



Universidad  
Zaragoza

Trabajo Fin de Grado

# Estudio y análisis de las poblaciones de *Quercus suber* de la Sierra de la Virgen

Modelos de predicción con machine learning: árbol de decisiones

Autora: Verónica García Magalhaes

Director: Dr. Luis Alberto Longares Aladrén

Facultad de Filosofía y Letras  
Departamento de Geografía y Ordenación del Territorio  
2021

## Resumen:

*Los ejemplares de Quercus suber (alcornoque) presentes en la provincia de Zaragoza se desarrollan en su límite ecológico, lejos de los principales núcleos de distribución. Esta característica les otorga una importancia genética reseñable que pone de manifiesto su necesidad de conservación y desarrollo. Con este como objetivo principal, es necesaria una intervención sobre el terreno que favorezca una regeneración natural atendiendo tanto a variables físicas como climáticas. En este trabajo se aplican modelos de predicción que emplean el aprendizaje automático (Machine Learning) para el estudio del comportamiento y desarrollo de esta población vegetal que sirvan de apoyo a la toma de decisiones en caso de intervención.*

**Palabras clave:** *Quercus suber, Sestrica, Aniñón, aprendizaje automático, modelos de predicción, árbol de decisiones*

## Abstract:

*The Quercus suber (cork oak) specimens present in the province of Zaragoza develop in their ecological limit, far from the main distribution centers. This characteristic gives them a remarkable genetic importance that highlights their need for conservation and development. With this as the main objective, an intervention on the ground is necessary that favors natural regeneration, taking into account both physical and climatic variables. In this work, prediction models that use machine learning are applied to study the behavior and development of this plant population to support decision-making in case of intervention.*

**Key words:** *Quercus suber, Sestrica, Aniñón, machine learning, prediction models, decision trees*

<b>1. INTRODUCCIÓN</b> .....	<b>5</b>
<b>1.1</b> Introducción general.....	<b>5</b>
<b>1.2</b> Antecedentes y justificación.....	<b>5</b>
<b>1.3.</b> Biología de la especie.....	<b>5</b>
<b>1.4.</b> Distribución geográfica .....	<b>6</b>
1.4.1 Área de estudio .....	7
<b>1.5</b> Caracterización ecológica .....	<b>9</b>
1.5.1 Parámetros fisiográficos .....	9
1.5.1.1 Altitud media.....	9
1.5.1.2 Pendiente .....	11
1.5.1.3 Orientación .....	12
1.5.1.4 Litología.....	12
1.5.1.5 Edafología.....	13
1.5.2 Parámetros climáticos.....	14
<b>2. OBJETIVOS</b> .....	<b>14</b>
<b>2.1</b> Objetivo general .....	<b>14</b>
<b>2.2</b> Objetivos específicos .....	<b>14</b>
<b>3. METODOLOGÍA</b> .....	<b>14</b>
<b>3.1</b> Delimitación área de estudio.....	<b>14</b>
<b>3.2</b> Modelos.....	<b>14</b>
3.2.1 Machine Learnig.....	14
3.2.2. Árbol de decisión.....	16
<b>4. RESULTADOS</b> .....	<b>17</b>
<b>4.1</b> Reducción de dimensionalidad .....	20
<b>4.2</b> Modelo .....	21
<b>5. CONCLUSIÓN</b> .....	<b>25</b>
<b>6. BIBLIOGRAFÍA</b> .....	<b>26</b>
<b>7. ANEXO</b> .....	<b>28</b>
<b>7.1</b> Mapas .....	<b>28</b>
<b>7.2</b> TABLAS .....	<b>36</b>
<b>7.3</b> OTROS.....	<b>36</b>

## ÍNDICE DE ILUSTRACIONES

Ilustración 1 Imagen del fruto, corteza y hoja. Fuente: arbolapp.es .....	6
Ilustración 2. Perfil 1. Elaboración propia.....	9
Ilustración 3. Perfil 2. Elaboración propia.....	9
Ilustración 4. Altitud media. Aladren et al. nd p 5. En color rojo aparecen las parcelas de Sestrica y Aniñón.....	10
Ilustración 5 Pendiente en %. Aladren et al. nd p 8. En color rojo aparecen las parcelas de Sestrica y Aniñón.....	11
Ilustración 6. Conceptos de subajuste, apropiado y sobreajuste. Elaboración propia .....	16
Ilustración 7. Agrupación de datos según salidas. Elaboración propia .....	19
Ilustración 8. Matriz de correlación. Elaboración propia .....	20
Ilustración 9. Árbol de regresión de 5 niveles. Elaboración propia .....	22

## ÍNDICE DE MAPAS

Mapa 1-1 Mapa de distribución en España.(Diaz-Esteban, M., 2009, p. 9) .....	6
Mapa 1-2. Mapa de localización. Elaboración propia .....	7
Mapa 1-3. Mapa de vegetación. Elaboración propia .....	8
Mapa 1-4. Presencia de Q. suber y perfiles realizados. Elaboración propia .....	8
Mapa 1-5. Mapa de Altitudes. Elaboración propia .....	10
Mapa 1-6. Mapa de Pendientes. Elaboración propia.....	11
Mapa 1-7. Mapa de Orientaciones. Elaboración propia.....	12
Mapa 1-8. Mapa de Unidades Morfolitológicas. Elaboración propia.....	13
Mapa 4-1. Mapa de localización final. Elaboración propia.....	25
Mapa 4-1. Mapa de localización final. Elaboración propia.....	25

## ÍNDICE DE TABLAS

Tabla 1. Presentación de datos. Tabla completa en el anexo. Elaboración propia.....	17
Tabla 2. Datos de entrada. Elaboración propia. Tabla completa en el anexo .....	18
Tabla 3. Tabla de entrenamiento. Tabla completa en anexo. Elaboración propia .....	23
Tabla 4. Tabla de predicción. Tabla completa en anexo. Elaboración propia .....	24

# 1. INTRODUCCIÓN

## 1.1 Introducción general

El presente trabajo analiza el espacio donde se localiza las poblaciones de *Quercus suber* de la provincia de Zaragoza, en las estribaciones de la Sierra de la Virgen, en concreto en los municipios de Sestrica y Aniñón, con la finalidad de establecer una posible intervención sobre el terreno para aumentar la presencia de esta especie en la zona.

Primero se lleva a cabo un análisis de los factores que influyen en la presencia de la especie en el territorio, para después establecer un modelo de predicción que facilite la toma de decisiones de cara a una posible repoblación.

## 1.2 Antecedentes y justificación.

Uno de los problemas medioambientales más importantes a nivel mundial, es la pérdida de biodiversidad, siendo esta definida como “la variabilidad entre los organismos vivientes de todas las fuentes, incluyendo, entre otros, los organismos terrestres, marinos y de otros ecosistemas acuáticos, así como los complejos ecológicos de los que forman parte: esto incluyen diversidad dentro de las especies, entre especies y de ecosistemas” (“Convention on biological diversity,” 1998, p. 1). Ante esta problemática se hace necesaria la puesta en marcha de planes de conservación, principal herramienta para la lucha por la pérdida de la biodiversidad.

Los bosques son una fuente de biodiversidad muy importante, al servir de refugio y alimento a gran número de especies, siendo uno de los ecosistemas que mayor variabilidad de estas albergan, de ahí la importancia de su conservación.

La presencia de un alcornoque en esta zona es una peculiaridad natural asombrosa que, junto a otras como que sea un enclave de nidificación de aves poco comunes como el Halcón abejero (*Pernis apivorus*), hacen de esta comarca del Aranda un lugar singular. (Hernández, J. Millán J. y Serra, A., 2007).

## 1.3. Biología de la especie

El conocido como bosque mediterráneo o Durisilva se encuentra vinculado al clima mediterráneo, con poco frecuentes y nunca profundas heladas, veranos cálidos, precipitaciones en torno a los 500mm (generalmente por debajo), invernales, equinocciales e irregulares.

En la Península Ibérica, y debido a su relieve, estas condiciones se distribuyen a lo largo de las depresiones del Guadalquivir y del Ebro, continentalizándose, denominándose clima mediterráneo continental, siendo también dominio de la Durilignosa (Rubio Recio, J.Manuel, 1989, p. 33).

Si bien el mayor exponente de la Durisilva, sin duda es el *Quercus ilex*, encina, también es significativa la importancia del *Quercus suber*, el alcornoque.

*Quercus suber* se caracteriza por sus hojas subpersistentes, con un ciclo de entre 13 y 23 meses (Rodríguez C., 2009), aovadas, enteras o dentadas. El alcornoque es un árbol inerme, fuerte y robusto que puede medir hasta 20 metros de altura en estado natural. Con raíces profundas, que se desarrollan de forma multidireccional. Es una especie tremendamente intervenida por los usos que de ella se realizan, por lo que es difícil ver en estado natural. Posee una corteza en principio lisa pero que pronto comienza a salirle prominencias lineales (Rodríguez C, 2009) agrietadas. Es el primer corcho, denominado bornizo. Este no se puede explotar, debe de ser extraído y descartado. Entre ocho y diez años después de esté primer descorche comenzará la explotación de estos árboles. En formaciones intervenidas sus copas alcanzan poco más de 4 metros, mientras que sus ramas son sinuosas y alargadas, pero en formaciones más salvajes se presentan erguidas. Esto puede dar una idea de lo cambiante que puede ser en estado natural.

La floración es difusa, en climas benignos incluso continua, empezando en abril y pudiéndose alargar hasta bien entrado el verano. Con amentos masculinos muy numerosos, en grupos de 5-6 y las flores femeninas pedunculadas y aisladas, dan lugar a bellotas de forma alargada, con la punta vellosa y un endocarpo libre de pelo.

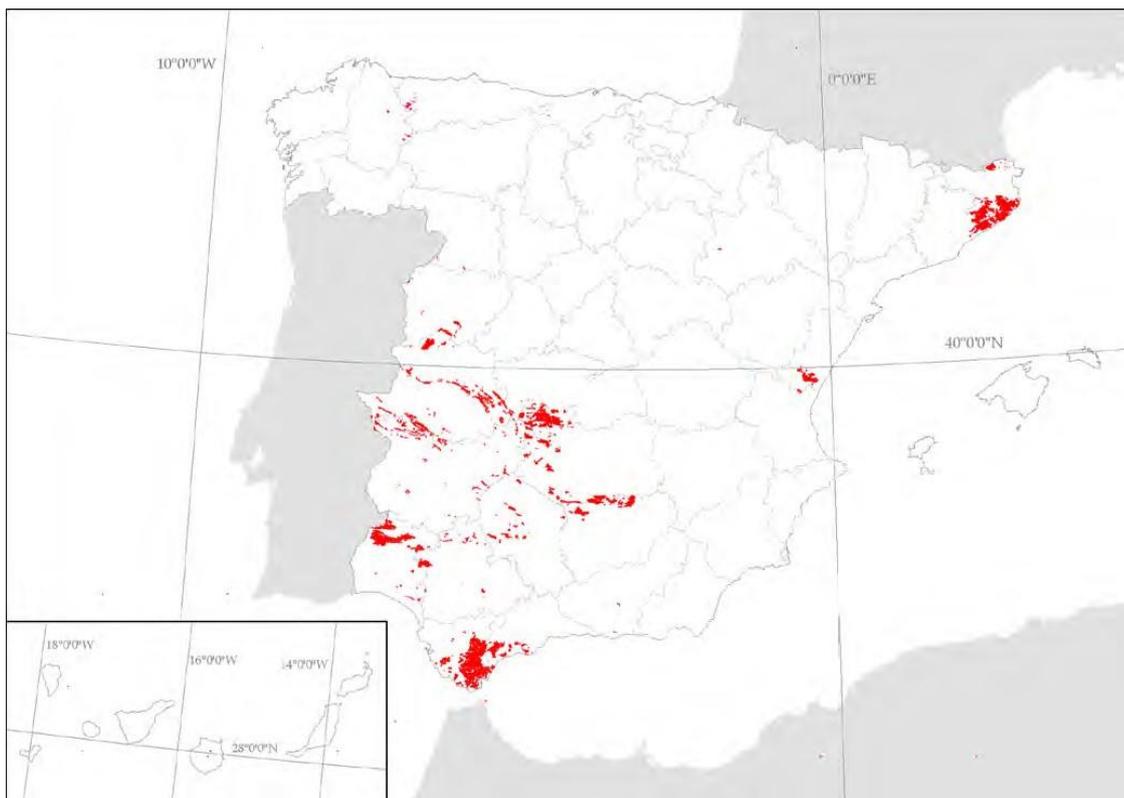


Ilustración 1 Imagen del fruto, corteza y hoja. Fuente: arbolapp.es

#### 1.4. Distribución geográfica

Especie propia de la región mediterránea, tiene su máxima representación en Portugal, seguido de Argelia y España. Los alcornoques en España son en su mayor parte (93%) (Blanco Castro, 2005, p. 300) de propiedad privada, siendo su máxima distribución en la zona suroeste del país.

De forma general, se puede decir que el alcornoque se distribuye en altitudes hasta los 1000m, con precipitaciones superiores a los 600mm y suelos silíceos. (Díaz-Esteban, M., 2009).



Mapa 1-1 Mapa de distribución en España. (Díaz-Esteban, M., 2009, p. 9)

Como se puede observar en el mapa de 1-1 de distribución del alcornoque en España, casi todos los ejemplares se dan en la zona suroeste y noreste. Hay algunos puntos representativos en Galicia, así como otros al este de España y de forma marginal en Zaragoza.

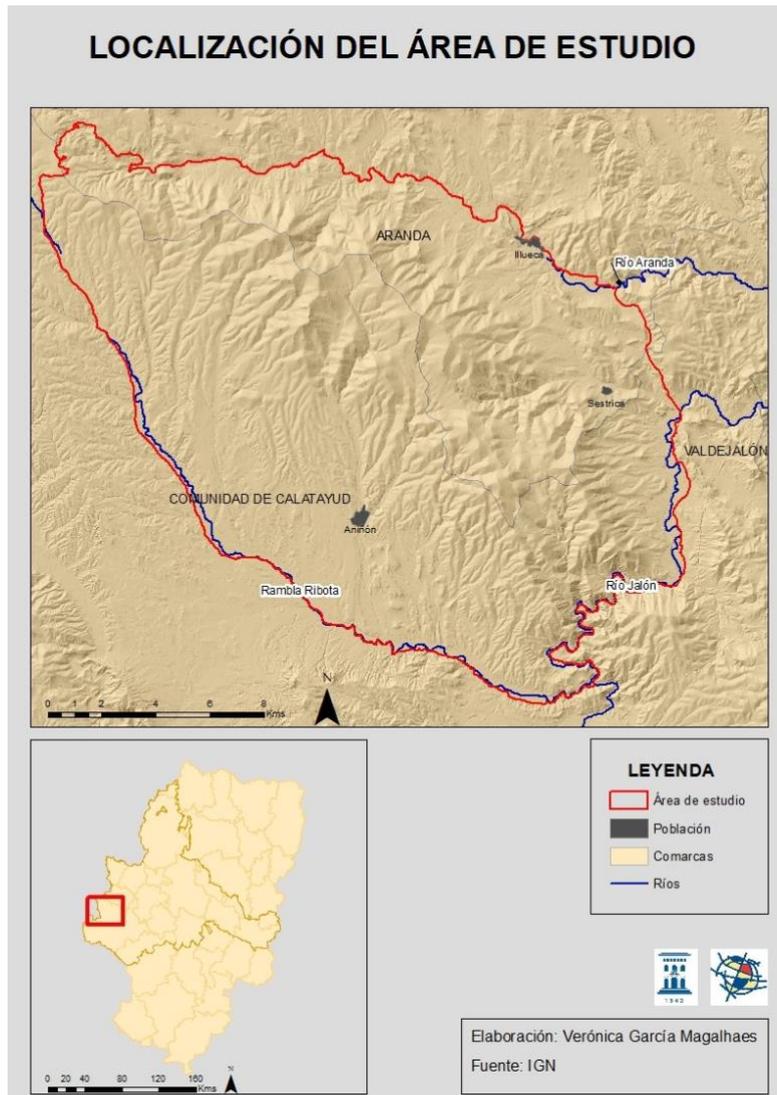
En el análisis territorial de este hábitat en la bioregión mediterránea, su estado de conservación se considera excelente. Sin embargo, la superficie total de este hábitat se distribuye en una única tesela de 410,30 ha y

En el análisis territorial de este hábitat en la biorregión mediterránea, su estado de conservación se considera excelente. Sin embargo, la superficie total de este hábitat se distribuye en una única tesela de 410,30 ha y con una cobertura baja de apenas el 12,5%, lo que nos da una idea de la calidad de esta formación en Aragón. Aparte de esa mancha relevante, que ha sido cartografiada, existen otros rodales de menor entidad que no han sido cartografiados y que presentan múltiples amenazas.

