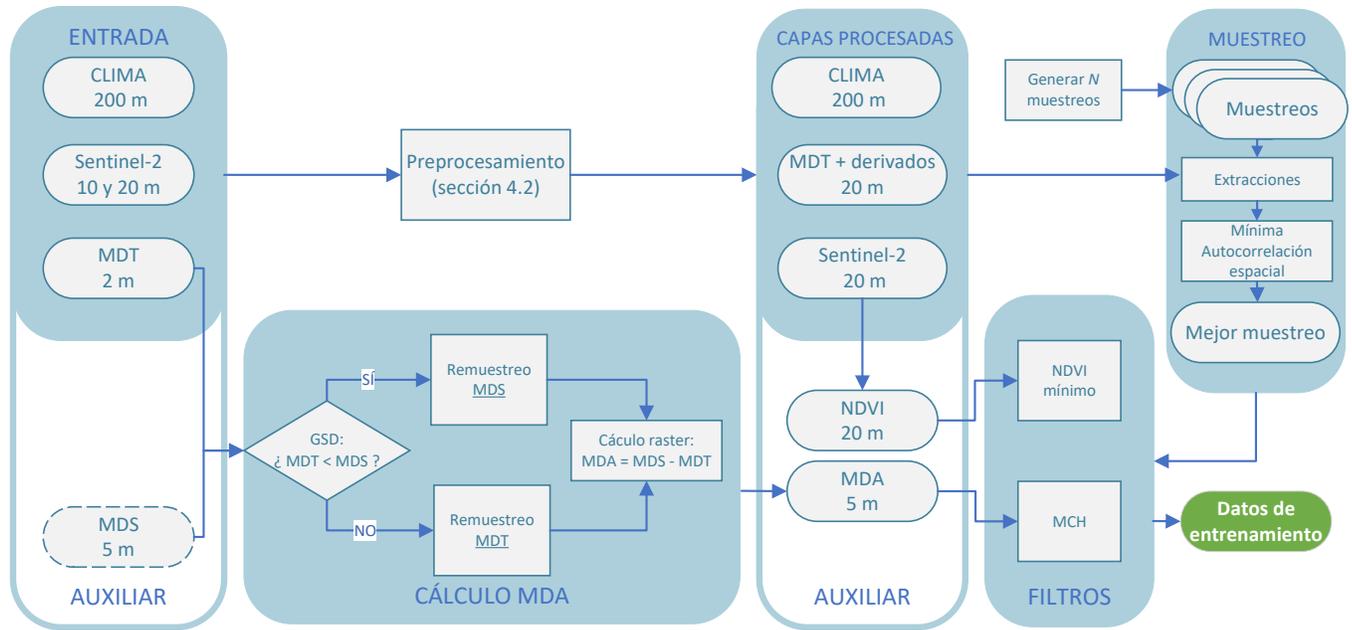


Figura 9: Procedimiento para generar los datos de entrenamiento a partir de los datos de entrada. En el cálculo del MDA (Modelo Digital de Alturas), GSD hace referencia a *Ground Sampling Distance* (resolución espacial). Las capas auxiliares son aquellas que solo intervienen en el control de calidad («Filtros»). «Generar N muestreos» toma como entrada el mapa de vegetación.

variándolo en distintas ejecuciones del programa y eligiendo según la mejor validación, incertidumbre y sensibilidad (sección 4.7) final de los modelos. El umbral podría variar según el ámbito geográfico de estudio y la naturaleza de los datos de entrada. Tal como ilustra la Figura 10, las autocorrelaciones espaciales en la familia de variables climáticas resultaron muy altas, incluso cuando se elevaba la distancia mínima de muestreo hasta 1200 m. La causa principal es que esta capa fue originalmente generada mediante interpolación (Ninyerola y col. 2007), con lo cual la autocorrelación espacial se vuelve una característica intrínseca e inevitable de los datos. De forma similar, la variable correspondiente al MDT remuestreado mediante promedio presentó alta autocorrelación espacial, en este caso porque los píxeles vecinos a 20 metros de resolución donde se distribuyen los bosques presentan valores cercanos ya que estos bosques se ubican generalmente en zonas con pendientes suaves. La Figura 11 muestra la relación esperada entre la autocorrelación espacial y la distancia mínima de muestreo. El complemento desarrollado genera automáticamente este tipo de gráfico para cada variable por dos motivos:

1. Comprobar que las operaciones de geoproceso están funcionando correctamente. Si los gráficos resultantes son consistentes, mostrando una tendencia de autocorrelación claramente decreciente con la distancia mínima de separación entre puntos de muestreo, es improbable que se estén dando fallos.
2. La visualización de esta información (Anexo B) ayuda al usuario a entender la naturaleza de los datos que se están manejando, facilitando la toma de decisiones futuras con el manejo de los mismos.

De este modo, el programa seleccionó el muestreo generado a partir de una **separación mínima de puntos de 280 m**, con un total de 1837 registros.

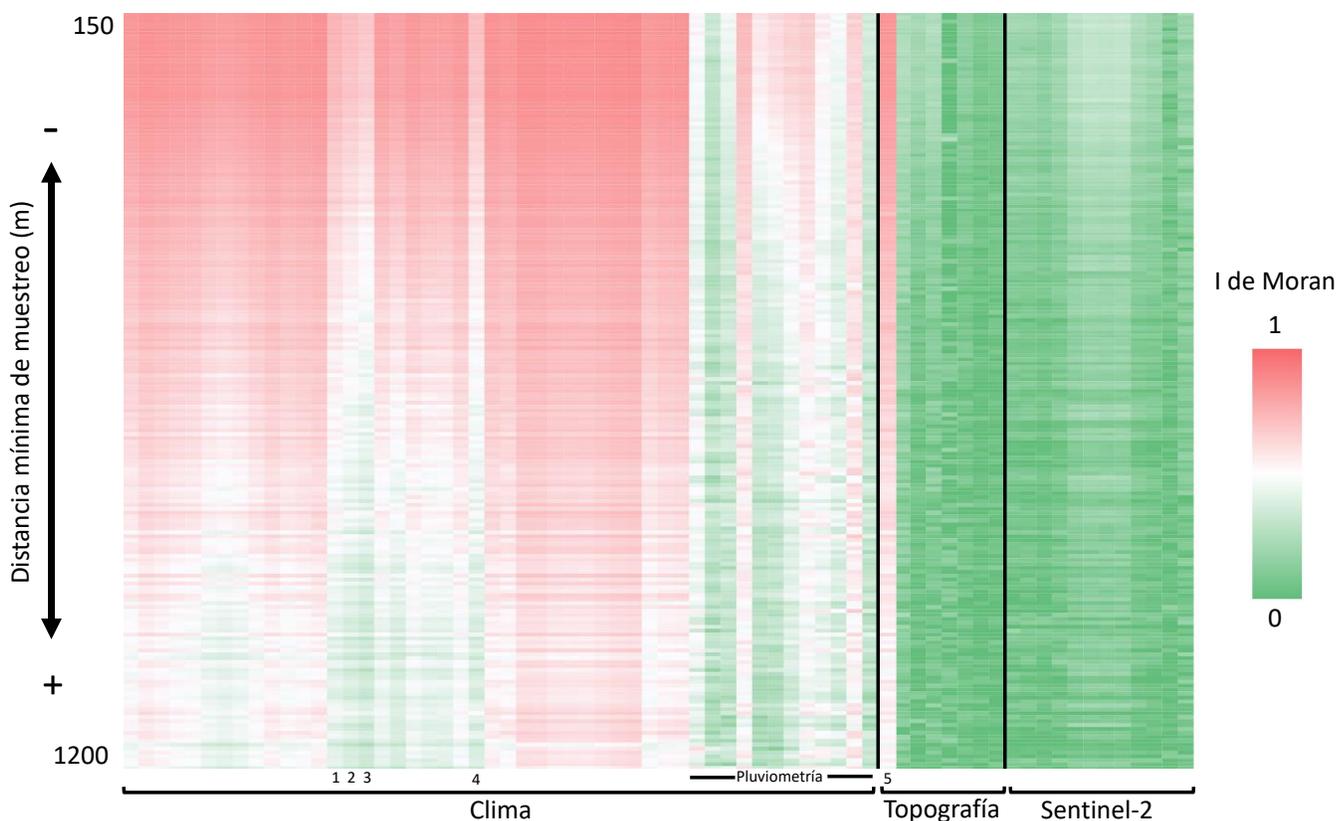


Figura 10: Autocorrelaciones espaciales de todas las variables medidas según la I de Moran para las 210 capas de muestreo generadas. Las bandas de Sentinel-2 y las variables topográficas presentaron bajas correlaciones (a excepción del MDT resmuestreado por el promedio). Las 48 variables climáticas presentaron altas autocorrelaciones espaciales, a excepción de las temperaturas máximas de febrero, marzo, abril y noviembre (1-3, 4), además de la pluviometría