#### 1.4.1 Área de estudio

Los alcornoques analizados en este estudio se encuentran en Aniñón y Sestrica, en la provincia de Zaragoza, dentro de la conocida como Sierra de la Virgen. Esta Sierra representa uno de los bloques elevados del sistema ibérico de edad paleozoica, caracterizada por su alienación ibérica armónica.

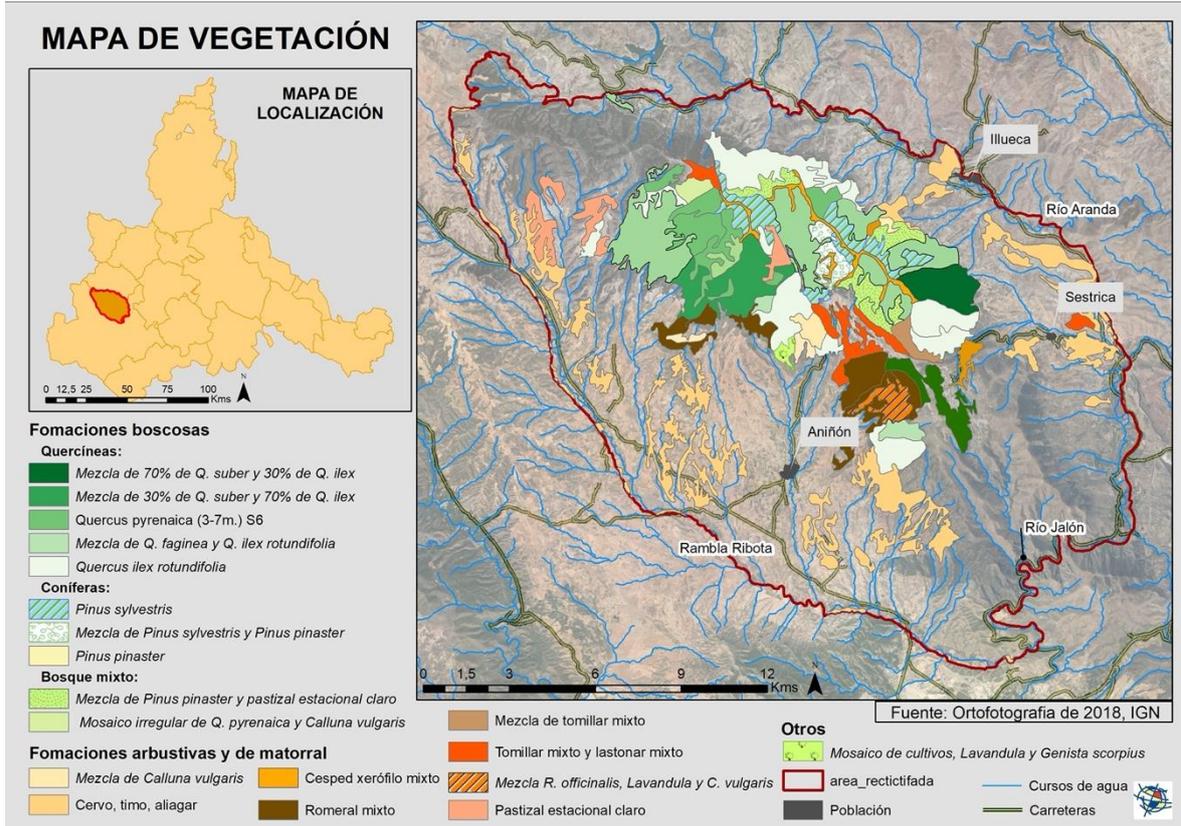
Una primera aproximación al área de estudio queda plasmada en el siguiente mapa de localización:



Mapa 1-2. Mapa de localización. Elaboración propia

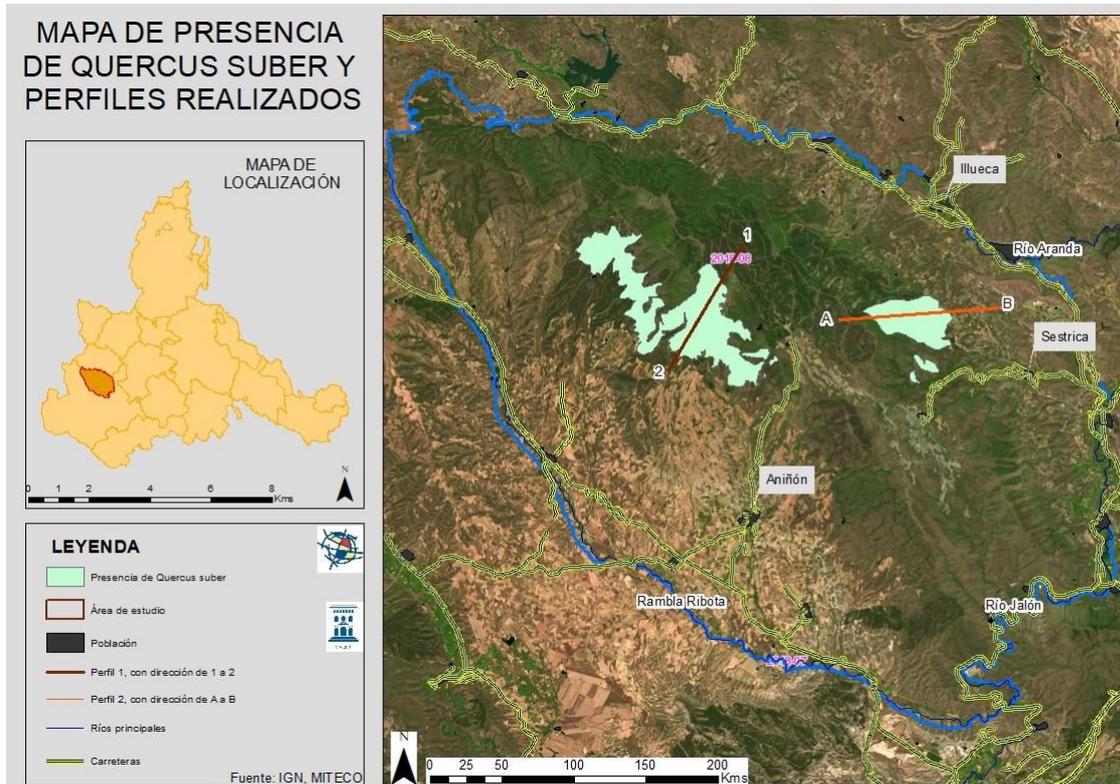
Situada en el Sistema Ibérico, la Comunidad de Calatayud se encuentra localizada al noroeste de la provincia de Zaragoza.

Se lleva a cabo la cartografía de la vegetación de la zona:



Mapa 1-3. Mapa de vegetación. Elaboración propia

Con la idea de conocer mejor las áreas donde se establecen las poblaciones que se pretenden estudiar y aumentar, se llevan a cabo dos perfiles desde Arcgis, trazando dos líneas que crucen las dos manchas, de modo que se pueda observar la altitud de los puntos y relacionarlos con la vegetación y litología de los mismos. Las dos manchas de *Quercus suber*, así como los dos perfiles analizados se representan en el siguiente mapa:



Mapa 1-4. Presencia de *Q. suber* y perfiles realizados. Elaboración propia

De esos dos perfiles realizados se han elaborado unos gráficos a partir de los mismos con información litológica y de vegetación, así como sus altitudes.

**Leyenda:**

**Vegetación**

-  *Quercus ilex*
-  Mezcla de 70% *Quercus ilex* y 30% *Quercus suber*
-  Mezcla de 70% *Q. suber* y 30% *Q. ilex*.
-  *Pinus pinaster*
-  *Pinus sylvestris*
-  Pastizal estacional
-  Césped xerófilo mixto

**Morfología**

-  Depresiones con limolitas, areniscas y dolomias
-  Laderas bajas sobre cuarcitas, areniscas y microconglomerados
-  Laderas medias y altas sobre areniscas y limolitas
-  Glacis
-  Laderas bajas y fondos de valle sobre pizarras
-  Depósitos aluviales

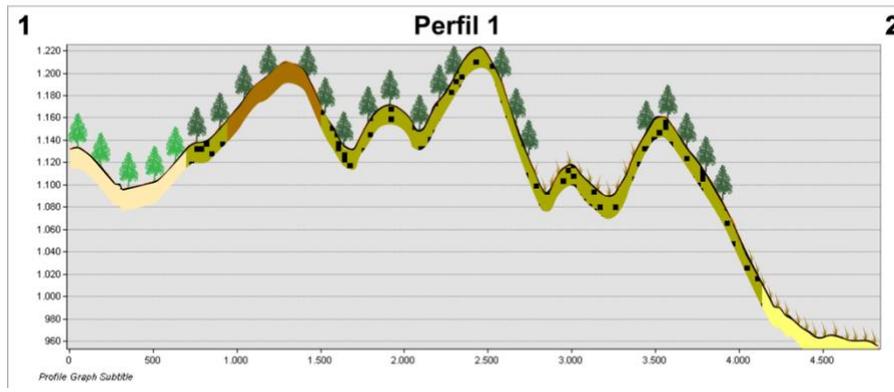


Ilustración 2. Perfil 1. Elaboración propia

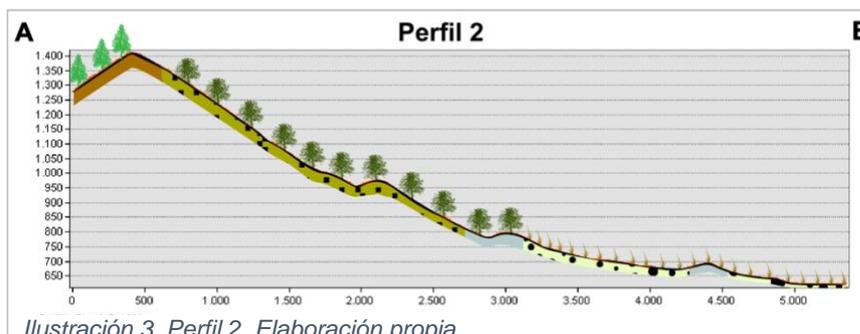


Ilustración 3. Perfil 2. Elaboración propia

En el perfil 1, situado de noreste a suroeste, se puede observar la presencia de *Quercus suber* a una altitud de 1000 m sobre el nivel del mar, en laderas bajas sobre cuarcitas, areniscas y microconglomerados.

En el perfil dos, desarrollado de oeste a este, la mezcla de *Quercus suber* se encuentra entre los 1300 y los 800 m de altitud en suelos de nuevo de laderas bajas tanto de cuarcitas, areniscas y micro conglomerados como fondos de valle sobre pizarras.

## 1.5 Caracterización ecológica

La caracterización ecológica de la zona de estudio se lleva a cabo con la elección de parámetros fisiográficos, climáticos y edáficos.

### 1.5.1 Parámetros fisiográficos

#### 1.5.1.1 Altitud media

Sin duda la altitud del área de estudio es peculiar con respecto a las estaciones analizadas en la monografía del INIA, con un coeficiente de variación elevado (Sánchez Palomares et al., 2007) situándose por encima de la media, tal como se puede ver en el siguiente gráfico analizado en (Aladrén et al., n.d., p. 5).

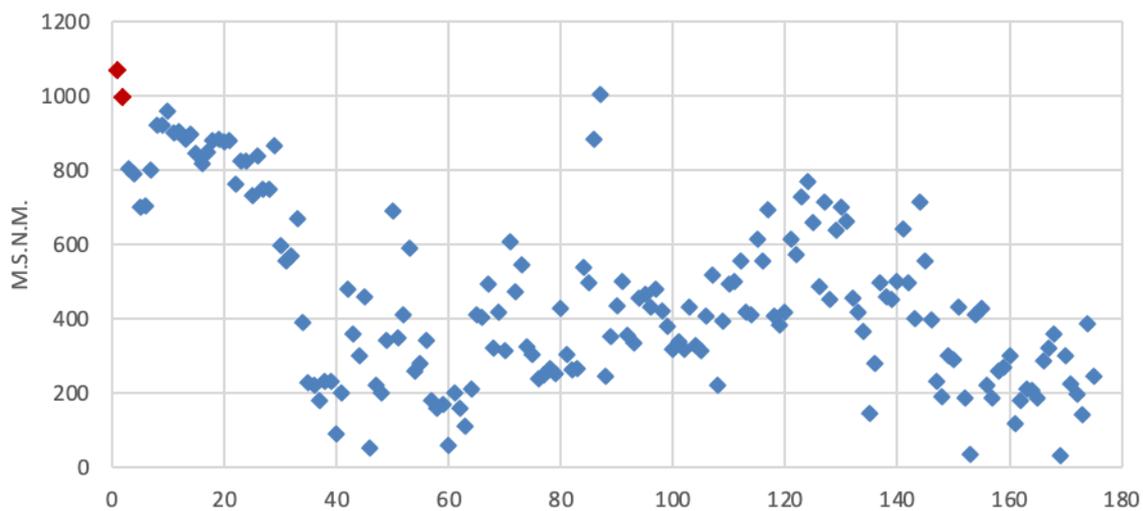
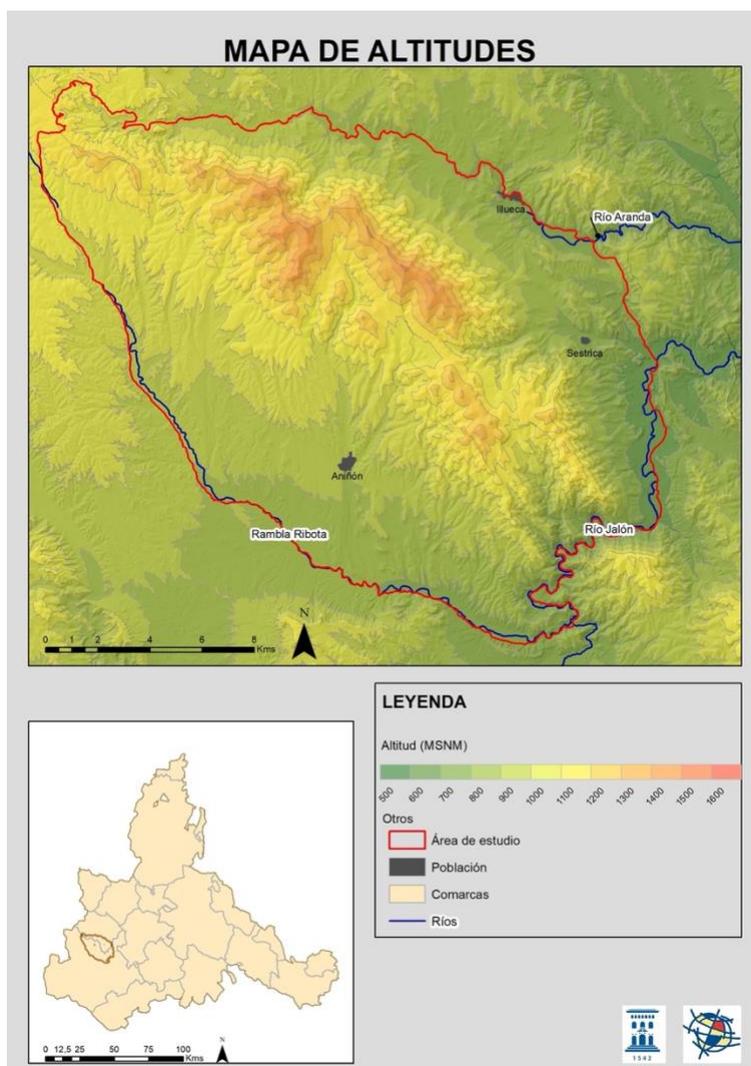


Ilustración 4. Altitud media. Aladren et al. nd p 5. En color rojo aparecen las parcelas de Sestrica y Aniñón

Una vez establecida la peculiaridad de la zona de estudio, el análisis de los diferentes puntos se puede ver en el siguiente mapa:



Mapa 1-5. Mapa de Altitudes. Elaboración propia

Efectivamente en el mapa se puede comprobar la situación de las manchas de Quercus sùber a latitudes mayores superiores a 1000 m sobre el nivel del mar

### 1.5.1.2 Pendiente

Así mismo este parámetro analizado, muestra valores en porcentajes de pendiente bajos, claramente inferiores a las estaciones estudiadas en la monografía del INIA (Aladrén et al., n.d., p. 7).

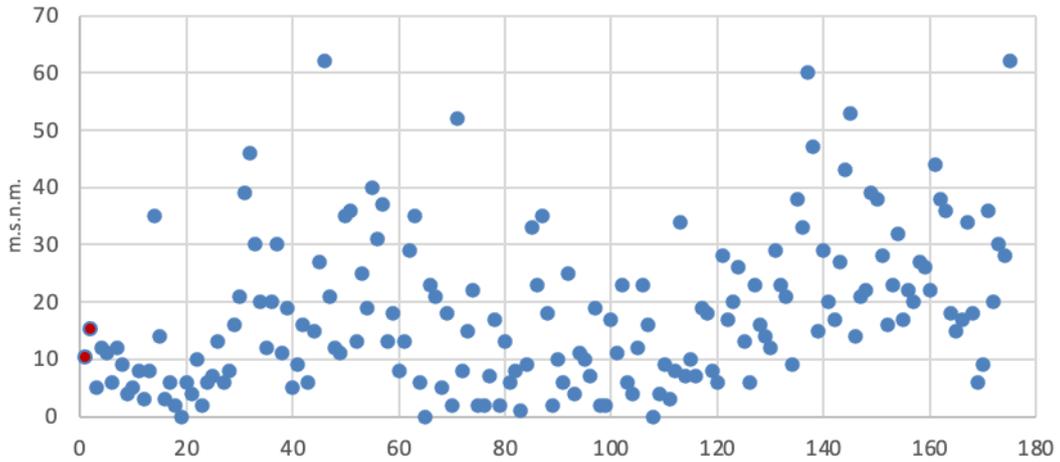
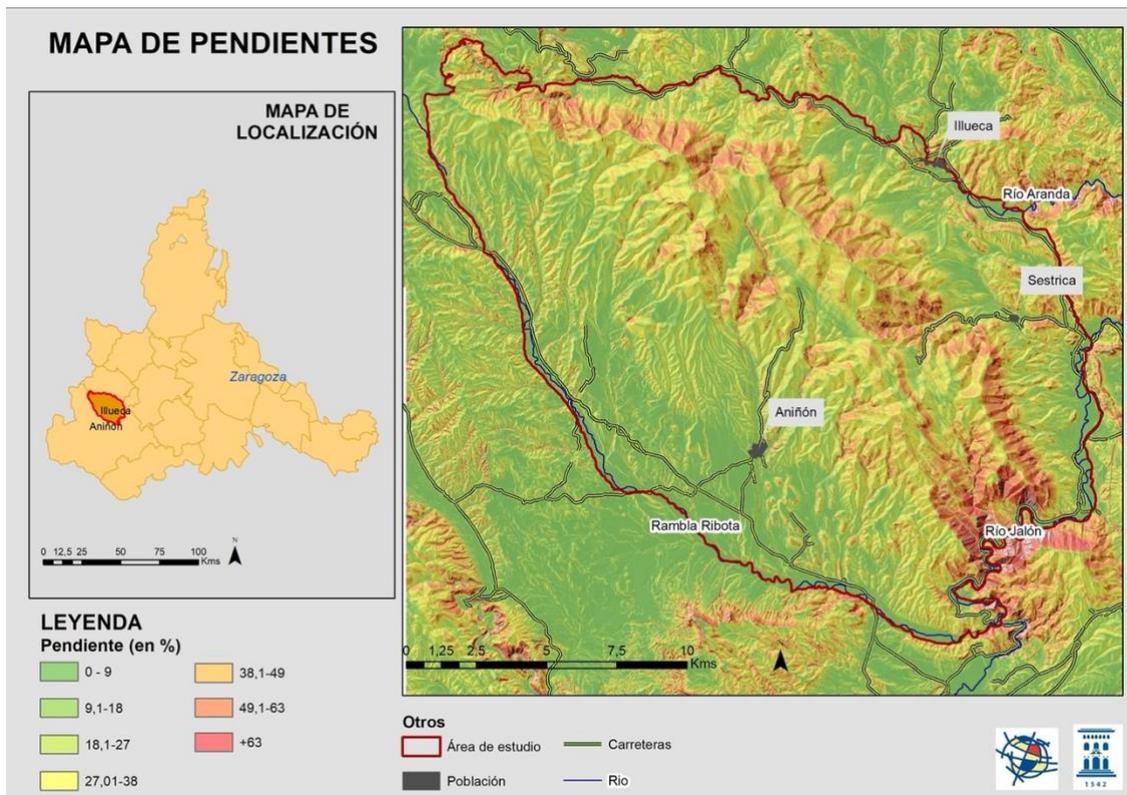


Ilustración 5 Pendiente en %. Aladren et al. nd p 8. En color rojo aparecen las parcelas de Sestrica y Aniñón

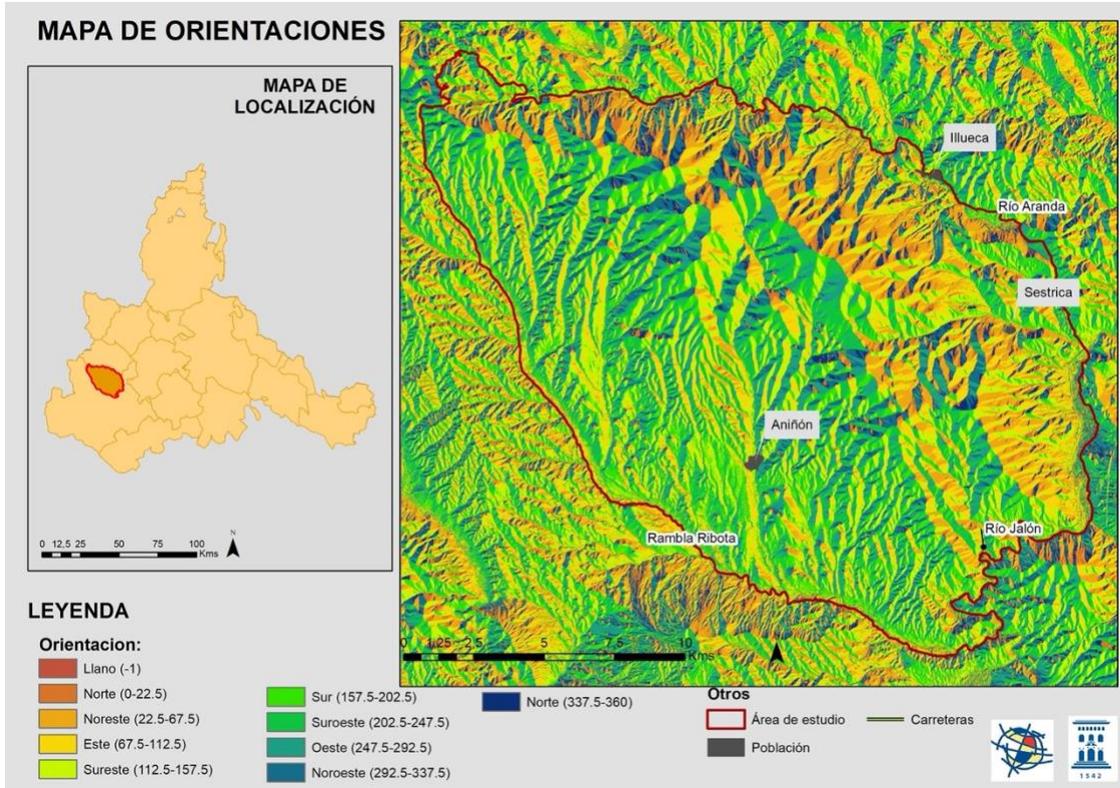


Mapa 1-6. Mapa de Pendientes. Elaboración propia

La zona de estudio no desarrolla pendientes con un porcentaje elevado, excepto en la zona central cruzando de noroeste a sureste donde sí se pueden observar pendientes superiores al 30%.

### 1.5.1.3 Orientación

La orientación del área no ha sido comparada con la monografía del INIA, donde sí analiza la insolación. Pero en este estudio se ha relacionado directamente con la orientación, con la idea de aumentar el estudio de este parámetro en trabajos posteriores.

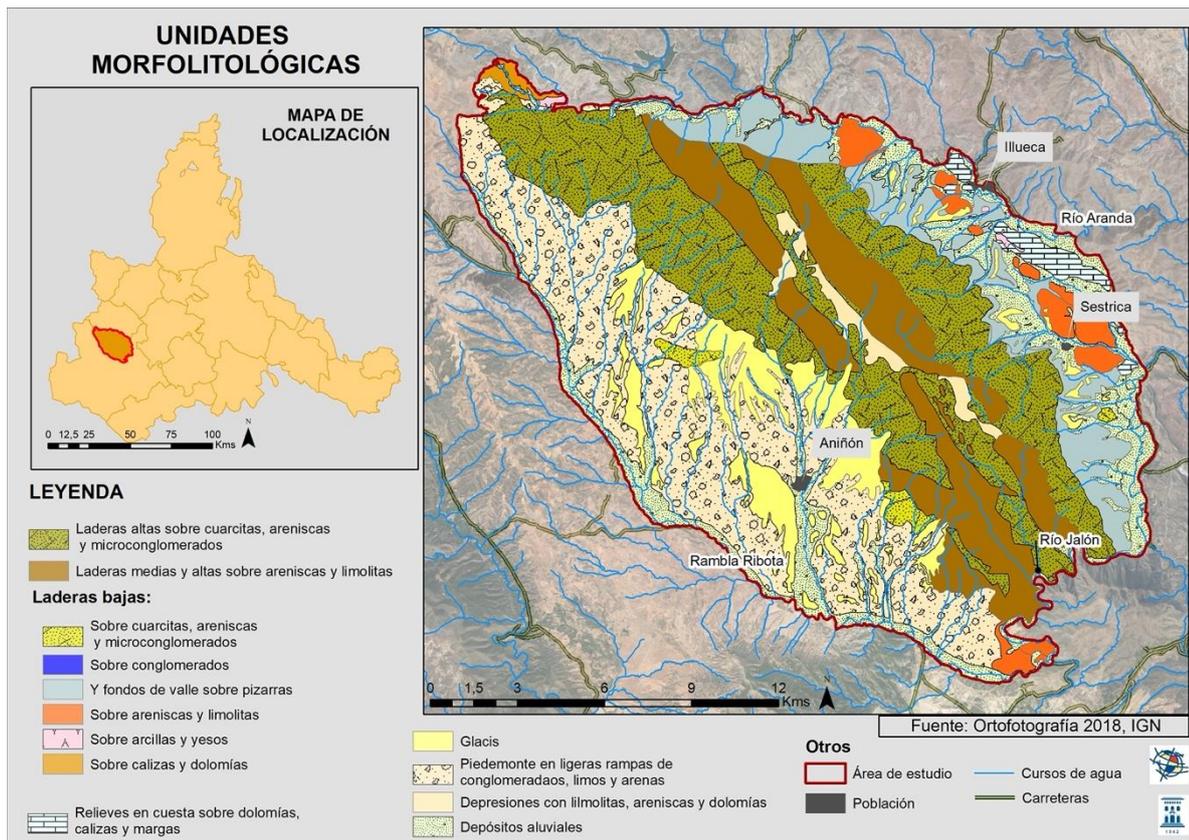


Mapa 1-7. Mapa de Orientaciones. Elaboración propia

### 1.5.1.4 Litología

En cuanto a las Unidades Morfolitológicas, la zona de estudio puede dividirse, en un primer vistazo en tres grandes sectores. Un primer sector situado al noreste de la zona, un segundo en el lado sur-noroeste y un último sector, dividiendo estos dos, situado de forma armoricana.

El sector noreste, donde discurre el río Aranda, es donde se observan la mayoría de las laderas bajas desarrolladas sobre suelos de areniscas y limolitas. Estas laderas se encuentran localizadas al noreste de la población de Sestrica, donde se establece la unidad más grande de todas las identificadas, que se une a otra unidad de tamaño también importante, a través de otra serie de laderas bajas de proporciones más humildes. Entre esta "cadena" de laderas bajas se entrelazan depósitos aluviales, así como depresiones que descansan en limolitas, areniscas y dolomías.



Mapa 1-8. Mapa de Unidades Morfolíticas. Elaboración propia

### 1.5.1.5 Edafología

Según el mapa de suelos de la provincia de Zaragoza (Guerra & Monturiol, 1970), los suelos del entorno de Sestrica están dominados en las vertientes N/W por tierras pardas forestales [Dystric Cambisols en nomenclatura FAO actualizada (IUSS Working Group WRB, 2015)] desarrolladas sobre pizarras y cuarcitas, y en las vertientes S/E por tierras pardas meridionales (Eutric Cambisols) también sobre pizarras principalmente y cuarcitas con áreas de xeroránkeros y litosuelos (Leptosols). Según la misma fuente, se encuentran también suelos pardos calizos (Calcisols) sobre terrazas erosionadas y sobre calizas, margas yesíferas y yesos.

Badía et al. (2018) analizaron los suelos de los alcornoques de Sestrica, los cuales identificaron como suelos con cierto grado de evolución formados a partir de areniscas y cuarcitas y cuyo tipo modal es probablemente de Dystric Cambisols. La característica más destacada de los suelos analizados en ese estudio fue una textura gruesa que, unida a una moderada pluviosidad, produce en los suelos un fuerte lavado, conducente a una translocación de las arcillas y a su acumulación en horizontes árgicos profundos en los suelos más gruesos (Fibla, 2019), que se clasifican en tal caso como Abruptic Luvisols. En las zonas más pendientes o erosionadas, por su parte, los suelos se adelgazan por debajo de un metro de espesor y pasan a clasificarse como Leptic Cambisols (Badía et al., 2018).

En todos estos suelos, la textura gruesa genera una relativa sequía edáfica en los horizontes menos profundos que, según Badía et al. (2018), tal vez puede beneficiar al alcornoque al reducir la competencia por parte de otras plantas no dotadas de raíces profundas. En base a esta conjetura, es posible hipotetizar que los suelos tendentes a una textura gruesa debido al carácter arenoso-gravoso y poco alterable de los materiales de origen pueden tener igualmente una buena disposición a albergar alcornoques. Tales suelos incluirían los desarrollados a partir de materiales con areniscas y cuarcitas ubicados en áreas más elevadas de la Sierra de la Virgen, y donde podemos esperar encontrar las tres tipologías de suelo identificadas en los alcornoques de Sestrica (Dystric Cambisols, Leptic Cambisols y Abruptic Luvisols), aunque seguramente deban excluirse aquellas zonas más erosionadas donde los Leptic Cambisols coexisten con suelos de espesor mínimo (Leptosols) y afloramientos rocosos. Otras tipologías edáficas presentes en la zona de Sestrica: los Calcisols, Regosols, Fluvisols y, probablemente también los Eutric Cambisols desarrollados sobre pizarras, son suelos poco lavados manifiestamente diferentes a aquellos donde se localizan los alcornoques, lo que unido a otras disparidades de tipo climático llevaría a no considerarlos de una alta potencialidad para el establecimiento del alcornoque.

Todo esto hace que la edafología, en un área tan reducida, no parezca que vaya a ser un parámetro que cambie en exceso.

### 1.5.2 Parámetros climáticos

Los parámetros climáticos analizados en este estudio será la media obtenida de: la temperatura máxima de cada uno de los meses, la temperatura mínima de cada uno de los meses y la precipitación media de los 12 meses en el rango de años que va desde 1960 hasta 2018.

## 2. OBJETIVOS

### 2.1 Objetivo general

Llevar a cabo un análisis de las poblaciones de *Quercus suber* ubicadas en el Sierra de la Virgen, analizando algunos de los parámetros principales que componen su ecología.