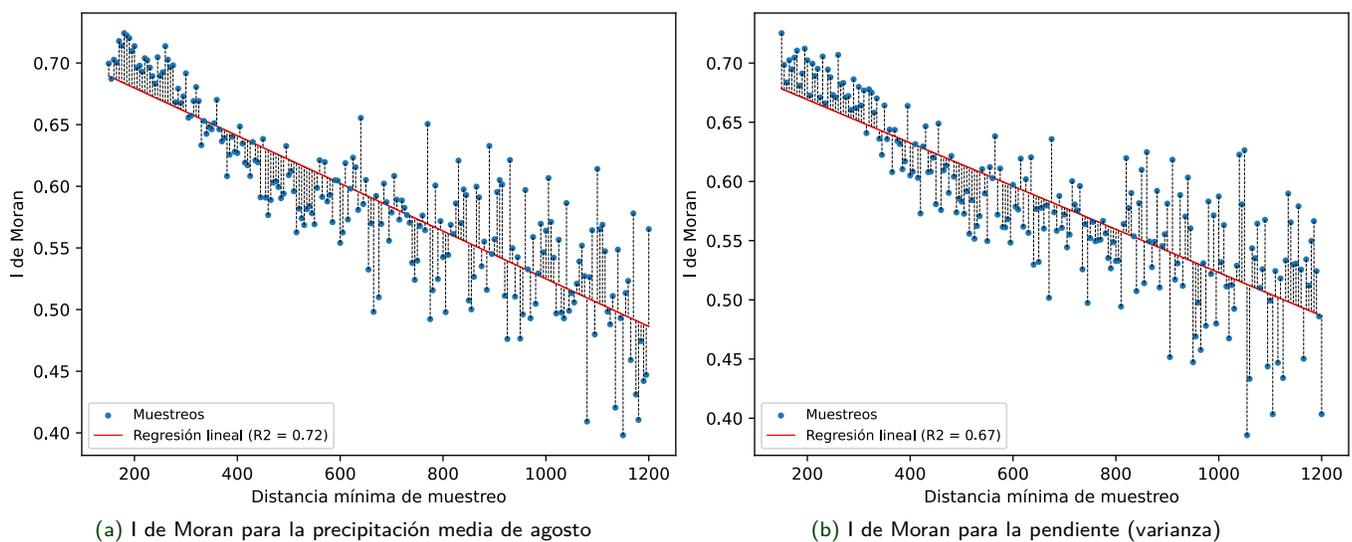


Figura 11: Comportamiento de la autocorrelación espacial (I de Moran) según la distancia mínima de muestreo a partir de una variable climática (a) y otra topográfica (b)

4.4.2. Filtros de calidad: altura mínima de la cobertura arbolada y NDVI mínimo

Con el fin de comprobar en cada punto de muestreo si la altura medida sobre el suelo supera un valor mínimo determinada por el usuario se utilizó el método MCH. Dicho valor mínimo de comparación representa la altura mínima que presenta una cobertura arbórea. En este estudio, se ha fijado el MCH en 3 metros. Para calcular el modelo digital de alturas (MDA), necesario para estimar la altura en cada punto, se calculó la diferencia entre el MDS y MDT. Se programó el complemento para calcular un remuestreo a partir de la capa de mayor resolución entre el MDT y el MDS (método del promedio ponderado) al tamaño de píxel de la de menor resolución y realizar la diferencia (MDS-MDT) a través de un cálculo ráster. La capa resultante se empleó para llevar a cabo las extracciones en los puntos de muestreo y eliminar aquellos registros muestreados en zonas sin la altura mínima requerida. Esto redujo el tamaño de la muestra a 1545 puntos, eliminando 292 (15.9%) de registros no válidos.

Sin embargo, este método presenta una limitación: el MDA no funciona en zonas donde la pendiente presenta alta variabilidad (p. ej.: zonas vecinas a paredes verticales o zonas con afloramientos rocosos irregulares). Esto provoca que el MDA falle por exceso o defecto. Si el error se produce por defecto a un valor por debajo del MCH, no tiene consecuencias graves ya que simplemente se pierde un registro. En caso de que se produzca por exceso, si la altura real es inferior al MCH y el exceso hace que supere el filtro, se seguiría teniendo un punto de muestreo erróneo. Por esta razón se lleva a cabo otro filtro basado en el NDVI mínimo aceptable según la estación fenológica de los bosques modelados. Como se ha comentado anteriormente, se ha escogido la imagen de satélite en una fecha coincidente con la estación fenológica de la vegetación estudiada de máximo verdor a partir de los resultados de Trullén y col. (2022). Se calculó NDVI a partir de las bandas del NIR y del rojo y se eliminaron todos aquellos registros que no presentaban una respuesta espectral de 0.75 en el índice. De esta forma, se redujo el tamaño del muestreo a **1469 puntos**, descartando 76 registros (4.1%) de localizaciones no aptas. Además de evitar puntos de no bosque (Figura 12a), esta metodología también elude puntos de muestreo en zonas de bosque con una baja fracción de cobertura vegetal (Figura 12b), donde el NDVI es inferior al mezclar vegetación y suelo en un mismo píxel de Sentinel-2. Este tipo de cobertura vegetal se excluyó de este trabajo ya que requiere otro tipo de metodología para ser modelado.

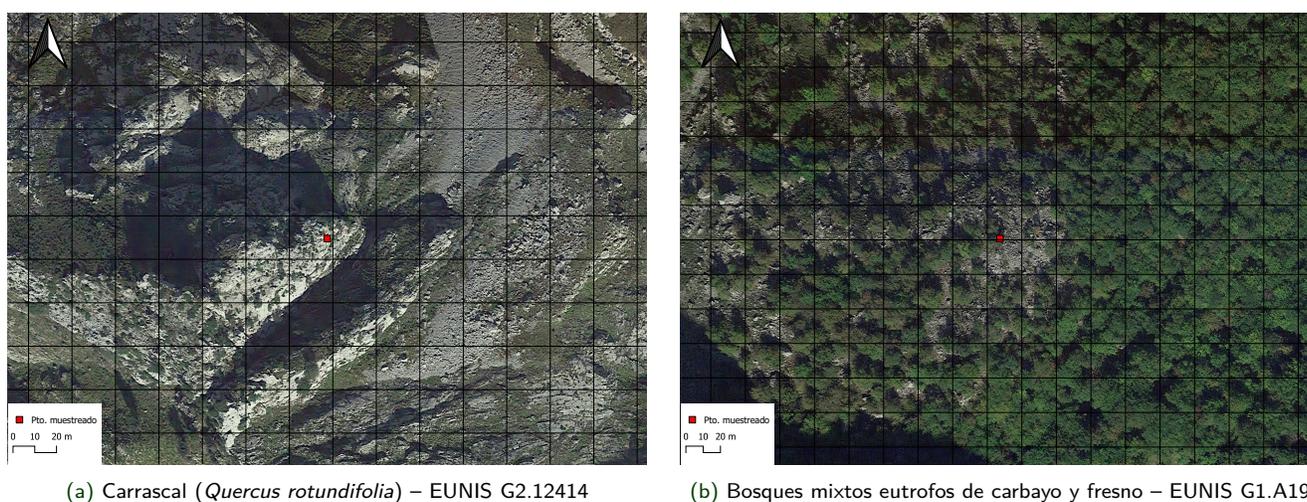


Figura 12: Dos ejemplos de puntos de muestreo filtrados: En (a), la localización superó incorrectamente el filtro MCH debido a la confusión del MDA por la complicación de la orografía, pero el NDVI (0.33) detecta el error. En (b), el punto se ubicó en un punto con mezcla de cobertura arbórea y suelo, también detectado por el NDVI (0.55). La rejilla representa los el *grid* de píxeles de Sentinel-2

4.5. Selección de variables de entrada al modelo

Para las variables climáticas, se seleccionaron aquellas que más limitan la presencia de una especie vegetal: temperatura media mínima del mes más frío, temperatura media máxima de más cálido y precipitación media del mes más seco (ODonnell y Ignizio, 2012). Para determinar estos valores, se ha desarrollado una función que genera el climograma correspondiente para el área de estudio a partir de las 48 capas climáticas, evitando correlaciones estadísticamente significativas (>0.7) entre las variables. Se seleccionaron los predictores de pendiente (media y varianza), MDT remuestreado por varianza y orientación (componente norte y componente sur medias) como variables predictoras topográficas por presentar bajas correlaciones. En términos ecológicos, la orientación determina el grado de insolación y el riesgo a heladas, mientras que la pendiente informa acerca de la capacidad de retención de agua y la disponibilidad de materia orgánica (Colmenero y col. 2005). De forma análoga a Álvarez-Martínez y col. (2018), se seleccionaron la banda 2 (azul), la 3 (verde), la 4 (rojo), la 8 (infrarrojo cercano), las 5, 6 y 7 (borde rojo) junto al índice NDVI (Rouse y col. 1974) y el ratio SWIR. Una vez calculado el AOO-P, esta se convirtió en otra variable predictora para el AOO-P, que también se empleó para predecir el AOO-R. La tabla 3 resume las variables predictoras seleccionadas para el modelo:

Tabla 3: Variables predictoras seleccionadas como entrada al modelo. En la familia de topografía, «media» y «varianza» están referidas al método de remuestreo con la que fueron calculadas. En la familia de satélite, se muestra el valor central de cada banda

Familia	Variable	Unidad	Fuente
Clima	Temperatura mes más frío	°C	Ninyerola y col. (2007)
	Temperatura mes más cálido	°C	
	Precipitación del mes más seco	mm	
Topografía	Pendiente (media)	grados	CNIG (2018)
	Pendiente (varianza)	-	
	Elevación (varianza)	-	
	Componente norte (media)	-	
	Componente este (media)	-	
Satélite	B2 - azul - 490 nm	-	USGS Copernicus Sentinel-2 (2017)
	B3 - verde - 560 nm	-	
	B4 - rojo - 665 nm	-	
	B5 - borde rojo - 705 nm	-	
	B6 - borde rojo - 740 nm	-	
	B7 - borde rojo - 783 nm	-	
	B8 - NIR - 842 nm	-	
	NDVI (Rouse y col. 1974)	-	
	Ratio SWIR	-	

4.6. Análisis de modelado espacial

Para generar cada modelo de cada ecosistema forestal, se evaluaron cuatro métodos de aprendizaje diferentes y se seleccionó aquel con mejor rendimiento como el mejor modelo (sección 4.7.1). En todos los modelos, se utilizaron las variables predictoras previamente seleccionadas. Como variable respuesta se empleó la presencia (valor 1) o ausencia (valor 0) de cada ecosistema forestal en las coordenadas de muestreo. Los distintos métodos de aprendizaje fueron: *Random Forest* (Breiman, 2001) y tres variaciones del mismo llamadas *Extra Trees* (Geurts y col. 2006), *XGBoost* (Chen y Guestrin, 2016) y *LightGBM* (Ke y col. 2017). *Random Forest* es un algoritmo de aprendizaje supervisado que construye múltiples árboles de decisión independientes utilizando subconjuntos aleatorios de variables de entrada. Esto aumenta la robustez del modelo y reduce la probabilidad de sobreajuste a los datos de entrenamiento.

Extra Trees introduce mayor aleatoriedad en la construcción de los árboles al seleccionar puntos de corte aleatorios para cada variable en lugar de buscar el mejor punto de corte. Esto puede disminuir el sesgo del modelo y mejorar su precisión.

XGBoost y *LightGBM* son algoritmos que emplean una técnica conocida como *boosting*, que consiste en crear una secuencia de árboles de decisión que se mejoran mutuamente al corregir los errores cometidos por los árboles anteriores. Estos algoritmos suelen tener un rendimiento superior al de *Random Forest*, pero requieren un ajuste más cuidadoso de los parámetros para evitar el sobreajuste. Estos dos métodos difieren en la forma en que construyen los árboles: *XGBoost* utiliza un algoritmo basado en gradientes que optimiza una función de pérdida definida por el usuario, mientras que *LightGBM* utiliza un algoritmo basado en histogramas que agrupa variables continuas en intervalos discretos. Esto hace que *LightGBM* sea más rápido y eficiente en el uso de memoria que *XGBoost*.

El resultado final de esta clasificación fue un modelo para cada tipo de bosque de bosque que predice la probabilidad de ese ecosistema forestal (AOO-RP). El AOO-R se calculó como la clase con mayor probabilidad de ecosistema forestal. Dado que esta clasificación se extiende a la superficie completa del PNPE, se le aplicó una máscara de zonas de bosque (Figura 13). Dicha máscara consta de valores binarios (bosque y no bosque) combinando la altura mínima de la cubierta (MCH) y el valor mínimo del índice NDVI. Se seleccionaron las áreas con una altura mínima de cobertura de 3 metros y un NDVI superior a 0.75. La máscara binaria se calculó a la misma resolución que la información satelital. Esto difiere de la sección Filtros de calidad: altura mínima de la cobertura arbolada y NDVI mínimo (4.4.2), donde se filtraron los puntos mediante cálculos independientes para cada registro muestreado, mientras que en este paso se calcula una capa ráster de máscara.

La importancia relativa global al modelo AOO-R de cada familia de variables (clima, topografía y satélite) se calculó como el promedio de la importancia relativa de cada ecosistema forestal ponderado según el porcentaje de extensión que ocupa. Con el fin de comparar el modelo resultante, se modificó el código original para calcular una clasificación alternativa (AOO-R no jerárquico) a partir de las mismas variables de respuesta y basado únicamente en datos de satélite Sentinel-2.

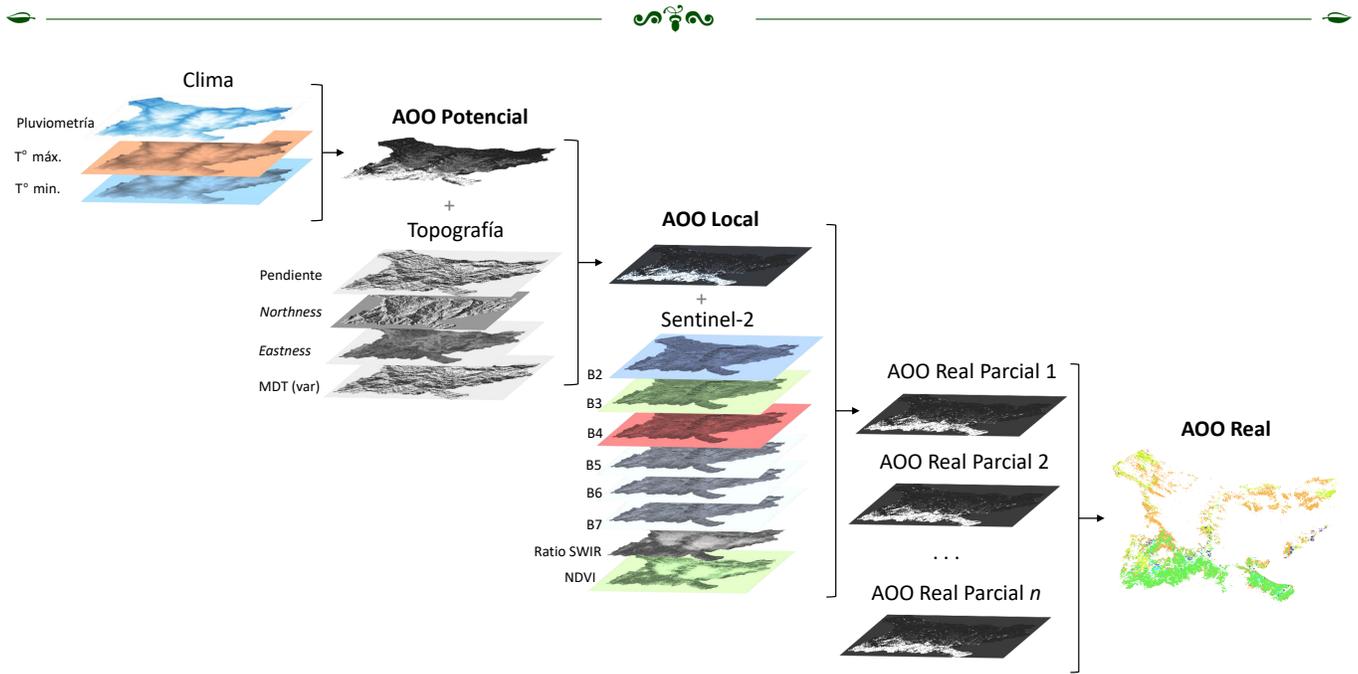


Figura 13: Resumen de la metodología de modelado jerárquico a partir de datos de presencia y ausencias de ecosistemas forestales. Las capas de clima, topografía y Sentinel-2 están a 200, 20 y 20 metros respectivamente. Los modelos AOO potencial, local y real parcial describen probabilidad de presencia para un tipo de ecosistema forestal, mientras que el AOO-R describe el tipo de bosque más probable

4.7. Validación

4.7.1. Validación cruzada

Para seleccionar el mejor modelo de aprendizaje automático entre los cuatro candidatos (Random Forest, Extra Trees, XGBoost o LightGBM), se utilizó un enfoque basado en la validación cruzada. La validación cruzada implica dividir el conjunto de datos de entrenamiento en k subconjuntos, utilizando $k-1$ subconjuntos para entrenar el modelo y el subconjunto restante para evaluar su rendimiento. Este proceso se repitió k veces, utilizando cada vez un subconjunto diferente para la evaluación. Luego, se calculó la media de los puntajes de rendimiento obtenidos en cada iteración para obtener una estimación del rendimiento del modelo. Este proceso se repitió para cada uno de los modelos candidatos y se selecciona el modelo con el mejor puntaje promedio de rendimiento como el mejor modelo. Para este estudio, se empleó un valor $k=10$. Esto significa que el conjunto de datos de entrenamiento se dividió en 10 subconjuntos y se realizó una validación cruzada de 10 iteraciones.