### 2.2 Objetivos específicos

- Realizar una cartografía predictiva de posibles localizaciones futuras de *Q. suber* en el área de estudio, como apoyo a una posible intervención territorial.
- Aplicar modelos predictivos de machine learning a la ordenación territorial

## 3. METODOLOGÍA

### 3.1 Delimitación área de estudio

La delimitación de la zona de estudio se ha llevado a cabo con el análisis del PNOA 0381 del año 2018 y MDT, descargados del IGN, a partir de Sistemas de Información Geográfica, en este caso Arcgis 10.7.1 bajo licencia de la Universidad, y mediante el análisis visual, siguiendo criterios estructurales, donde el relieve ha ido marcando los límites del mismo. La delimitación este y sur de la zona se da por el río Ribota. La delimitación oeste viene dada por el río Jalón, así como la norte la da el río Aranda. Es en la zona más occidental del área, donde se ha llevado a cabo una delimitación del todo visual, siguiendo criterios de relieve, de manera que no quedase ninguna unidad dividida.

Igualmente, para la realización de la cartografía de las unidades morfolitológicas, se ha realizado a partir de la información tanto del PNOA como el MDT.

La zona posee una geomorfología a simple vista sencilla, pero que se complica extraordinariamente en el análisis visual, por lo que se decide llevar a cabo un análisis morfolitológico, donde se han tenido en cuenta los elementos geomorfológicos en combinación con la litología, dando lugar a las siguientes unidades morfolitológicas:

- Laderas altas sobre cuarcitas, areniscas y microconglomerados
- Laderas medias y altas sobre areniscas y limolitas
- Laderas bajas:
  - Sobre cuarcitas, areniscas y microconglomerados
  - Sobre conglomerados
  - Y fondos de valle sobre pizarras
  - Sobre areniscas y limolitas
  - Sobre arcillas y yesos
  - Sobre calizas y dolomías
- Relieves en cuesta sobre dolomías, calizas y margas
- Glacis
- Piedemonte en ligeras rampas de conglomerados, limos y arenas
- Depresiones con limolitas, areniscas y dolomías

- Depósitos aluviales

Esto es el resultado del estudio tanto del PNOA, como del mapa de suelo (“Mapa de suelos de las provincias de Zaragoza, Huesca y Logroño,” 1970) y el geológico del IGME, Magna50-381.

Una vez delimitada la zona, se realiza una malla de 0,2 m mediante SIG, desde donde se lleva a cabo la extracción de datos de los parámetros analizados.

## 3.2 Modelos

### 3.2.1 Machine Learnig

Puede parecer que Machine Learning y el modelado estadístico son muy diferentes, pero en la última década esas diferencias se han reducido hasta desaparecer. El aprendizaje no requiere suposiciones previas, ni conocer las relaciones subyacentes entre variables, el algoritmo procesará los datos y establecerá los patrones. De forma general, el algoritmo aprenderá de los datos, por lo que cuantos más tenga, más afinado será el patrón de relaciones.

No se hace necesario comprender cómo se recopilan los datos, ni las propiedades estadísticas, ni la distribución poblacional de los mismos.

Por todo esto la elaboración de los modelos predictivos, se decide utilizar machine learning, aprendizaje automático.

Para la realización de los modelos se han utilizado las siguientes librerías:

- <https://scikit-learn.org/stable/index.html>, para el análisis de datos predictivos
- <https://matplotlib.org/>, para la visualización de datos
- <https://numpy.org/>, para el cálculo de datos de computación científica
- <https://pandas.pydata.org/>, para el análisis y manipulación de datos
- <https://seaborn.pydata.org/>, para la visualización de datos estadísticos.

La elección de estas librerías se lleva a cabo porque son las librerías más utilizadas con Python para la predicción de modelos que es lo que se va a realizar en este trabajo. Una vez seleccionadas las librerías y cargadas en el programa Jupyter (<https://jupyter.org/try>) utilizado durante el desarrollo y análisis de los datos se cargan las tablas con los datos que se han obtenido en la malla 0,2 de Arcgis.

Una vez en el entorno virtual hay que crear un notebook.

Los modelos se pueden agrupar de la siguiente manera:

- **Regresión** o de **clasificación**. Para usar modelos de regresión, los valores de las salidas tienen que ser continuos, mientras que, en los modelos de clasificación, los valores de las salidas son categóricos, es decir, los datos de salida se pueden dividir en clases específicas.
- **Aprendizaje reforzado**: Tiene como objetivo desarrollar un sistema que mejore su rendimiento. El sistema aprenderá de su propia experiencia y para ello, necesita iterar con el entorno hasta conseguir el comportamiento deseado. El entorno le proporcionará información al sistema para que reaccione ante él, durante este proceso, el sistema irá obteniendo recompensas positivas o negativas, que harán que refuerce su comportamiento.
- **Aprendizaje no supervisado**: El sistema maneja datos sin etiquetar o desconocidos, pero analizando los datos aportados al modelo, se puede extraer información significativa y así clasificar o predecir las salidas.

En este trabajo se ha optado por el aprendizaje supervisado porque el conjunto de datos tiene salidas/etiquetas conocidas mediante un modelo de clasificación, ya que los datos con los que contamos están etiquetados y las salidas que se quieren predecir se dividen en clases:

- Existencia de alcornoques al 30% de cobertura
- Existencia de alcornoques al 70%
- Sin la existencia de alcornoques.

El conjunto de algoritmos utilizado no podrá predecir salidas distintas a las descritas porque no se ha

entrenado el modelo con dicho objetivo. Para poder clasificar un nuevo conjunto de datos con otras etiquetas a las anteriormente mencionadas, tendríamos que aportar más datos de entrada y volver a entrenar el modelo.

Una vez que ya se tiene claro que tipo de aprendizaje se quiere usar en el modelo, se ha de decidir que algoritmo usar. Los algoritmos más usados para este tipo de modelo son:

- Regresión logística: Es un modelo lineal usado para clasificación binaria, aunque también puede ser usado para clasificación multiclase. Este tipo de modelos usa una función logística que acepta como entrada cualquier número real y devuelve como salida un número real con valor entre 0 y 1.
- K Vecinos más cercanos (kNN): Este modelo busca en las observaciones más cercanas que se están tratando de predecir. Si la mayoría de sus k vecinos pertenecen a una clase, el punto que se está observando se clasificará en dicha clase.
- Máquinas de vectores de soportes (SVM): Este modelo construye en un espacio multidimensional, un hiperplano para la separación de las distintas clases, cuyo objetivo es encontrar el hiperplano marginal máximo que mejor divida los datos en las clases.
- Árboles de decisión
- Bosques aleatorios

De todos los modelos posibles se decide por un modelo de árbol de decisiones por ser estos fáciles de entender, permitir la exploración de datos e identificar importancia de variables a partir de cientos de ellas, así como la limpieza de datos. También se ha tenido en cuenta un modelo de regresión logística, que finalmente se decide no poner en el trabajo.

### 3.2.2. Árbol de decisión

Los árboles de decisiones son representaciones gráficas de posibles soluciones a una decisión basada en ciertas condiciones, es uno de los algoritmos más utilizados en machine learning. Se pueden realizar tareas de clasificación o regresión. La comprensión del funcionamiento es simple a la vez que potente.

Los árboles de decisión tienen un primer nodo que se denomina ROOT (nodo raíz) que se va descomponiendo en el resto de atributos de entrada en dos ramas, planteando una cuestión que puede ser falsa/cierta y vuelve a dividirse hasta llegar a la profundidad máxima posible y dar lugar a las llamadas hojas o nodos finales, que son las respuestas a la solución planteadas en el root.

Un árbol de decisión es una estructura de árbol similar a un diagrama de flujo. Es una visualización que imita fácilmente el pensamiento humano, por lo que se hace fácil de entender e interpretar.

Un ejemplo de esto: se suponen unos atributos como valores de género (masculino/femenino) y edad en rangos para tomar una decisión. Se puede crear un árbol dividido en género y después en edad, o puede ser al revés, primero edad y después género. Será el mismo algoritmo el que devolverá, una vez analizados los datos y salidas, la mejor forma de hacer las divisiones (split) entre los nodos. Analizará todas las opciones y devolverá el árbol óptimo para la toma de decisiones más acertada desde un punto de vista probabilístico.

Para llevar a cabo el análisis de todos los árboles y nodos posibles, el algoritmo debe medir todas las predicciones logradas y valorarlas, para poder compararlas entre sí y obtener la mejor. Para llevar esto a cabo utiliza diferentes funciones, destacando el "índice de Gini" y "Ganancia de información", que utiliza la conocida como "entropía".

El índice de Gini se usa para atributos categóricos. Intenta estimar la información que da cada uno de los atributos. Para medir la aleatoriedad de incertidumbre de un valor aleatoria se define la Entropía. Calculando la entropía de cada uno de los atributos se puede calcular la ganancia de información del árbol, que se debe maximizar.

La entropía mide el grado de incertidumbre de una muestra, es decir, la impureza o aleatoriedad que hay en un conjunto de datos. Una muestra completamente homogénea, en la que todos los datos están clasificados de la misma manera, la incertidumbre es mínima. En este caso entropía o incertidumbre será cero. Sin embargo, una muestra que se ha distribuido de igual manera, es decir, que cada clasificación que hay tiene el mismo número de ejemplos, tendría una incertidumbre máxima. En este caso la entropía o incertidumbre sería 1.

Decir que la entropía es cero, es lo mismo que decir que la impureza es cero. Cuando una muestra contiene el mismo número de ejemplos positivos y negativos la entropía es 1. En el caso de que la muestra contenga un número diferente de ejemplos positivos y negativos la entropía estará entre 0 y el 1. En las muestras cuanto

menor sea la entropía, mejor será nuestro modelo.

La entropía es una medida de información, o más bien falta de ella. Calcula las ganancias de información haciendo una división. Esto mide cómo reduce la incertidumbre cada etiqueta.

Antes de entrenar el modelo se ha de realizar un preprocesamiento de los datos. (igual hay entradas que nos sobran).

El sobreajuste (overfitting) y el subajuste (underfitting) son problemas que se pueden dar en los modelos de machine learning, pero que pueden ser evitados, o al menos reducir.

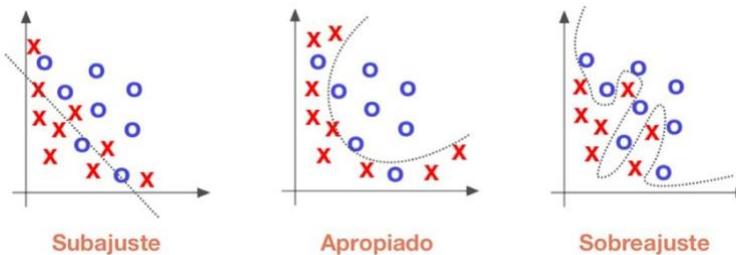


Ilustración 6. Conceptos de subajuste, apropiado y sobreajuste. Fuente: [www.epicalsoft.com](http://www.epicalsoft.com)

El sobreajuste se produce cuando el modelo se adapta bien a los datos de entrenamiento, pero no a los datos de prueba, en este caso el modelo tendrá una varianza alta.

Sin embargo, el subajuste se produce cuando nuestro modelo no tiene suficientes datos de entrenamiento y no es capaz de generalizarlos, en este caso el modelo tendrá un alto sesgo o bias (lo que te separas de la línea)

En muchos algoritmos se puede introducir regularización y modificar sus parámetros (entropía, gini...), utilizar validación cruzada y la detección temprana.

La detección temprana es cuando entrenas repetidamente un modelo se va observando si el rendimiento del modelo mejora o no. El modelo tendrá mejor rendimiento en los datos de entrenamiento conforme van aumentando las iteraciones, pero puede tener un bajo rendimiento en los datos de prueba. Si esto pasa, hay que detener las iteraciones de entrenamiento para que no se produzca un ajuste excesivo.

La *Validación Cruzada* o *k-fold Cross Validation* consiste en tomar los datos originales y crear a partir de ellos dos conjuntos separados: un primer conjunto de entrenamiento (y prueba), y un segundo conjunto de validación.

Luego, el conjunto de entrenamiento se va a dividir en  $k$  subconjuntos y, al momento de realizar el entrenamiento, se va a tomar cada  $k$  subconjunto como conjunto de *prueba* del modelo, mientras que el resto de los datos se tomará como conjunto de *entrenamiento*.

Este proceso se repetirá  $k$  veces, y en cada iteración se seleccionará un conjunto de *prueba* diferente, mientras los datos restantes se emplearán, como se mencionó, como conjunto de *entrenamiento*. Una vez finalizadas las iteraciones, se calcula la precisión y el error para cada uno de los modelos producidos, y para obtener la precisión y el error final se calcula el promedio de los  $k$  modelos entrenados.

Una vez se cuenta con esta precisión promedio para un modelo, se puede repetir entonces el procedimiento del *Cross Validation* para todos los demás modelos de clasificación que se estén evaluando, y se seleccionará al final aquel que produzca el mejor valor de precisión y menor error promedio.

El primer análisis llevado a cabo es un análisis exploratorio de los datos. A continuación, se definen las entradas (features o inputs) y las salidas (targets) siendo estas últimas las coberturas de Quercus tanto al 70% como al 30% como la inexistencia de la especie analizada. Los inputs hacen referencia a todos los parámetros analizados, tanto los climáticos como los físicos, así como los geomorfotológicos, siendo descartados algunos de ellos en el análisis exploratorio de los datos.

## 4. RESULTADOS

En este apartado se expondrán los resultados de los modelos de estadísticos. Los resultados de los parámetros analizados se deciden colocar en la introducción.

Una vez cargadas las librerías necesarias para llevarla a cabo el análisis estadístico se elabora una tabla de presentación de datos donde aparecen reflejados los mismos:

	P. Enero	TM Enero	tm Enero	pendiente	Altitud	orientación	Quercus
Count	339	339	339	339	339	339	339
mean	32,78	7,49	-1,60	33,14	156,07	170,67	1,17
std	3,77	0,55	0,58	12,77	43,63	87,24	0,89
min	25,00	6,10	-2,80	2,86	52,00	1,10	0,00
25%	31,00	7,10	-1,90	24,64	122,00	105,60	0,00
50%	33,00	7,40	-1,70	32,49	152,00	165,85	1,00
75%	35,00	7,80	-1,40	40,17	190,00	236,23	2,00
max	42,00	8,80	0,00	70,24	242,00	359,22	2,00

Tabla 1. Presentación de datos. Tabla completa en el anexo. Elaboración propia

Para los datos numéricos, el índice del resultado incluirá recuento (count), media (mean), desviación típica (std), mínimo (min) y máximo (max), así como percentil inferior, 50 y superior. Por defecto, el percentil inferior es 25 y el percentil superior es 75. El percentil 50 es el mismo que la mediana. Las precipitaciones de cada mes vendrán reflejadas por P más el mes, la temperatura máxima por TM y la mínima por tm.

Una vez llevado a cabo en el análisis exploratorio de los datos, la tabla de input a analizar queda de la siguiente manera:

OBJECTID	precip_01	tmax_01	tmin_01	pendiente	altitud	orientación
1,00	36,00	7,00	-1,90	3,81	176,00	193,43
2,00	35,00	7,10	-1,90	32,19	122,00	108,39
3,00	35,00	7,10	-1,90	25,72	183,00	236,68
4,00	35,00	7,10	-1,90	34,98	113,00	151,77
5,00	36,00	7,00	-1,90	16,90	164,00	198,28
6,00	36,00	7,00	-1,90	22,31	162,00	204,36
7,00	36,00	7,00	-1,90	33,63	121,00	166,02
8,00	37,00	6,80	-2,20	33,14	209,00	267,12
9,00	37,00	6,80	-2,20	36,20	230,00	308,08
10,00	33,00	7,40	-1,60	31,17	186,00	240,11
11,00	35,00	7,10	-1,90	28,68	134,00	173,86
12,00	35,00	7,10	-1,90	33,97	113,00	142,47
13,00	35,00	7,10	-1,90	28,63	143,00	187,87
14,00	35,00	7,10	-1,90	30,11	222,00	330,42
15,00	36,00	7,00	-1,90	2,86	175,00	109,80
16,00	36,00	7,00	-1,90	35,82	110,00	149,85

17,00	36,00	7,00	-1,90	42,09	234,00	326,91
18,00	33,00	7,60	-1,50	34,59	211,00	1,10
19,00	33,00	7,40	-1,60	33,14	205,00	8,01
20,00	33,00	7,40	-1,60	30,65	224,00	324,39
21,00	33,00	7,40	-1,60	20,27	211,00	303,02
22,00	33,00	7,40	-1,60	32,33	122,00	108,15
23,00	34,00	7,30	-1,80	10,82	193,00	271,27
24,00	34,00	7,30	-1,80	22,61	191,00	248,78
25,00	34,00	7,30	-1,80	24,07	143,00	175,71
26,00	32,00	7,80	-1,50	34,32	196,00	250,98
27,00	32,00	7,80	-1,50	41,03	110,00	170,30
28,00	33,00	7,50	-1,70	43,71	227,00	284,11
29,00	33,00	7,50	-1,70	35,96	223,00	285,95
30,00	33,00	7,50	-1,70	32,49	126,00	170,59
31,00	36,00	7,10	-2,10	16,36	152,00	165,46
34,00	36,00	7,00	-2,20	31,49	175,00	229,07
35,00	36,00	7,00	-2,20	63,94	82,00	90,59

Tabla 2. Datos de entrada. Elaboración propia. Tabla completa en el anexo

En la siguiente tabla se puede ver cómo están agrupados los datos según las salidas analizadas:

Cubierta Q. suber al 30%	110 datos
Cubierta Q. suber al 70%	60 datos
Otras cubiertas	169 datos

Estos mismos datos se pueden observar en la siguiente gráfica:

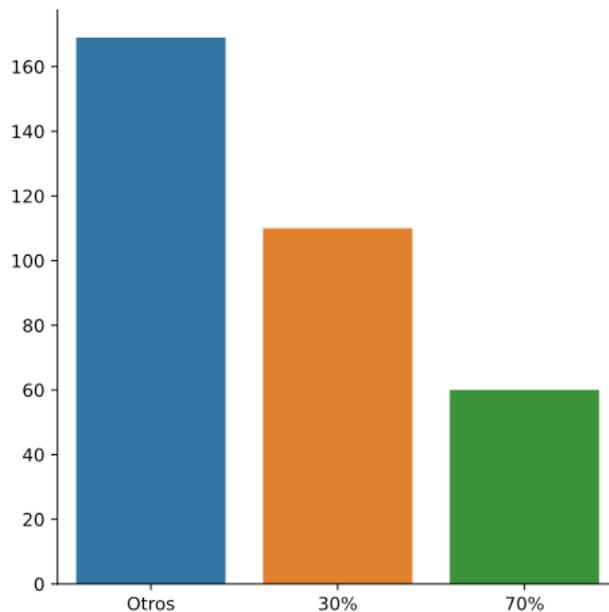


Ilustración 7. Agrupación de datos según salidas. Elaboración propia

Cómo se observa la mitad de los datos pertenecen a datos relacionados con la cubierta vegetal estudiada mientras el resto pertenecen a otras cubiertas tales como *Pinus pinaster*, *Pinus sylvestris*,

tomillares mixtos, romeral mixto, tomillar mixto, pastizal estacional.

Una vez importada la tabla al programa se realiza una limpieza de datos, donde una de las cosas que se busca es si alguna de las variables analizadas o de los parámetros estudiados, tiene una desviación estándar de cero, ya que estaríamos ante una constante por lo que no aportaría nada al modelo. En el área estudiada se observa que ninguno de los parámetros analizados cumple con este requisito por lo que todos los parámetros metidos en la tabla serán analizados.

Se estudia la correlación entre los parámetros de entrada dando como resultado la siguiente matriz de correlación:

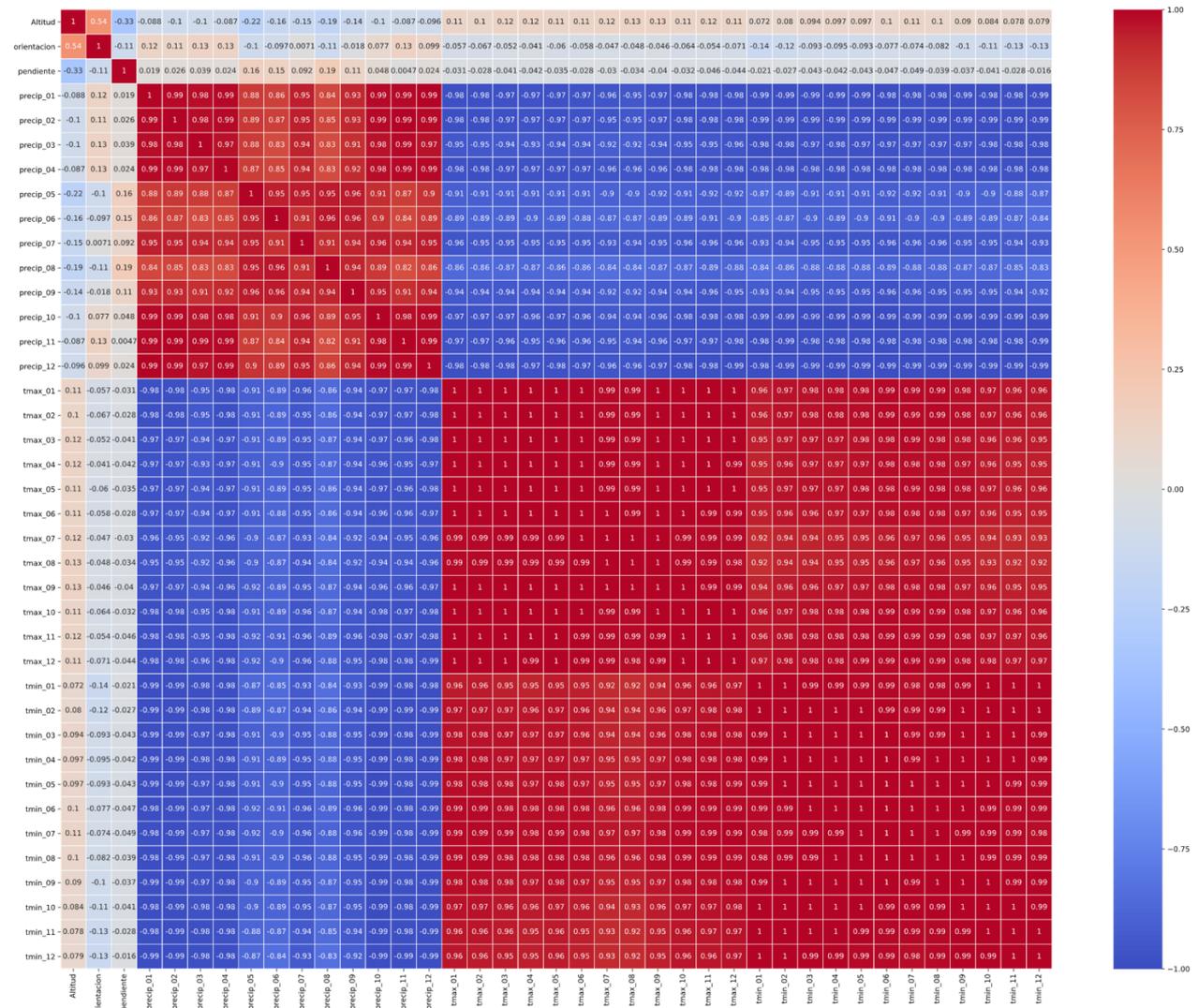


Ilustración 8. Matriz de correlación. Elaboración propia

Gracias a esta matriz se puede observar que los parámetros de altitud, orientación y pendiente están poco correlacionados con el resto de las de los parámetros. Sin embargo, el resto de los parámetros tienen una alta correlación.

#### 4.1 Reducción de dimensionalidad

Al tener parámetros altamente correlacionados se estudia la posibilidad de una nueva filtración finalmente se decide no quitar ninguna entrada debido a que se considera que el coeficiente de correlación debe ser mayor de 0,99 para que la entrada sea borrada se considera este coeficiente de correlación tan elevado ya que durante el entrenamiento del modelo se han probado diferentes valores de coeficiente y la precisión del modelo era más baja que la conseguida con este coeficiente de correlación.

## 4.2 Modelo

Para la realización del modelo se hace necesario pasar las etiquetas de la existencia de cobertura de *Quercus* a representación numérica quedando de la siguiente manera:

- *Quercus* suber en cubiertas de más del 30%: 0
- *Quercus* suber en cubiertas de más del 70%: 1
- Otras cubiertas: 2

Los valores de 0, 1 y 2 son dados de forma automática por el programa.

Una vez aplicada la función de evaluación se ajusta al modelo para que sea lo más preciso posible y se define el árbol de decisiones a través de las librerías sklearn.

Se lleva a cabo el modelo de árbol de regresión con profundidad 5, para cuya visualización se utiliza la función de `export_graphviz` y se convierte de un gráfico a un fichero png, quedando de la siguiente manera:



Durante todo el desarrollo del modelo, se irán probando diferentes niveles de árbol, cuyo desarrollo se puede consultar en el Anexo.

El siguiente paso realizado es el análisis del árbol para ver si es posible su mejora. En la gráfica se puede ver un nodo raíz que hace una primera división por las temperaturas máximas en noviembre, dirigiéndose las salidas a la izquierda por True, las que son menores a 30,5 es decir la clase 70% que son las que más *Quercus* tienen y a la derecha hacia False para las clases del 30% y de otras coberturas. En los diferentes niveles se ven divisiones por diferentes tipos tanto de precipitación, pendiente, temperaturas mínimas etc. Se puede observar que algunas de las hojas finalizan antes de llegar al último nivel. Esto se debe a que alcanzan un nivel de entropía cero o bien a que tienen un número de muestras menor al mínimo permitido para hacer la separación (16). La precisión alcanzada por el árbol da unos valores de 81,06%. Un porcentaje de precisión bastante elevado considerando el número de muestras analizadas.

A lo largo del estudio se han ido cambiando los parámetros de configuración del árbol para conseguir un porcentaje de precisión mayor. Estos parámetros no son los parámetros analizados sino los propios de la función tales como la profundidad (*max\_depth*), el número mínimo de muestras de hojas (*min\_samples\_leaf*) y el número mínimo de muestras por separación (*min\_samples\_split*).

No se consigue mejorar el porcentaje de precisión por lo que se da por bueno el árbol anterior. Para comprobar este modelo de predicción se lleva a cabo el uso de los datos reservados para test. Estos datos son los datos de los parámetros analizados cuyas salidas de presencia de *Quercus* suber se conocen.

El resultado es la siguiente tabla:

OBJECTID	<i>Quercus</i>	Predicción
77	0	0
271	1	1
225	0	0
160	0	0
242	2	0
330	1	1
16	2	2
255	2	2
65	0	0
282	1	1
90	2	0
176	0	0
69	0	0
60	0	2
171	2	0
233	2	2
27	0	0
200	2	2
64	0	0

Tabla 3. Tabla de entrenamiento. Tabla completa en anexo. Elaboración propia

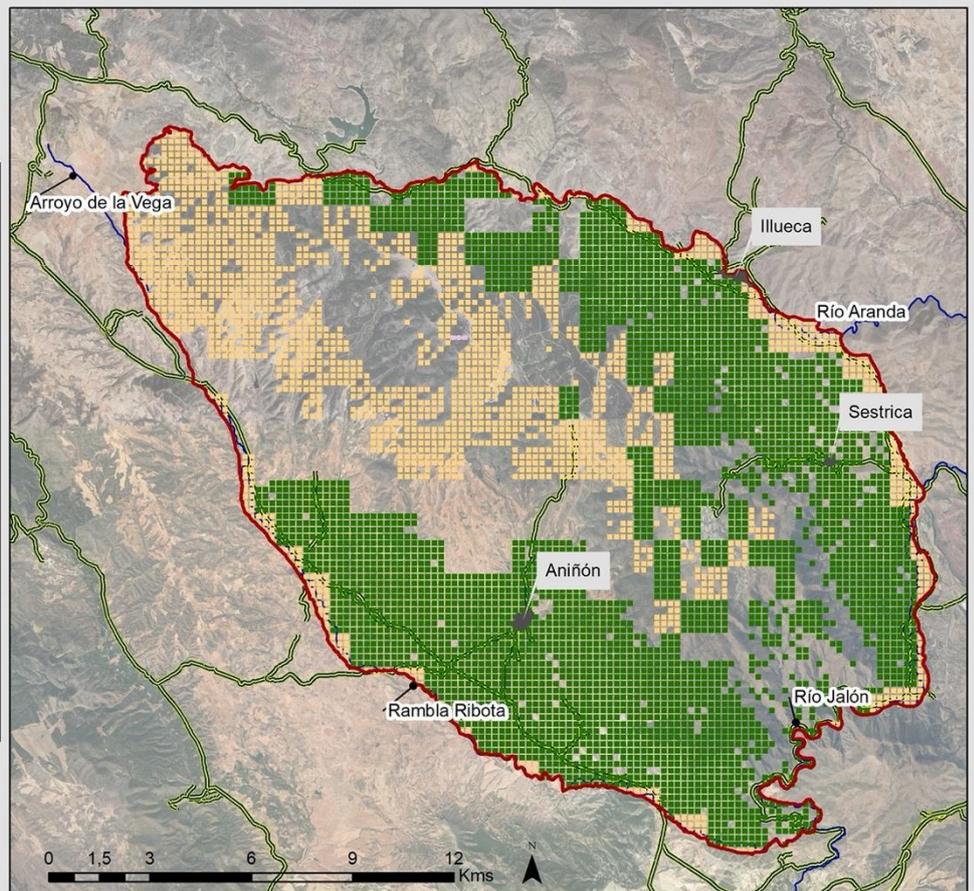
La tabla final de predicción es la siguiente:

OBJECTID	tmin_11	tmax_04	precip_05	tmin_03	Quercus
2	3,10	17,10	53,00	2,40	1
3	3,10	17,10	53,00	2,40	1
4	3,10	17,10	53,00	2,40	1
5	3,10	17,10	53,00	2,40	1
6	3,10	17,10	53,00	2,40	1
7	3,10	17,10	53,00	2,40	1
8	3,10	17,10	53,00	2,40	1
9	3,10	17,10	53,00	2,40	1
10	3,10	17,10	53,00	2,40	1
12	3,10	17,10	53,00	2,40	1
13	3,10	17,10	53,00	2,40	1
14	3,10	17,10	53,00	2,40	1
15	3,10	17,10	53,00	2,40	1
16	3,00	17,00	53,00	2,30	1
17	2,90	17,00	53,00	2,20	1
18	3,10	17,10	53,00	2,40	1
19	3,10	17,10	53,00	2,40	1
20	3,10	17,10	53,00	2,40	1
21	3,00	17,00	53,00	2,30	1
22	3,00	17,00	53,00	2,30	1
23	3,00	17,00	53,00	2,30	1
24	2,90	17,00	53,00	2,20	1
25	2,90	17,00	53,00	2,20	1
26	2,90	17,00	53,00	2,20	1
27	2,90	17,00	53,00	2,20	1
29	3,00	17,00	53,00	2,30	1
30	3,00	17,00	53,00	2,30	1
31	3,00	17,00	53,00	2,30	1
32	2,90	17,00	53,00	2,20	1
33	2,90	17,00	53,00	2,20	1
34	2,90	17,00	53,00	2,20	1
35	2,90	17,00	53,00	2,20	1

Tabla 4. Tabla de predicción. Tabla completa en anexo. Elaboración propia

La tabla obtenida nos da como resultado el siguiente mapa:

# MAPA DE LOCALIZACIÓN FINAL



## LEYENDA

Zonas con mayor probabilidad de albergar la especie:

- Cubierto de *Q. suber* al 30%
- Cubierto de *Q. suber* al 70%

## Otros

- Área de estudio
- Carreteras
- Población
- Río

Fuente: Ortofotografía año 2018, IGN



Mapa 4-1. Mapa de localización final. Elaboración propia

El mapa final da una idea de las posibles localizaciones de una intervención territorial, con la finalidad de aumentar las poblaciones de *Quercus suber*. Como se puede observar, las cubiertas con *Quercus suber* al 70% serían más propicias situarlas en las zonas de menos altitud, con pendientes menos acusadas y unas litologías menos duras, en la zona noreste y la parte más meridional.