4.7.2. Análisis de incertidumbre

Se calculó la certeza del modelo final como la diferencia de probabilidad entre la primera y segunda clase más probable entre los modelos AOO-RP. Si en un mismo píxel se solapan dos clases de bosque de similar probabilidad, la predicción se considera dudosa. La incertidumbre de la clasificación se calculó como el valor opuesto de la certeza (Álvarez-Martínez y col. 2010), tal y como muestra la Ecuación (1):

$$I = 1 - (P_{C1} - P_{C2}) \quad (1)$$

Donde:

- I = Incertidumbre
- P_{C1} = Probabilidad de la clase más probable entre los AOO-RP
- P_{C2} = Probabilidad de la segunda clase más probable entre los AOO-RP

4.7.3. Matriz de confusión

A partir de la comparación de una rasterización a 20 metros de resolución del mapa de vegetación (Prieto y Bueno, 2013) con el modelo AOO resultante, se calculó la exactitud (AC , Ecuación (2)) y la sensibilidad (TP , Ecuación (3)) del AOO-R en cada clasificación del ecosistema forestal. La exactitud representa por la proporción entre el número de predicciones correctas, mientras que la sensibilidad mide la proporción de casos positivos que fueron correctamente identificados.

$$AC = \frac{TN + TP}{TN + TP + FP + FN} \quad (2)$$

$$TP = \frac{TP}{FN + TP} \quad (3)$$

Donde:

- AC = Exactitud
- TP = Sensibilidad
- TN = Número de píxeles negativos reales⁵
- FP = Número de píxeles falsos positivos⁶
- FN = Número de píxeles falsos negativo⁷
- TP = Número de píxeles positivos reales⁸

La similitud del AOO-R frente al mapa de vegetación se calculó como el número de píxeles del AOO-R coincidentes con los de la rasterización del mapa de vegetación, es decir, el porcentaje de clases (ecosistemas forestales) similares.

⁵Predicción como «no bosque» correcta

⁶Predicción como ecosistema forestal errónea

⁷Predicción como «no bosque» errónea

⁸Predicción como ecosistema forestal correcta

4.8. Complemento QGIS

Toda la metodología que se acaba de describir se aplica de forma automatizada a través de un complemento de QGIS desarrollado para este trabajo (Figura 14). Se introdujeron los datos (variables respuesta y predictoras) explicados en la sección 4.2 (Figura 14a), una altura mínima de la cobertura arbolada (MCH) de 3 metros (sección 4.4.2 – Figura 14b), un NDVI mínimo de 0.75 (sección 4.4.2 – Figura 14c) con resoluciones objetivo de 200-20-20 metros para clima, topografía y Sentinel-2 en los distintos niveles jerárquicos (Figura 14d).

La programación completa del flujo de procesos (Anexo A) en QGIS⁹ se logra mediante la biblioteca PyQGIS en el entorno de desarrollo integrado de QGIS, en combinación con otros módulos como Pandas y GeoPandas (Anexo A.1). PyQGIS es una biblioteca que permite a los desarrolladores escribir códigos y complementos para QGIS utilizando el lenguaje de programación Python. Esta biblioteca proporciona una

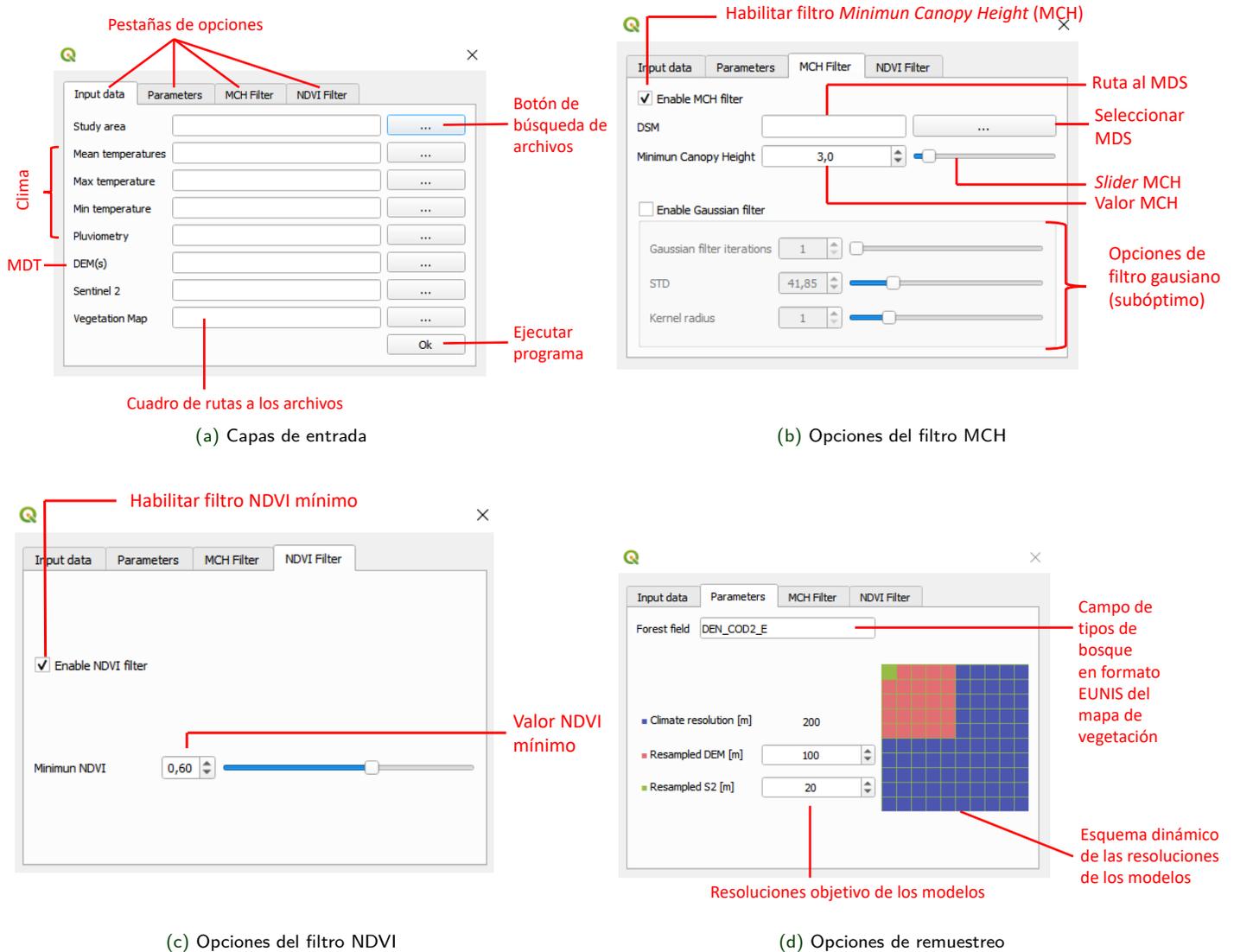


Figura 14: Guía gráfica del complemento desarrollado para QGIS

⁹ Versión 3.28.5-Firenze

amplia gama de funcionalidades para interactuar con QGIS y automatizar tareas de geoprocésamiento. La interfaz gráfica (Figura 14) permite al usuario modificar fácilmente varios parámetros que influyen en el procesado de los datos. Además de generar los resultados finales (modelo AOO-R, incertidumbre y contribución relativa de cada variable), el complemento también genera el proyecto SIG correspondiente, agregando las capas de entrada, generadas y finales en diferentes grupos desde el navegador al espacio de trabajo. El complemento también genera una serie de directorios donde se almacenan todas las capas de datos intermedias y las figuras generadas durante el proceso de forma ordenada. Esto permite al usuario realizar un seguimiento detallado del proceso generado por el programa y aprovechar los datos intermedios para futuros análisis.

El complemento desarrollado calcula el modelo AOO-R y sus incertidumbres, así como un archivo CSV que resume las contribuciones de cada predictor en cada modelo para cada bosque e indica el método de aprendizaje más eficiente. También proporciona información sobre los resultados de cada operación realizada a través de la consola de comandos (p. ej: la separación mínima entre los puntos de muestreo seleccionados). Las capas intermedias generadas por el programa se agregan al espacio de trabajo de QGIS, agrupadas según su origen y función en los modelos, y se les aplican estilos automáticamente (mismo esquema de colores para los mapas de vegetación y la clasificación final, *hillshade* para el MDT como capa base). Cada ejecución del programa crea un nuevo directorio con subdirectorios que contienen todas las capas intermedias y gráficos generados.