En las zonas más elevadas, con unas pendientes más pronunciadas y litologías más duras serán más apropiadas para las coberturas con un 30% de *Quercus suber*.

Este análisis predictivo tan solo da información de la viabilidad de las plantaciones desde un punto de vista climático y físico.

## 5. CONCLUSIÓN

Se han obtenido unos porcentajes de precisión de los datos entrenamiento de un 81,06%. Un porcentaje bastante bueno, pero los porcentajes de precisión que se han obtenido con los datos de prueba no ha sido tan bueno: un 70,54%. Aunque no es un porcentaje malo teniendo en cuenta la cantidad de datos con los que se contaba. Probablemente con un mayor número de datos se podría tener un porcentaje de acierto mucho mayor.

En el manejo de gran cantidad de datos cabe la posibilidad de usarlos de forma reactiva, pero gracias al machine learning, se puede llevar a cabo análisis proactivos, actuando antes de que existe el problema, como ya lo hacen las empresas de telefonía móvil, por ejemplo, donde se ha pasado de analizar los datos para la simple facturación, a encontrar patrones de comportamiento que detectan clientes en riesgo de darse de baja.

Ese manejo de cantidad ingente de datos no es posible de analizar por una persona para sacar conclusiones y menos todavía para hacer predicciones. Sin embargo, los algoritmos sí pueden detectar patrones de comportamiento. En la actualidad no se hace necesario ser un experto en datos para poder aprovechar estas tecnologías. Existen herramientas de uso sencillas y asequibles que ayudan en el análisis de los datos.

El mapa final da un margen muy amplio de actuación sobre el terreno. Se hace necesario afinar el modelo con datos de humedad, pero las probabilidades de aumentar los especímenes de Quercus suber en la Sierra de la Virgen son amplias, tanto en manchas con una cubierta de más del 70%, como en cubiertas del 30%.

Las primeras son propicias en terrenos más llanos y con menos altitud, algo que corresponde con la ecología de la especie. Cubiertas del 30% es más probable que se puedan desarrollar en altitudes más elevadas.

El aumento de estas dehesas puede ser un punto a favor en la lucha contra la despoblación rural, fijando población a partir de la explotación de corcho.

Uno de los objetivos de los ODS es el desarrollo sostenible, y el cambio de la utilización de corchos de plásticos a corchos naturales sin duda contribuirá a ello. Como también a darle una mayor calidad a la D.O. de Cariñena, de piedras el vino, de piedras el corcho.

## 6. BIBLIOGRAFÍA

- Álvarez, E.T., 2003. Experiencias sobre regeneración natural de alcornoque (*Quercus suber* L.). Actas III Reunión Sobre Regen. Nat.-IV Reunión Sobre Ord. Montes 11.
- Bejarano Palma, R., Rubio Recio, J.M., 2018. Biogeografía: una revisión sobre su papel, su trayectoria y sus contenidos. *Estud. Geográficos* 60, 545. <https://doi.org/10.3989/egeogr.1999.i237.577>
- Blanco Castro, E., 2005. Los Bosques ibéricos: una interpretación geobotánica. Planeta, Barcelona.
- Centro de Estudios Bilbilitanos, Institución “Fernando el Católico.” (Eds.), 1997. Cuarto Encuentro de Estudios Bilbilitanos: [Calatayud y comarca]: actas, Publicación núm. 55 del Centro de Estudios Bilbilitanos y núm 1,816 de la Institución “Fernando el Católico.” Presented at the Encuentro de Estudios Bilbilitanos, Centro de Estudios Bilbilitanos ; Institución “Fernando el Católico,” Calatayud (Zaragoza).
- Convention on biological diversity [WWW Document], 1998. URL <https://www.cbd.int/doc/legal/cbd-en.pdf> (accessed 11.9.20).
- Díaz-Esteban, M., 2009. Bases ecológicas preliminares para la conservación de los tipos de hábitat de interés comunitario en España[. Ministerio de Medio Ambiente y Medio Rural y Marino.
- Díaz-Fernández, P.M., Gallardo MUÑOZ, M., 1996. Alcornocales marginales en España. Estado actual y perspectivas de conservación de sus recursos genéticos. *Ecología* 10, 21–47.
- Fernández-Cancio, A., Sánchez-Salguero, R., Gil, P.M., Manrique Menéndez, E., Fernández Fernández, R., Navarro-Cerrillo, R.M., 2012. Efectos del cambio climático sobre la distribución de los alcornocales españoles. Una aproximación fitoclimática para la futura gestión. *Ecosistemas* 21, 50–62. <https://doi.org/10.7818/ECOS.2012.21-3.07>
- Fibla, N.C., 2019. Aplicación de la micromorfología al estudio de suelos forestales del Sistema Ibérico.
- Gento Arranz, I., 2016. Entre naturaleza y cultura: el alcornocal de la dehesa de Foncastín (Trabajo Fin de Grado). Universidad de Valladolid, Valladolid.
- Gutiérrez Hernández, O., Cámara Artigas, R., Senciales González, J.M., García, L.V., 2018a. Modelos predictivos en Biogeografía: aplicación para la modelización de nichos ecológicos en Geografía Física. *Bol. Asoc. Geógrafos Esp.* 88–126. <https://doi.org/10.21138/bage.2395>
- Gutiérrez Hernández, O., Cámara Artigas, R., Senciales González, J.M., García, L.V., 2018b. Modelos predictivos en Biogeografía: aplicación para la modelización de nichos ecológicos en Geografía Física. *Bol. Asoc. Geógrafos Esp.* 88–126. <https://doi.org/10.21138/bage.2395>
- Hernández, J. Millán J. y Serra, A., 2007. Comarca del Aranda. Diputación General de Aragón, Departamento de Presidencia Y Relaciones Institucionales, Zaragoza.
- Longares Aladrén, L.A., Girona García, A., Martí Dalmau, C., BadíaVillas, D., Ortiz Perpiñá, J.O., Mora Hernández, J.L., De Luis Arrillaga, M., Saz Sánchez, M.Á., Zufiaurre Galarza, R., n.d. Caracterización ecológica de los alcornocales de la provincia de Zaragoza.
- Maldonado J., Benito M, Sánchez de Dios R., Sainz H, 2014. Evolución reciente de las áreas de los bosques esclerófilos Ibéricos. Cambios deducidos a partir de la cartografía foresta.
- Manual de gestión del Habitat: Ficha de manejo y conservación, 2011.
- Mapa de suelos de las provincias de Zaragoza, Huesca y Logroño, 1970.
- Moreno, C.E., n.d. Métodos para medir la biodiversidad 86.
- Quezel, P., Tomaselli, R., Morandini, R., Crespo, M., 1982. Bosque y maquia mediterráneos: ecología, conservación y gestión. Serbal, Barcelona.
- Raschka, S., Mirjalili, V., 2019. Python machine learning: aprendizaje automático y aprendizaje profundo con Python, scikit-learn y TensorFlow.
- Rodríguez Urbieto, T.I., 2008. Estructura, dinámica y regeneración de los bosques mixtos de alcornoque (*Quercus suber*) y quejigo moruno (*Quercus canariensis* Willd.) del sur de la Península Ibérica. Universidad de Alcalá de Henares, Madrid.
- Rodríguez-Barbero, C., 2009. Fenología de *Quercus ilex* L. y *Quercus suber* L. en una dehesa del centro peninsular.
- Rubio Recio, J.Manuel, 1989. Biogeografía. Paisajes vegetales y vida animal, Síntesis. ed, Geografía de España. Síntesis, Madrid.
- Sánchez Palomares, O., Jovellar, L., Sarmiento, L., Rubio, A., Gandullo, J., 2007. Las estaciones ecológicas de los alcornocales españoles, Monografías INIA. Serie forestal. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid España.
- Sanmartin, I., 2012. Biogeografía (Capítulo 45). pp. 457–474.
- Tellería, J.L., 2013. Pérdida de biodiversidad. causas y consecuencias de la desaparición de las especies. *Mem. Real Soc. Esp. Hist. Nat.* 10, 13–25.

Datos climáticos:

- CRU-TS 4.03 (Harris et al., 2014) downscaled with WorldClim 2.1 (Fick and Hijmans, 2017).

Recursos web:

- <https://aprendeia.com/sobreajuste-y-subajuste-en-machine-learning/>
- <https://epicalsooft.com>

Librerías para python:

- <https://scikit-learn.org/stable/index.html>, para el análisis de datos predictivos
- <https://matplotlib.org/>, para la visualización de datos
- <https://numpy.org/>, para el cálculo de datos de computación científica
- <https://pandas.pydata.org/>, para el análisis y manipulación de datos
- <https://seaborn.pydata.org/>, para la visualización de datos estadísticos

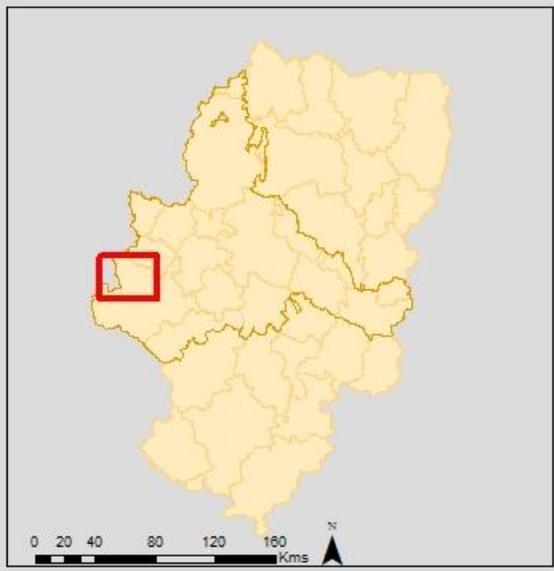
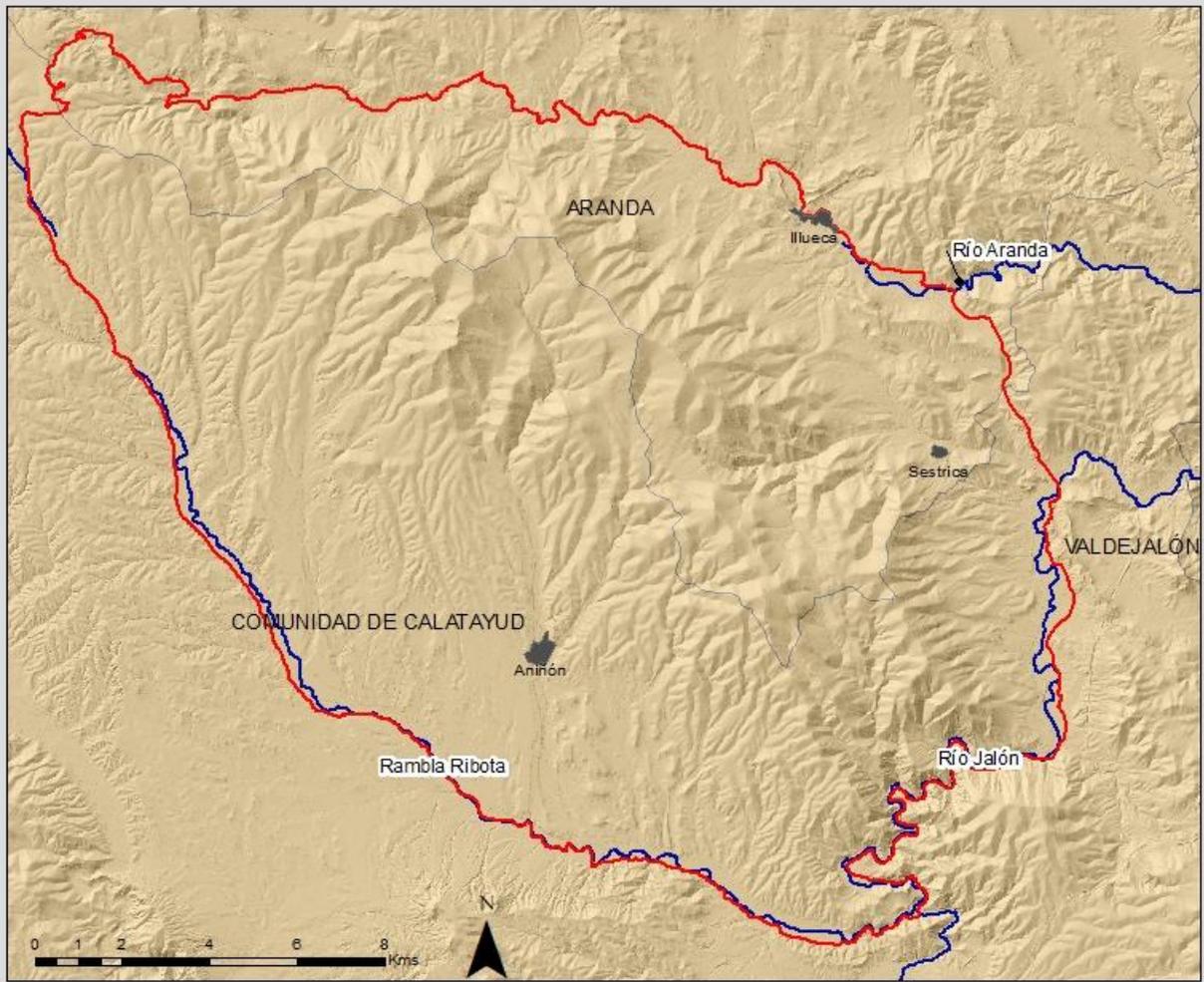
Ilustraciones:

- [www.arbolapp.es](http://www.arbolapp.es)
- [www.freepik.es/home](http://www.freepik.es/home) (iconos encinas/pinos y cesped)

# 7. ANEXO

## 7.1 Mapas

### LOCALIZACIÓN DEL ÁREA DE ESTUDIO



**LEYENDA**

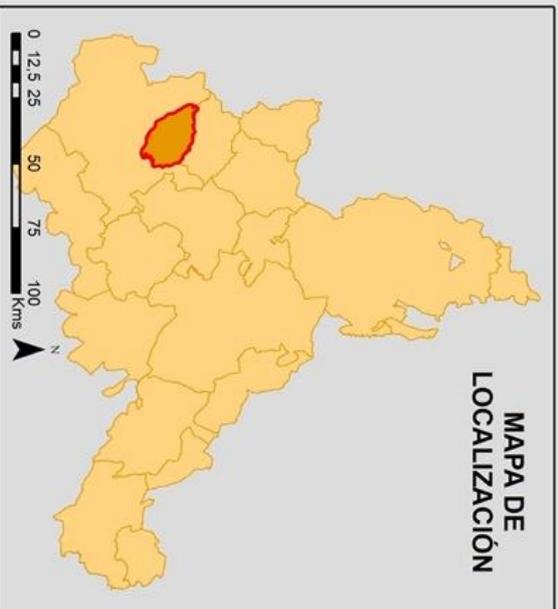
- Área de estudio
- Población
- Comarcas
- Ríos



Elaboración: Verónica García Magalhaes  
Fuente: IGN

# MAPA DE VEGETACIÓN

MAPA DE LOCALIZACIÓN



## Fomaciones boscosas

### Quercíneas:

- Mezcla de 70% de Q. suber y 30% de Q. ilex
- Mezcla de 30% de Q. suber y 70% de Q. ilex
- Quercus pyrenaica (3-7m.) S6
- Mezcla de Q. faginea y Q. ilex rotundifolia
- Quercus ilex rotundifolia

### Coníferas:

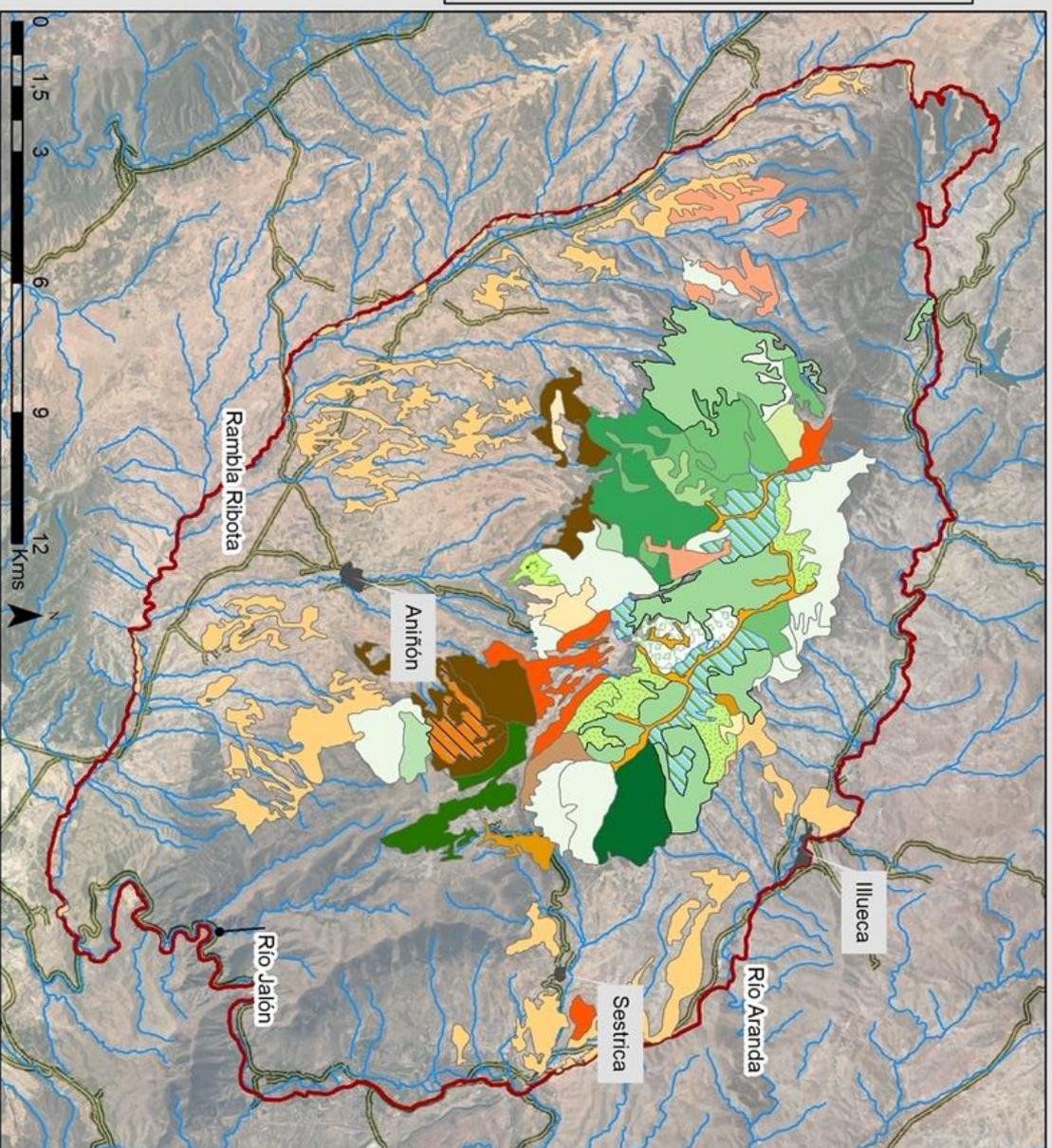
- Pinus sylvestris
- Mezcla de Pinus sylvestris y Pinus pinaster
- Pinus pinaster

### Bosque mixto:

- Mezcla de Pinus pinaster y pastizal estacional claro
- Mosaico irregular de Q. pyrenaica y Calluna vulgaris

## Fomaciones arbustivas y de matorral

- Mezcla de Calluna vulgaris
- Césped xerófilo mixto
- Cervo, timo, allagar
- Romeral mixto



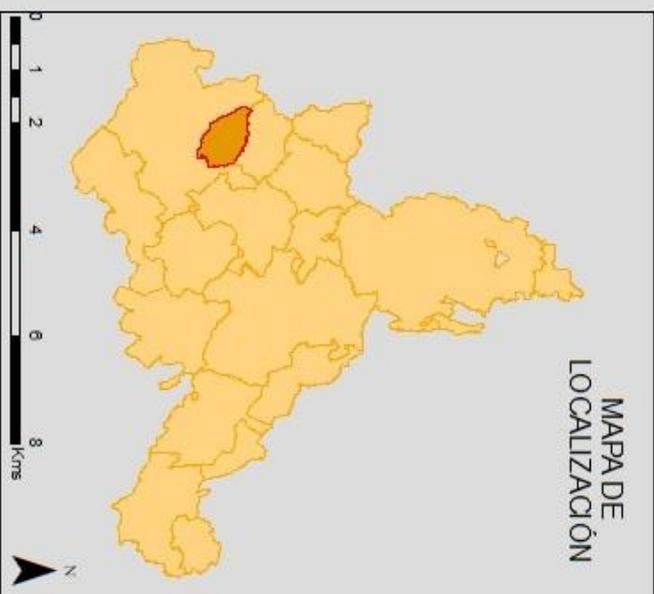
### Otros

- Mezcla de tomillar mixto
- Tomillar mixto y lastonar mixto
- Mezcla R. officinalis, Lavandula y C. vulgaris
- Pastizal estacional claro
- Mosaico de cultivos: Lavandula y Genista scorpius
- area\_rectificada
- Población
- Cursos de agua
- Carreteras

Fuente: Ortofotografía de 2018, IGN



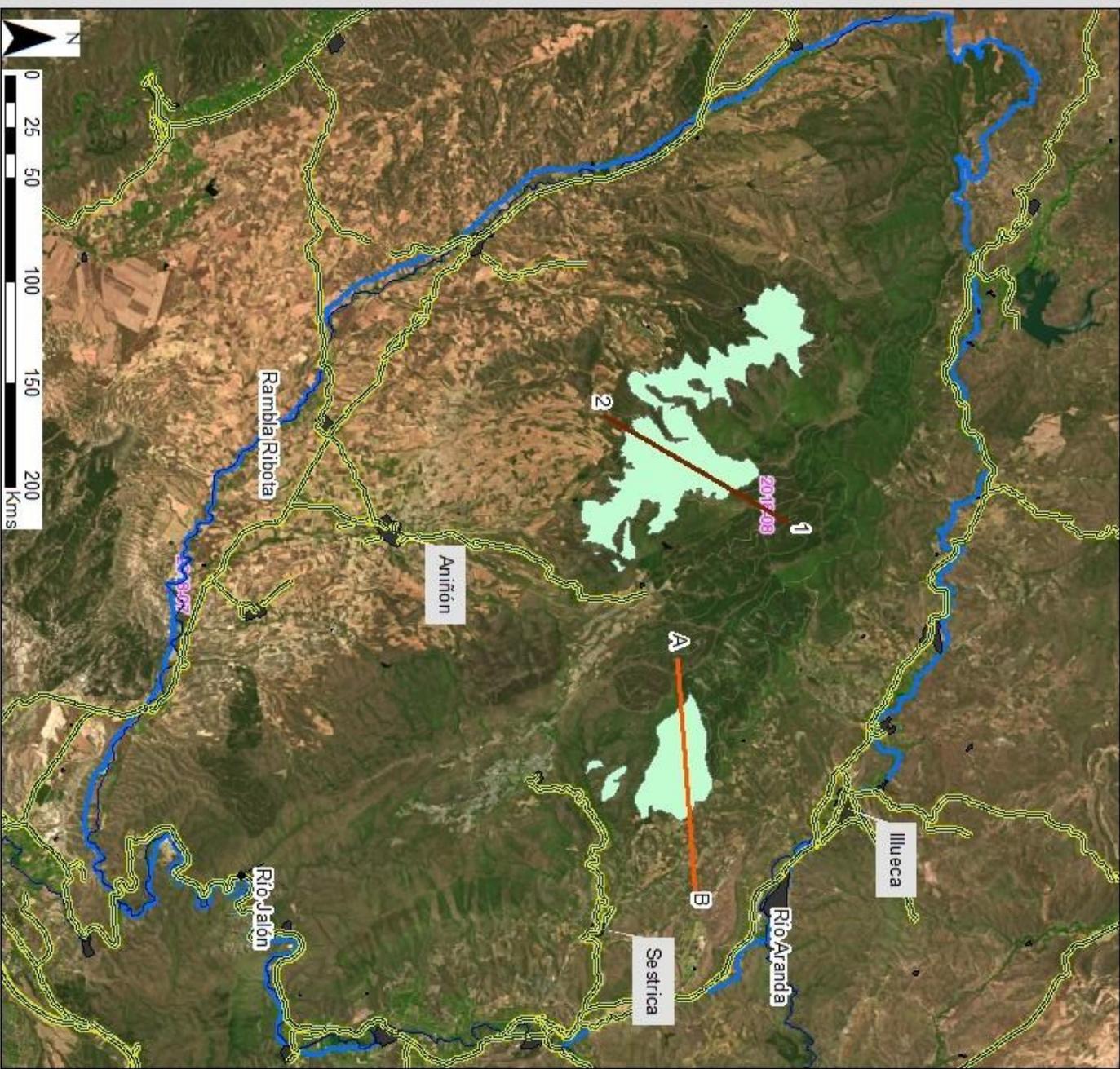
# MAPA DE PRESENCIA DE QUERCUS SUBER Y PERFILES REALIZADOS



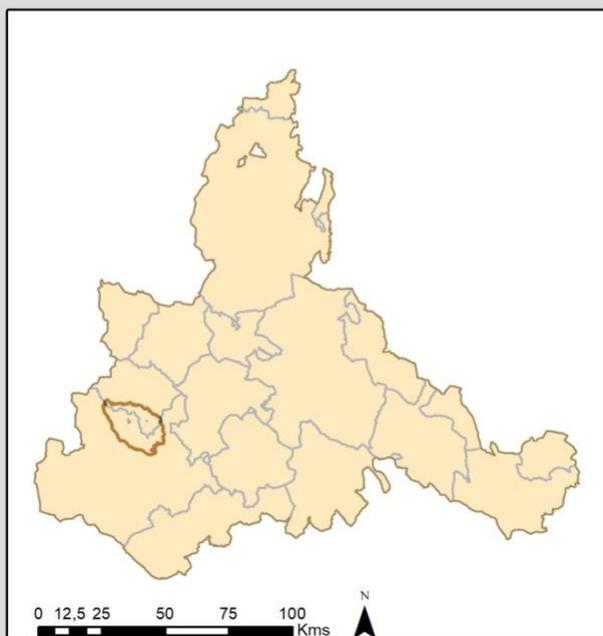
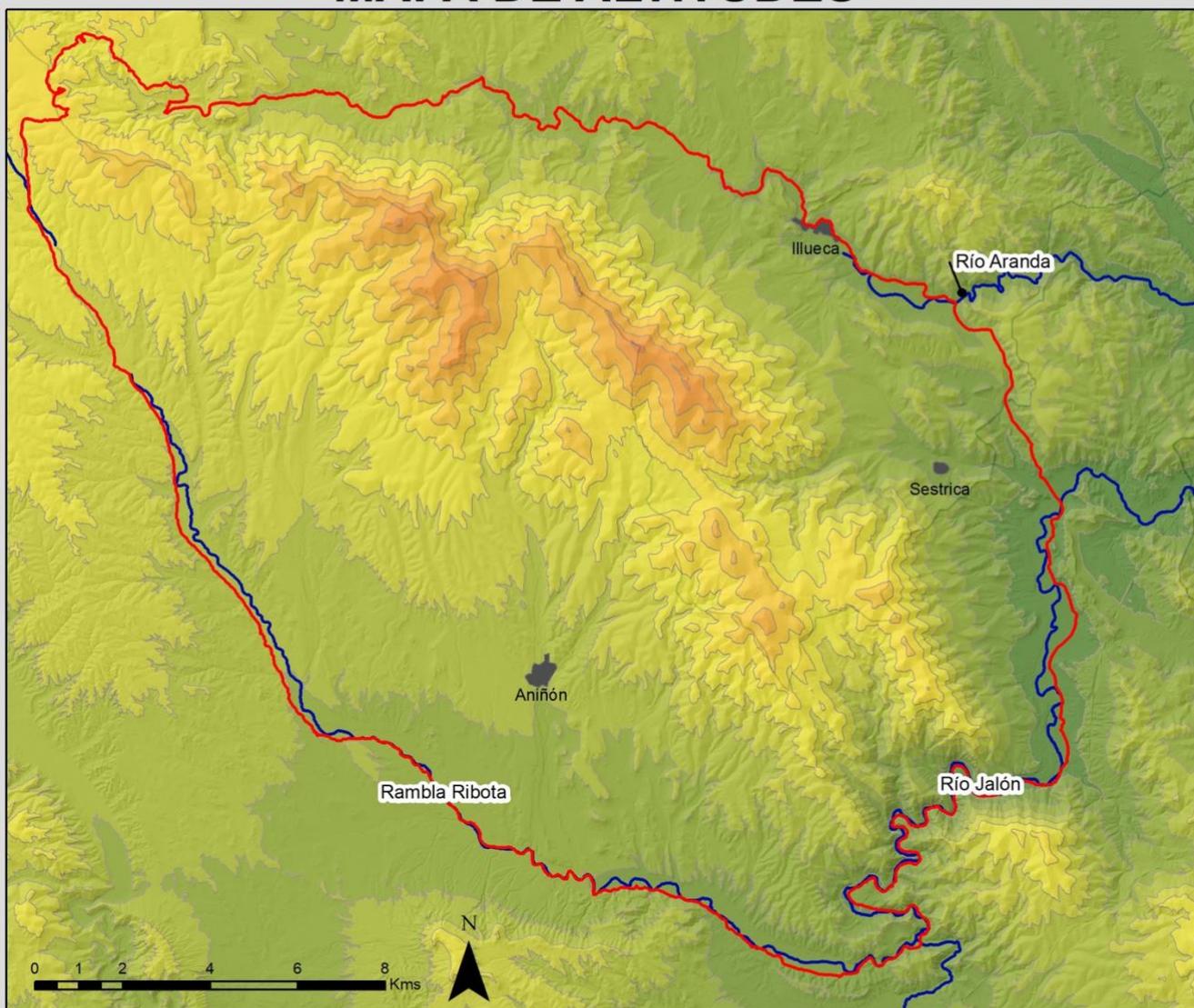
**LEYENDA**

- Presencia de Quercus suber
- Área de estudio
- Población
- Perfil 1, con dirección de 1 a 2
- Perfil 2, con dirección de A a B
- Ríos principales
- Carreteras

Fuente: IGN, MITTECO

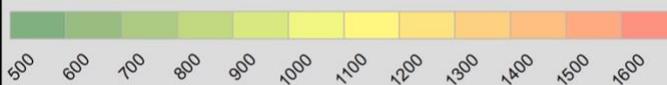


# MAPA DE ALTITUDES



## LEYENDA

Altitud (MSNM)