Este enfoque proporciona un control total sobre los parámetros de geoprocésamiento utilizados y permite realizar cambios en ellos evitando repetir el trabajo. La automatización del flujo de procesos mejora la eficiencia y precisión del análisis, ya que reduce el riesgo de errores humanos y permite procesar grandes cantidades de datos en poco tiempo. En este tipo de estudio, es esencial contar con la ayuda de la programación debido al gran volumen de datos de entrada y la inviabilidad de realizar todas las operaciones y geoprocésos necesarios manualmente.

5. Resultados

De acuerdo con las aportaciones relativas de las variables predictoras al modelo AOO-R jerárquico, y teniendo en cuenta la extensión de cada ecosistema forestal, el clima explicó el 83.83 % de los patrones de distribución de los ecosistemas forestales analizados. Por otro lado, la información de satélite explicó un 11.44 % y la topográfica un 4.72 %. La Tabla 4 muestra las aportaciones relativas de cada familia de variables en cada ecosistema forestal junto con los resultados de validación cruzada en cada modelo. En todos los casos la contribución relativa del grupo de variables predictoras de satélite fue superior a la correspondiente a los datos topográficos.

Tabla 4: Validaciones cruzadas para cada ecosistema forestal y aportaciones de las familias de variables a la clasificación final

Ec. forestal forestal	Código EUNIS	Validación cruzada (%)			Aportación a AOO-R (%)		
		AAO-P	AAO-L	AAO-R	Clima	Topografía	Satélite
Hayedos oligótrofos	T182	81.20	93.20	96.48	87.85	2.32	9.83
Hayedos eutrofos	T172	79.61	91.04	95.3	80.60	5.94	13.46
Bosques mixtos oligótrofos	T1BX	99.18	100	100	68.57	8.76	22.67
Bosques mixtos eutrofos	T1E1	85.09	95.45	98.73	84.15	6.35	9.50
Abedulares	T1D7	97.94	100	100	74.19	11.21	14.60
Robledales oligótrofos	T1BY	95.06	99.52	100	85.14	5.22	9.64
Carrascales	T21Y	99.79	99.93	100	72.72	8.57	18.71
Robledales termófilos	T197	95.81	99.86	100	74.13	8.83	17.04

Los resultados de las validaciones cruzadas mostraron valores máximos en las clasificaciones de abedulares, robledales oligótrofos, carrascales y robledales termófilos. Se observó un aumento en la puntuación de validación a medida que se desciende en la jerarquía de los modelos según las familias de variables, mejorando cada ajuste en cada jerarquía consecutiva. En la mayoría de los casos, el método de aprendizaje Random Forest obtuvo las mejores puntuaciones de validación cruzada, con excepciones de Extra Trees y LightGBM, pero con una diferencia en ambos casos de solo 0.01 % (Tabla 5).

Tabla 5: Mejor modelo según su validación cruzada (RF: *Random Forest*; LGBM: *LightGBM*; ET: *Extra Trees*)

Bosque	Código EUNIS	Mejor modelo		
		AAO-P	AAO-L	AAO-R
Hayedos oligótrofos	T182	RF	LGBM	RF
Hayedos eutrofos	T172	RF	RF	RF
Bosques mixtos oligótrofos	T1BX	RF	ET	RF
Bosques mixtos eutrofos	T1E1	RF	RF	LGBM
Abedulares	T1D7	RF	RF	RF
Robledales oligótrofos	T1BY	RF	RF	ET
Carrascales	T21Y	RF	RF	RF
Robledales termófilos	T197	RF	RF	RF

El modelo AOO-R jerárquico (Figura 16) mostró una similitud global del 78.06 % en comparación con el Mapa de vegetación (Prieto y Bueno, 2013). Al evaluar la sensibilidad, precisión e incertidumbre de las clasificaciones de ecosistemas forestales individualmente (tabla 6), las clasificaciones de hayedos eutrofos (T182) y hayedos oligótrofos (T172), que representan el 82.44 % de la superficie total arbolada, alcanzaron sensibilidades del 81.55 % y 72.02 % respectivamente. Sin embargo, la sensibilidad de las demás clases de ecosistemas forestales disminuyó significativamente, excepto para los bosques mixtos eutrofos, que representan el 15.39 % de la superficie total.

Tabla 6: Exactitud, sensibilidad e incertidumbre (media y desviación típica) de la predicción de distribución de cada ecosistema forestal del modelo AOO-R jerárquico comparado con el Mapa de vegetación (Prieto y Bueno, 2013)

Ec. Forestales	T182	T172	T21Y	T1BX	T1E1	T1D7	T1BY	T197
Exactitud (%)	84.13	81.03	98.54	99.62	89.23	98.22	95.01	96.42
Sensibilidad (%)	81.55	72.02	5.02	19.04	56.02	15.27	18.34	22.83
Incertidumbre μ (%)	8.38	8.51	0.09	13.08	4.47	0.31	1.46	0.87
Incertidumbre σ (%)	21.01	21.64	2.69	3.20	16.53	4.68	10.87	7.63

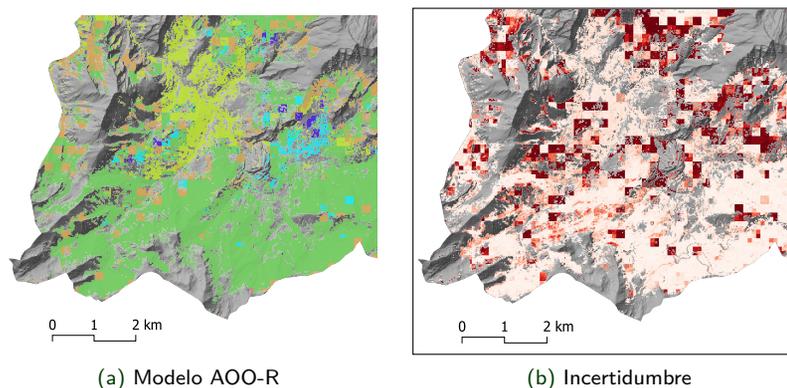


Figura 15: Detalle de píxeles con alta incertidumbre del AOO-R

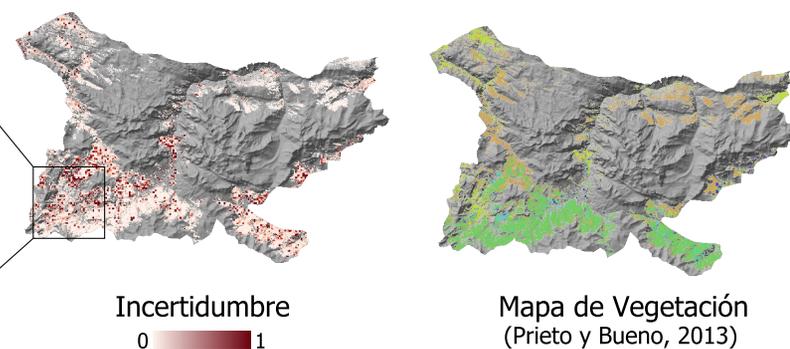
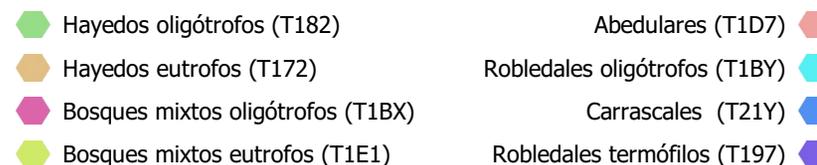
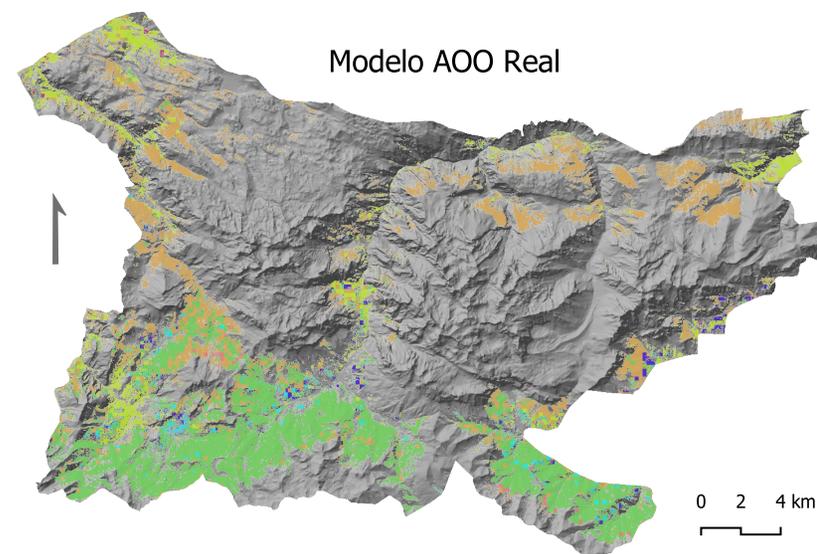


Figura 16: Clasificación AOO-R, Mapa de vegetación (Prieto y Bueno, 2013) e incertidumbre del modelo

La incertidumbre del modelo jerárquico AOO-R se presentó siguiendo un patrón de regiones de 200 x 200 metros (Figura 15, 17.a y 17.b), con una media de 24.25 % y una desviación estándar de 31.25 %. La mediana y el percentil 90 de la incertidumbre se encontraban en el 17.02 % y en el 71.32 %, respectivamente, y su distribución de frecuencias desplazada hacia la izquierda. Al limitar el modelo a las áreas donde la predicción difiere con el mapa de vegetación de referencia, las regiones que presentan incertidumbre se redujeron significativamente (Figura 17.b) y se produjo un aumento relativo en la frecuencia de alta incertidumbre en su distribución. El modelo AOO-R no jerárquico calculado a partir de datos de Sentinel-2 dio lugar a mayor incertidumbre, con una media de 71.31 % y una desviación estándar de 21.53 %, uniformemente distribuida por todos los ecosistemas forestales forestales.

En comparación con el modelo jerárquico, el modelo no jerárquico produjo una disminución media del 17.72 ± 6.20 % en la sensibilidad de las clasificaciones del ecosistema forestal y un aumento progresivo en la incertidumbre (tabla 7). Aunque los valores de incertidumbre fueron similares en ambos modelos para los ecosistemas forestales de menor superficie (Figura 18a), a medida que aumenta la superficie del ecosistema forestal modelado, la incertidumbre se vuelve significativamente mayor en el modelo no jerárquico (Figura 18b). Asimismo, mientras que la clasificación del modelo jerárquico produjo límites claramente definidos entre los diferentes tipos de ecosistemas forestales (Figura 19a), la clasificación del modelo no jerárquico resultó dispersa y sin agrupaciones claras (Figura 19c).

Tabla 7: Exactitud, sensibilidad e incertidumbre (media y desviación típica) de la predicción de distribución de cada ecosistema forestal del modelo AOO-R no jerárquico comparado con el Mapa de vegetación (Prieto y Bueno, 2013)

Ec. Forestales	T182	T172	T21Y	T1BX	T1E1	T1D7	T1BY	T197
Exactitud (%)	70.50	68.77	98.54	99.69	82.26	98.34	95.43	96.26
Sensibilidad (%)	70.35	49.17	0.03	0.56	24.49	1.26	0.02	2.43
Incertidumbre (%)	32.11	27.19	0	0	11.25	0	0.36	0.31
Incertidumbre (%)	37.31	37.12	0	0	28.50	0	5.77	5.34

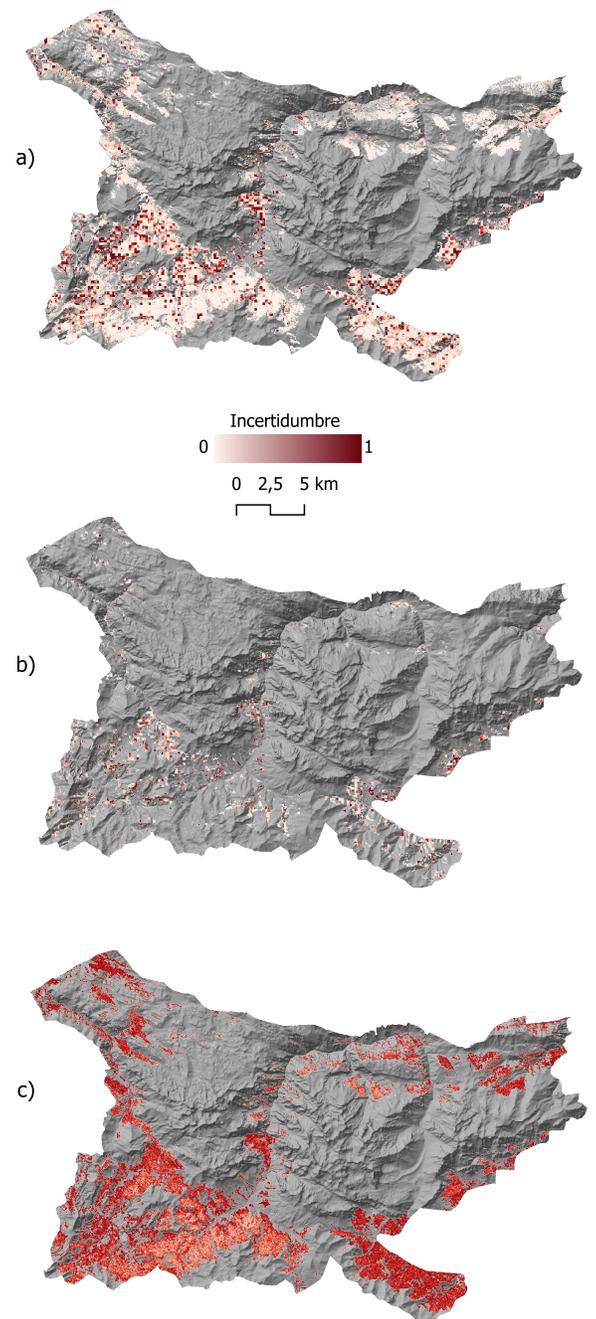


Figura 17: Comparación de la incertidumbre global del modelo AOO-R (a) con la incertidumbre de las regiones del modelo que diferían del Mapa de Vegetación original (b) y la incertidumbre global del modelo AOO-R con solo datos de Sentinel-2 (no jerárquico)

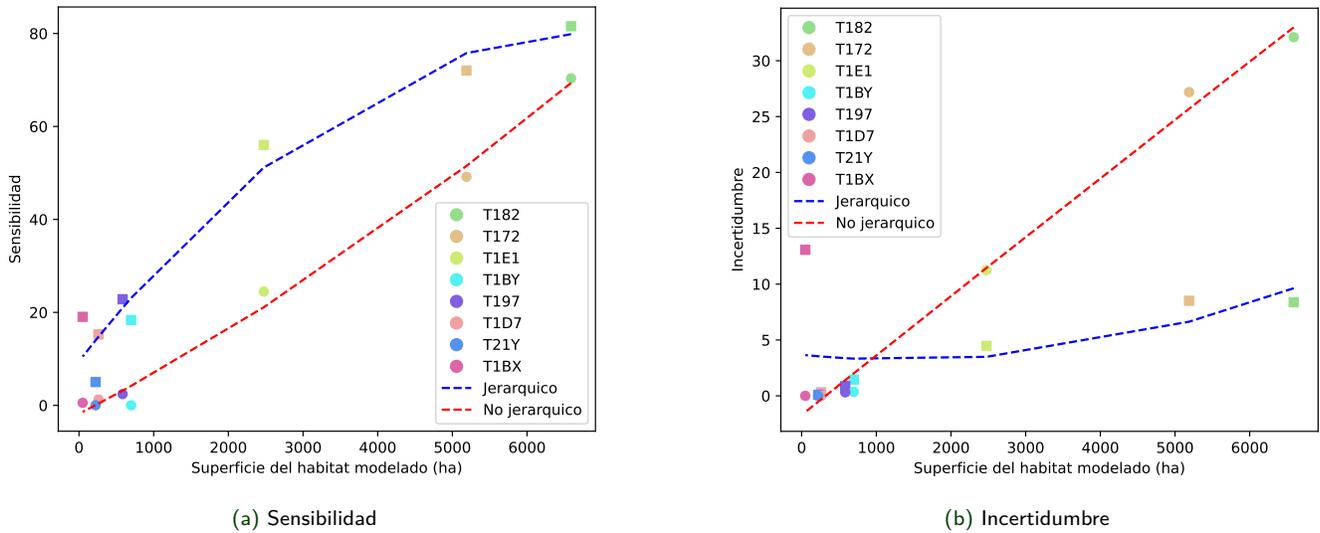


Figura 18: Sensibilidad e incertidumbre en los modelos AOO-R jerárquico y no jerárquico en función de la superficie del ecosistema forestal modelado. Los puntos cuadrados representan registros del modelo jerárquico, circulares del no jerárquico

En el modelo jerárquico AOO-R, las mayores contribuciones relativas de Sentinel-2 se encontraron en los bosques mixtos oligótrofos (T1BX) y en los bosques de *Quercus rotundifolia* eutrofos (T21Y), con 22.67 % y 18.71 % respectivamente (Figura 20). A medida que aumenta la superficie, las contribuciones de las variables de Sentinel-2 tienden a disminuir. No obstante, existen excepciones a esta tendencia, como en el caso de los bosques mixtos eutrofos (T1E1), que contaron con una contribución relativa de Sentinel-2 del 9.50 % y una superficie de 2475 ha. Por otro lado, los hayedos oligótrofos tuvieron una contribución del 9.84 % con una superficie mayor de 6589.25 ha. No se identificó un patrón en la aportación relativa de cada predictor del satélite Sentinel-2 en cada ecosistema forestal, cada tipo de ecosistema contó con aportaciones relativas diferentes. Con todo, el ratio SWIR y el NDVI representaron las mayores contribuciones en la mayoría de los ecosistemas forestales:

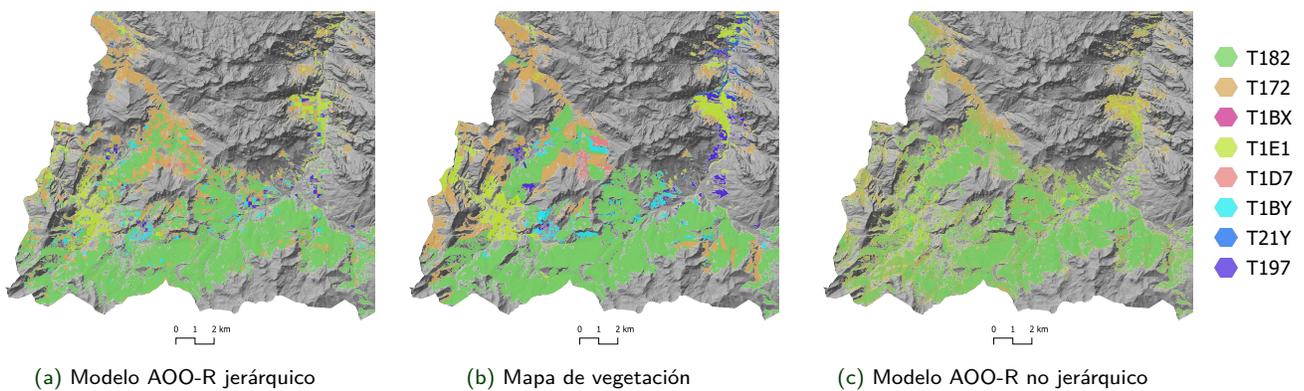


Figura 19: Detalles de los modelos AOO-R jerárquico (a) y no jerárquico (b) junto al mapa de vegetación (Prieto y Bueno, 2013) de referencia – Ámbito geográfico ≡ Figura 15

6. Discusión

Este estudio proporciona una evidencia cuantitativa del potencial de combinar datos del satélite Sentinel-2 con información climática y topográfica mediante una metodología de modelado jerárquico. De esta forma, se logra reducir significativamente la incertidumbre y aumentar la sensibilidad del modelo en comparación con un modelo no jerárquico calculado exclusivamente a partir de datos de Sentinel-2.

Las mayores puntuaciones de validación cruzada en ecosistemas de menor superficie se explican por su menor cantidad de puntos de entrenamiento para el modelo. Como consecuencia, los grupos de validación cruzada no contaron con suficientes registros para obtener una medida confiable. La sensibilidad del modelo, definida en la sección 4.7.3 como la proporción de píxeles de hábitat forestal correctamente identificados, depende de la superficie del hábitat clasificado. Esto se debe a que el área del hábitat determina la cantidad de puntos de entrenamiento disponibles, de acuerdo con la distancia mínima establecida en la sección 4.4.2. En el caso de ecosistemas forestales con pocos puntos de entrenamiento, como los robledales termófilos (T197) o los robledales oligótrofos (T1BY), el modelo jerárquico produce parches de 200 x 200 metros con baja incertidumbre, lo que resulta en un mosaico de ecosistemas forestales con límites de bosques bien definidos, incluso en aquellos bosques con pocos puntos de entrenamiento (Figura 19a). Por otro lado, el modelo no jerárquico demuestra una menor sensibilidad y mayor incertidumbre. Cuando este modelo acierta en la predicción de píxeles de hábitat, estos suelen presentarse aislados o mezclados, especialmente en áreas donde el hábitat modelado está en contacto con otros bosques y ocupa una superficie menor. En comparación con el modelo jerárquico, el modelo no jerárquico da lugar a una clasificación menos precisa y coherente, con mezcla de ecosistemas forestales, límites difusos y ausencia de bosques de menor superficie (Figura 19c). Este efecto se explica por la topografía irregular, que origina un continuo de ecosistemas y condiciones (insolación, pendiente, microclima), generando incertidumbre espectral y errores de clasificación entre los diferentes tipos de ecosistemas forestales (Álvarez-Martínez y col. 2018; Álvarez-Martínez y col. 2010). Por este mismo motivo, la incertidumbre en el modelo no jerárquico aumenta en mayor medida con el tamaño del hábitat modelado en comparación con el modelo jerárquico (Figura 18b). En este ámbito de estudio, el tamaño del hábitat refleja la heterogeneidad del mosaico de hábitat, de modo que a mayor tamaño corresponde un mayor número de bosques contiguos y zonas de transición donde se puede generar más incertidumbre si no se tiene en cuenta las condiciones

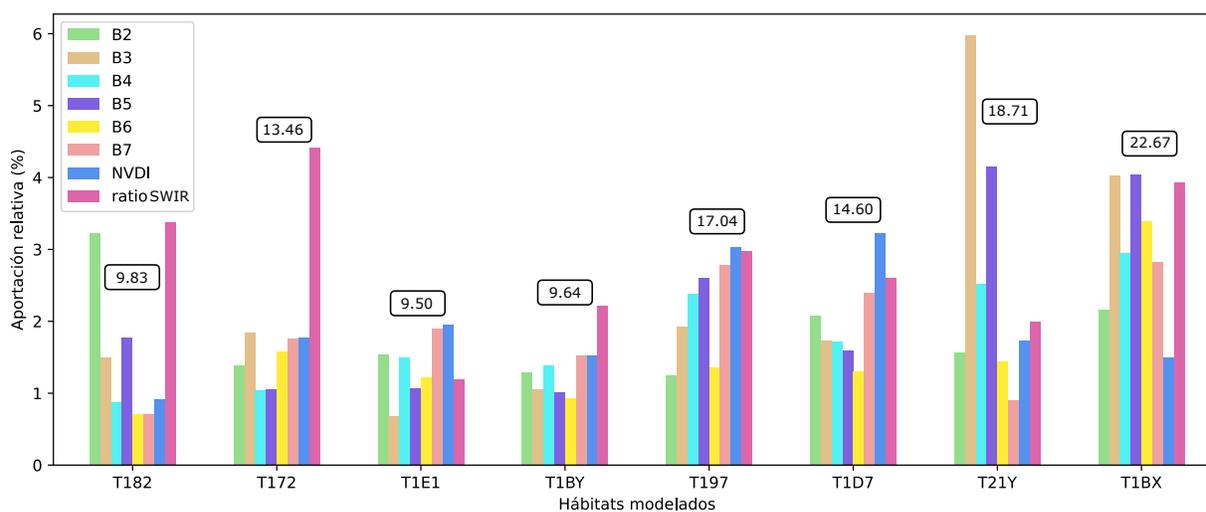


Figura 20: Aportación de las variables de Sentinel-2 seleccionadas para el modelo AOO-R jerárquico según cada ecosistema forestal modelado. Los números en la zona superior indican la aportación total de Sentinel-2 al modelo AOO-R



climáticas y topográficas. El patrón del modelo jerárquico de alta incertidumbre en regiones de 200 x 200 metros (Figura 15b) denota dos posibles fenómenos: (i) Los datos climáticos correspondientes al píxel donde existe alta incertidumbre indican condiciones favorables para la presencia de múltiples tipos de ecosistemas forestales, lo que coincide con una zona de transición donde es plausible encontrar diferentes tipos de bosques. (ii) El dato climático en el píxel con alta incertidumbre podría ser erróneo. Esta interpretación se basa en la metodología seguida por Ninyerola y col. (2007) para generar el atlas climático empleado en este estudio. A pesar de que cuenta con correcciones topográficas, el método de interpolación es sensible a la alta irregularidad del terreno y no tiene en cuenta la influencia de la vegetación en el clima, identificada por Frey y col. (2016), Geiger (1965) y Rosenberg (1974).

Al limitar la incertidumbre a las regiones donde el modelo jerárquico AOO real se desvía del mapa de vegetación (Figura 17.b), se identifican las áreas con mayor probabilidad de errores en el modelado. Esto revela zonas de alta incertidumbre que se corresponden con el caso 2 descrito anteriormente. En este sentido, la metodología desarrollada es aplicable como método de validación de modelos climáticos a partir de datos espacializados de presencia de ecosistemas forestales, topografía y satélite, identificando zonas potencialmente mal interpoladas.

Al evaluar el rendimiento de ambos modelos mediante una validación a partir del mapa de vegetación del PNPE (Prieto y Bueno, 2013) se asume que no se han producido cambios significativos en la distribución de la masa forestal irregular en los cuatro años transcurridos entre la creación del mapa de vegetación y la adquisición de la imagen Sentinel-2 utilizada en el estudio (sección 4.2.2). Esto implica asumir ciertos errores derivados de la simplificación del mapa forestal en los límites de los ecosistemas forestales y de la falta de representación de superficie no arbolada dentro de sus teselas. No obstante, dado que estos errores afectan indistintamente tanto al modelo jerárquico como al modelo no jerárquico, la comparación entre ambos modelos sigue siendo válida. Además, se ha calculado la incertidumbre como medida de índice de confusión (Álvarez-Martínez y col. 2010), cuyo cálculo es independiente del mapa de vegetación de referencia y da lugar a resultados consistentes con el resto de las métricas de validación.



7. Conclusiones

De acuerdo con los objetivos propuestos y las hipótesis planteadas para este trabajo, las conclusiones obtenidas son:

- (i) La herramienta metodológica desarrollada en este estudio ofrece una solución automatizada y eficiente para la identificación de patrones de distribución de ecosistemas forestales en espacios geográficos con interés ecológico. Al permitir la modificación de parámetros y la visualización inmediata de los resultados, esta herramienta proporciona flexibilidad en el análisis y la interpretación de los resultados. Además, al organizar y gestionar automáticamente los datos generados, se facilita enormemente el proceso de revisión y validación de los resultados. La posibilidad de realizar correcciones en la metodología a través de cambios en el código sin necesidad de trabajo manual adicional supone una gran ventaja.
- (ii) El modelo jerárquico basado en datos climáticos, topográficos y espectrales permite obtener una clasificación de los tipos de ecosistemas forestales más precisa, con menos incertidumbre y mayor sensibilidad, que el modelo no jerárquico.
- (iii) Los predictores más relevantes a la hora de predecir las distribuciones de los ecosistemas forestales fueron los climáticos. Los predictores espectrales y topográficos tuvieron mucha menos capacidad predictiva.
- (iv) El análisis de incertidumbre permitió identificar zonas de transición entre ecosistemas forestales o potenciales zonas incorrectamente interpoladas en los datos climáticos de entrada.

Referencias

- Abril, A. D. G., Cañete, J. G., & Irastorza, P. (1996). Las masas irregulares y mixtas: ¿Un objetivo fundamental de la gestión forestal futura? *Cuadernos de la Sociedad Española de Ciencias Forestales*, (3), 139-145.
- Álvarez-Martínez, J. M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A., & Juanes, J. A. (2018). Modelling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution*, 9, 580-593.
- Álvarez-Martínez, J. M., Stoorvogel, J. J., Suárez-Seoane, S., & de Calabuig, E. L. (2010). Uncertainty analysis as a tool for refining land dynamics modelling on changing landscapes: A case study in a Spanish Natural Park. *Landscape Ecology*, 25, 1385-1404.
- Arquer, P. F., & Valderrábano Luque, J. (2007). El Parque Nacional de los Picos de Europa: naturaleza en el entorno de los Lagos de Covadonga.
- Blanco, E., Augier, S., Velázquez, J., García-Abril, A., Rodríguez, R., & González, C. (2012). Distribución y alternativas de gestión de los robledales (*Quercus robur* y *Quercus petraea* (Matts) Liebl.) de Castilla y León.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Centro Nacional de Información Geográfica. (2018). Modelo Digital del Terreno - MDT02.
- Centro Nacional de Información Geográfica. (2020). Modelo Digital de Superficies - MDS05.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.
- Chuvieco, E. (2016). *Fundamentals of Satellite Remote Sensing*. CRC Press.
- Colmenero, G. S., Bastida, J., Zamora, P. J. R., & Alcántara, J. M. (2005). Modelos predictivos de distribución de especies de vegetación potencial en el incendio del "Puerto de Las Palomas"(Sierra de Cazorla, Jaén). *Cuadernos de la Sociedad Española de Ciencias Forestales*, (20), 167-172.
- Cruz-Cárdenas, G., López-Mata, L., Villaseñor, J. L., & Ortiz, E. (2014). Potential species distribution modeling and the use of principal component analysis as predictor variables. *Revista Mexicana de Biodiversidad*, 85, 189-199.
- Datos de Copernicus Sentinel-2 [Recuperado 5/04/2023, procesado por ESA]. (2017).
- Felpete, J. I. A. (2011). *Catálogo florístico del Parque Nacional Picos de Europa*. Jardn Botánico Atlántico.
- Frey, S. J., Hadley, A. S., Johnson, S. L., Schulze, M., Jones, J. A., & Betts, M. G. (2016). Spatial models reveal the microclimatic buffering capacity of old-growth forests. *Science Advances*, 2.
- García, J. M., & González, L. A. (2000). La geología y el relieve de los Picos de Europa. *Naturalia Cantabricae*, (1), 3-19.
- García-González, A., & Goldsmith, F. (1988). Cantabrian mountain beechwoods: a survey and the case for their conservation. *Biological conservation*, 45(2), 121-134.
- Geary, W. L., Bode, M., Doherty, T. S., Fulton, E. A., Nimmo, D. G., Tulloch, A. I., Tulloch, V. J., & Ritchie, E. G. (2020). A guide to ecosystem models and their environmental applications. *Nature Ecology & Evolution*, 4(11), 1459-1471.

- Geiger, R. (1965). The climate near the ground. Har.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63, 3-42.
- Guisan, A., & Thuiller, W. (2005). *Predicting species distribution: Offering more than simple habitat models*.
- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., y col. (2013). Predicting species distributions for conservation decisions. *Ecology letters*, 16(12), 1424-1435.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, 135(2-3), 147-186.
- Hefley, T. J., & Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1, 87-97.
- Kang, E. L., Liu, D., & Cressie, N. (2009). Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models. *Computational Statistics & Data Analysis*, 53(8), 3016-3032.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Keil, P., Belmaker, J., Wilson, A. M., Unitt, P., & Jetz, W. (2013). Downscaling of species distribution models: a hierarchical approach. *Methods in Ecology and Evolution*, 4(1), 82-94.
- Kumar, S., Arya, S., & Jain, K. (2022). A SWIR-based vegetation index for change detection in land cover using multi-temporal Landsat satellite dataset. *International Journal of Information Technology*, 1-14.
- Lengyel, S., Déri, E., Varga, Z., Horváth, R., Tóthmérész, B., Henry, P.-Y., Kobler, A., Kutnar, L., Babij, V., Selikar, A., y col. (2008). Habitat monitoring in Europe: a description of current practices. *Biodiversity and Conservation*, 17, 3327-3339.
- Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, 405(6783), 243-253.
- Moran, P. A. P. (1950). *Notes on Continuous Stochastic Phenomena* (1).
- Ninyerola, M., Pons, X., & Roure, J. M. (2007). Atlas climático digital de la Península Ibérica: metodología y aplicaciones en bioclimatología y geobotánica.
- ODonnell, M. S., & Ignizio, D. A. (2012). Bioclimatic predictors for supporting ecological applications in the conterminous United States. *US geological survey data series*, 691(10), 4-9.
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution*, 20(9), 503-510.
- Prieto, J. F., & Bueno, A. (2013). Mapa de vegetación 1: 10.000 del Parque Nacional Picos de Europa. *Memoria de análisis global de la vegetación del Parque Nacional Picos de Europa - Organismo Autónomo Parques Nacionales, Ministerio de Agricultura, Alimentación y Medio Ambiente. Madrid*.

- Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., y col. (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote sensing of environment*, 239, 111626.
- Rivas, S. (2007). Mapa de series, geoseries y geopermaseries de vegetación de España:[Memoria del mapa de vegetación potencial de España]. Parte I.[Salvador Rivas Martínez y colaboradores]. *Itinera geobotánica*, (17), 5-436.
- Rodríguez, J. P., Keith, D. A., Rodríguez-Clark, K. M., Murray, N. J., Nicholson, E., Regan, T. J., Miller, R. M., Barrow, E. G., Bland, L. M., Boe, K., y col. (2015). A practical guide to the application of the IUCN Red List of Ecosystems criteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1662), 20140003.
- Rosenberg, N. J. (1974). *The biological environment*. John Wiley & Sons Limited.
- Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., y col. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*, 351(1), 309.
- Ruiz de la Torre, J. (2001). Árboles y arbustos de la España peninsular.
- Serrada, R. (1993). *Apuntes de repoblaciones forestales* (inf. téc.).
- Stein, A., Gerstner, K., & Kreft, H. (2014). *Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales*.
- Trueba, R. A., & Navarro, E. G. (2012). Interpolación en zonas de montaña. Clima del Parque Nacional de Picos de Europa. *Revista Tiempo y Clima*, 5(35).
- Trullén, D. L., Álvarez-Martínez, J. M., Labrador, J. D. S., Jiménez-Alfaro, B., Pérez-Silos, I., Hernández-Romero, G., & Barquín, J. (2022). Espectrofenología con datos Sentinel 2: definición de curvas de referencia para la caracterización de ecosistemas forestales. *Ecosistemas*, 2411.
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Aroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, 44, 1731-1742.
- Zhang, L., Huettmann, F., Liu, S., Sun, P., Yu, Z., Zhang, X., & Mi, C. (2019). Classification and regression with random forests as a standard method for presence-only data SDMs: A future conservation example using China tree species. *Ecological Informatics*, 52, 46-56.