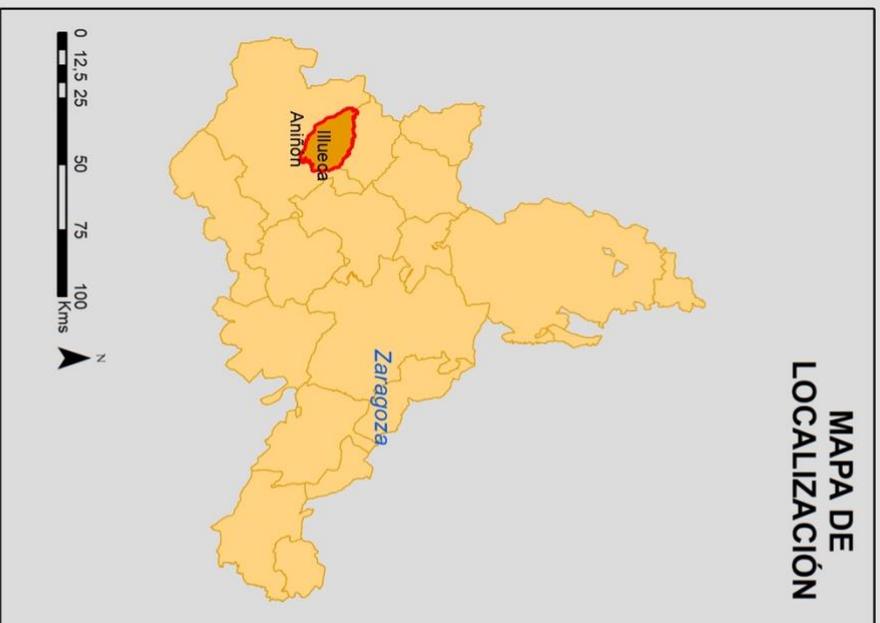
Otros

- Área de estudio
- Población
- Comarcas
- Ríos



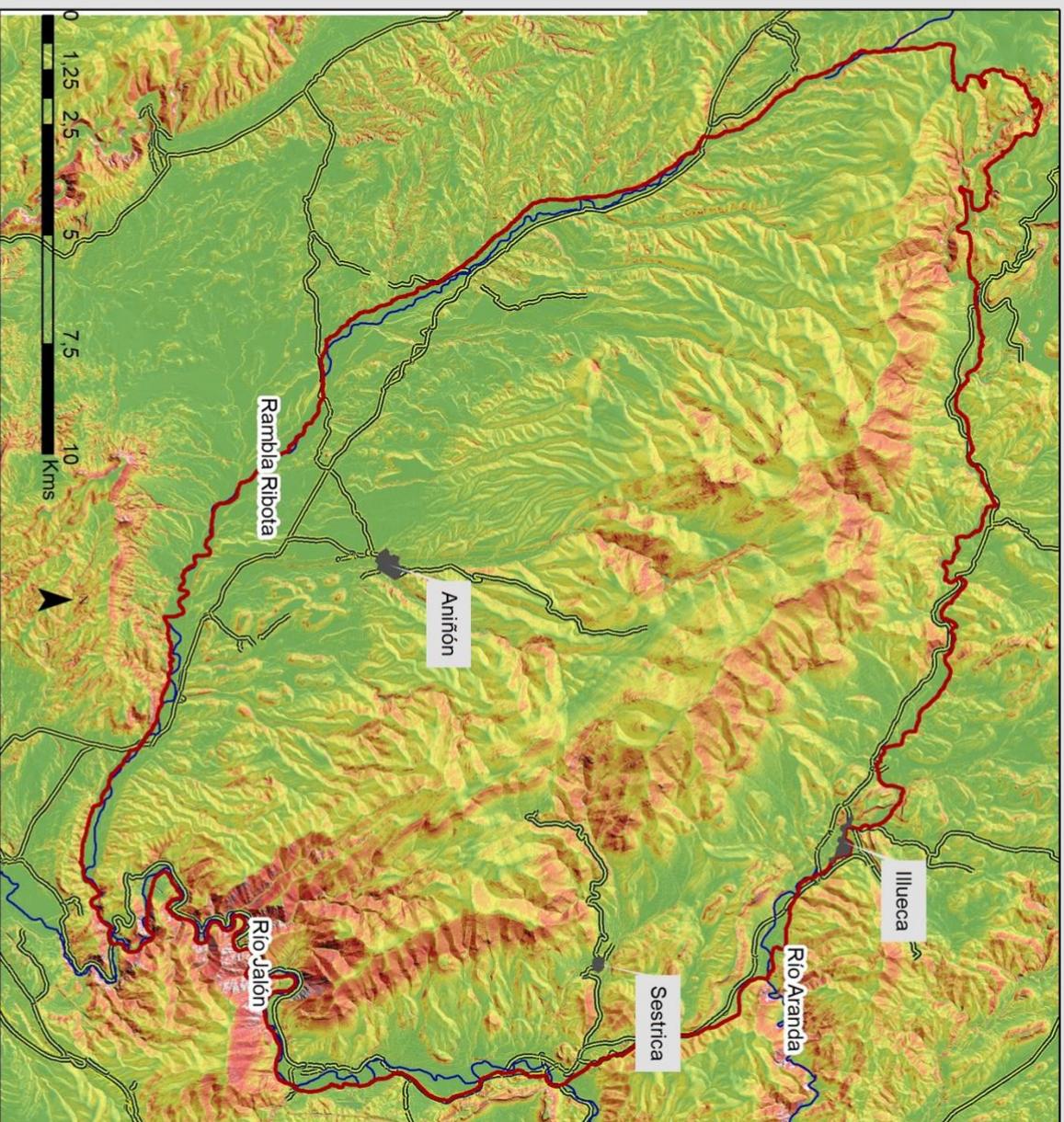
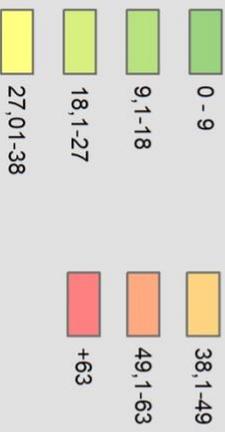
# MAPA DE PENDIENTES

## MAPA DE LOCALIZACIÓN

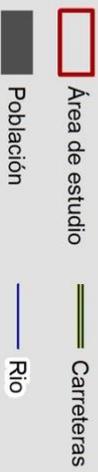


## LEYENDA

### Pendiente (en %)

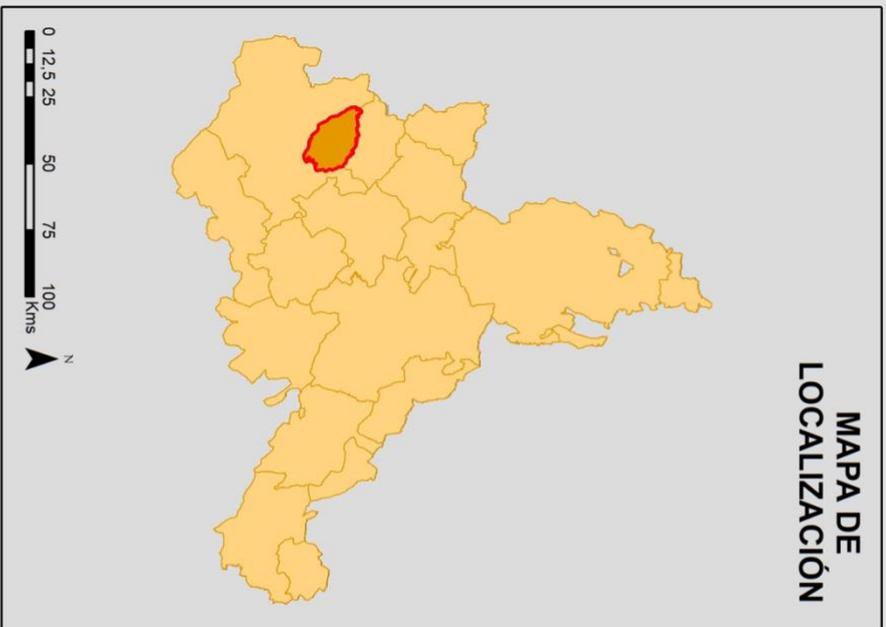


## Otros



# MAPA DE ORIENTACIONES

## MAPA DE LOCALIZACIÓN



## LEYENDA

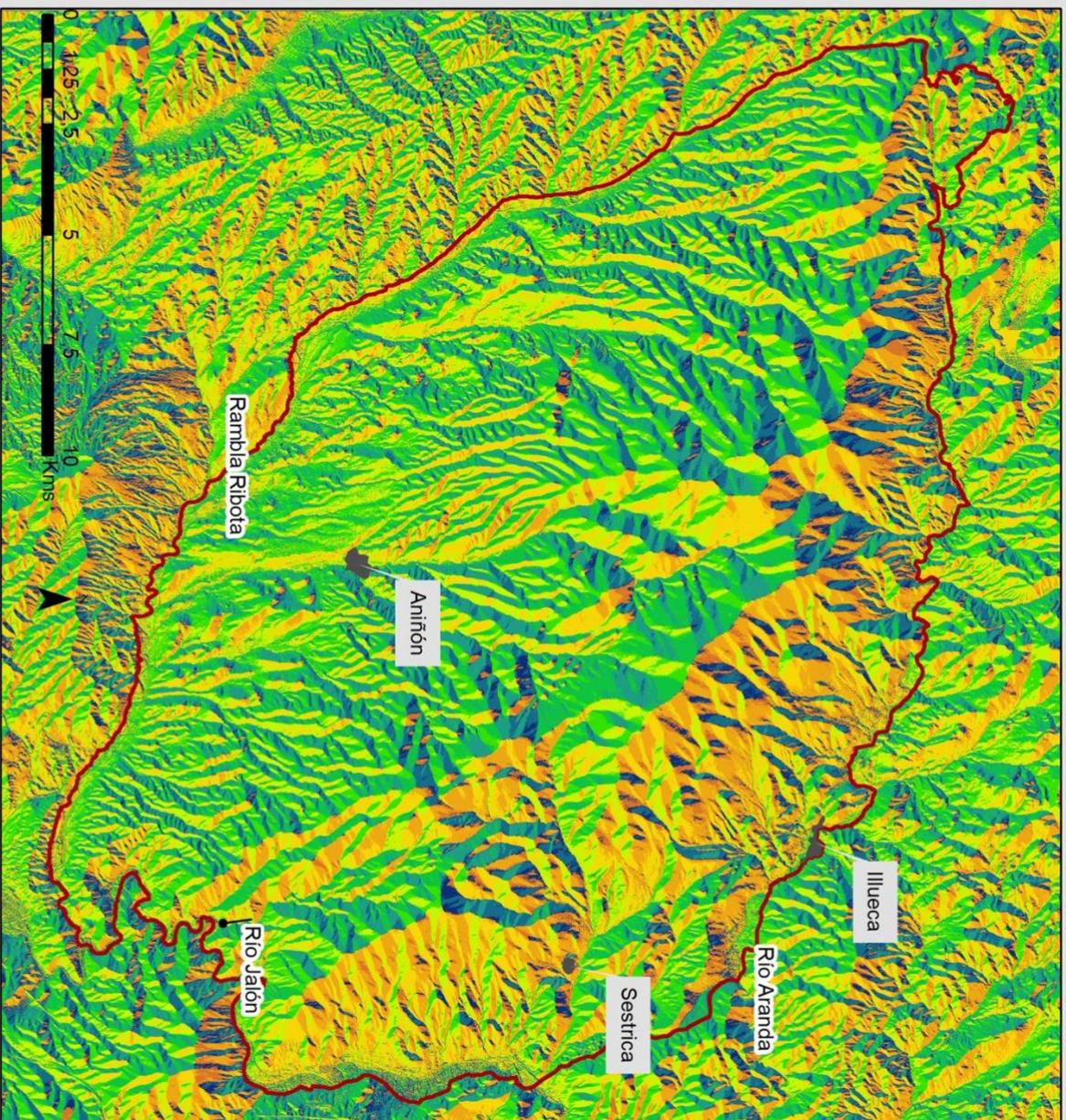
### Orientación:



### Otros

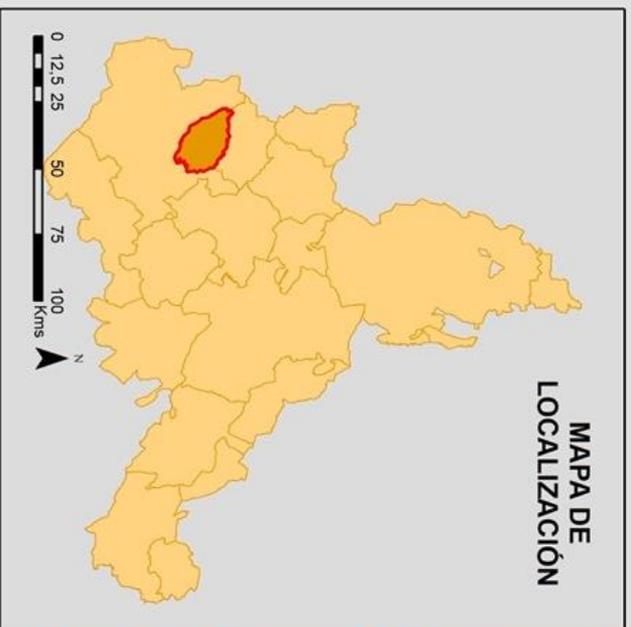


Carreteras



# UNIDADES MORFOLITOLÓGICAS

MAPA DE LOCALIZACIÓN



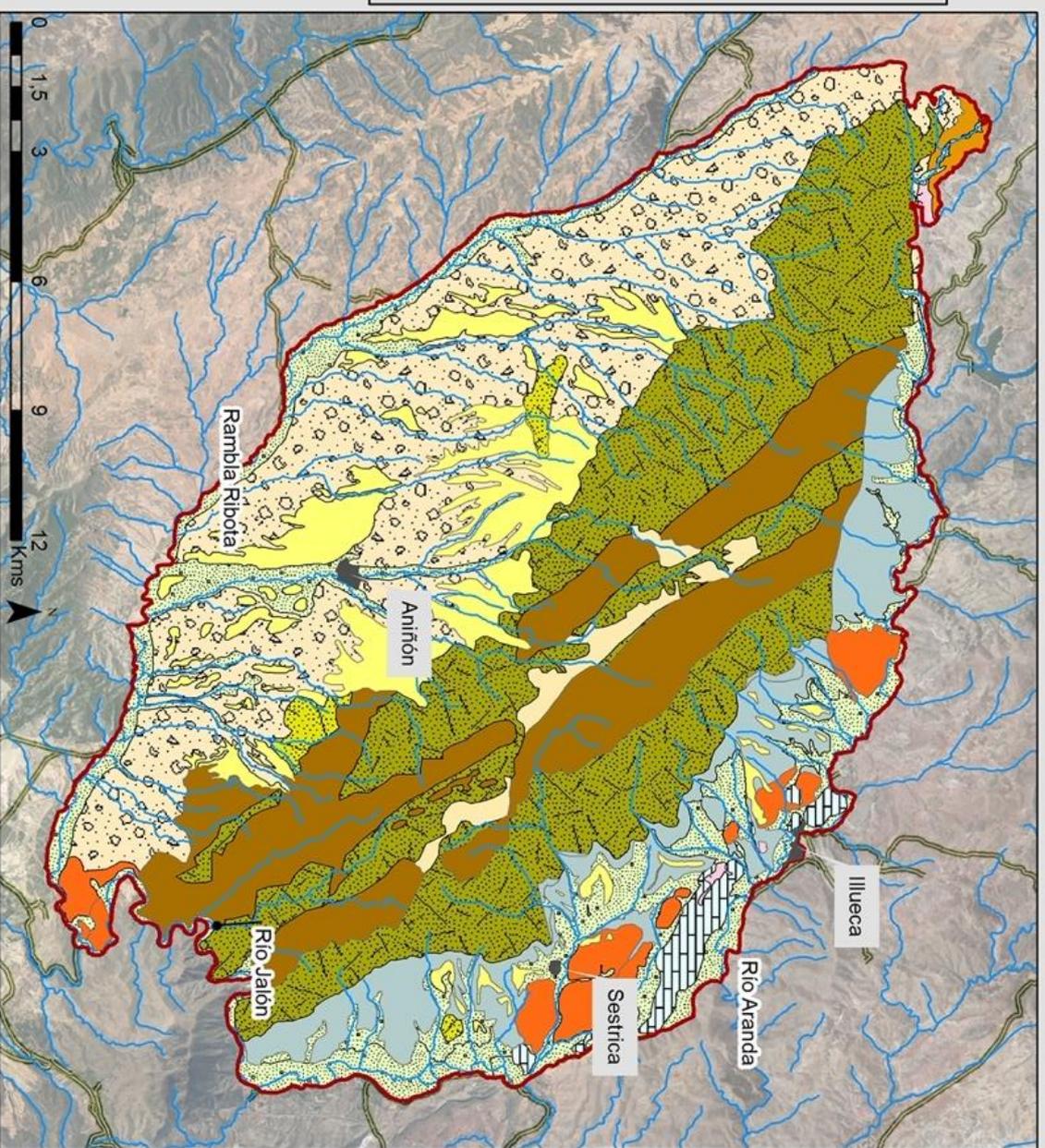
## LEYENDA

- Laderas altas sobre cuarcitas, areniscas y microconglomerados
- Laderas medias y altas sobre areniscas y limolitas

### Laderas bajas:

- Sobre cuarcitas, areniscas y microconglomerados
- Sobre conglomerados
- Y fondos de valle sobre pizarras
- Sobre areniscas y limolitas
- Sobre arcillas y yesos
- Sobre calizas y dolomías

- Relieves en cuesta sobre dolomías, calizas y margas



- Glacis
- Piedemonte en ligeras rampas de conglomerados, limos y arenas
- Depresiones con limolitas, areniscas y dolomías
- Depósitos aluviales

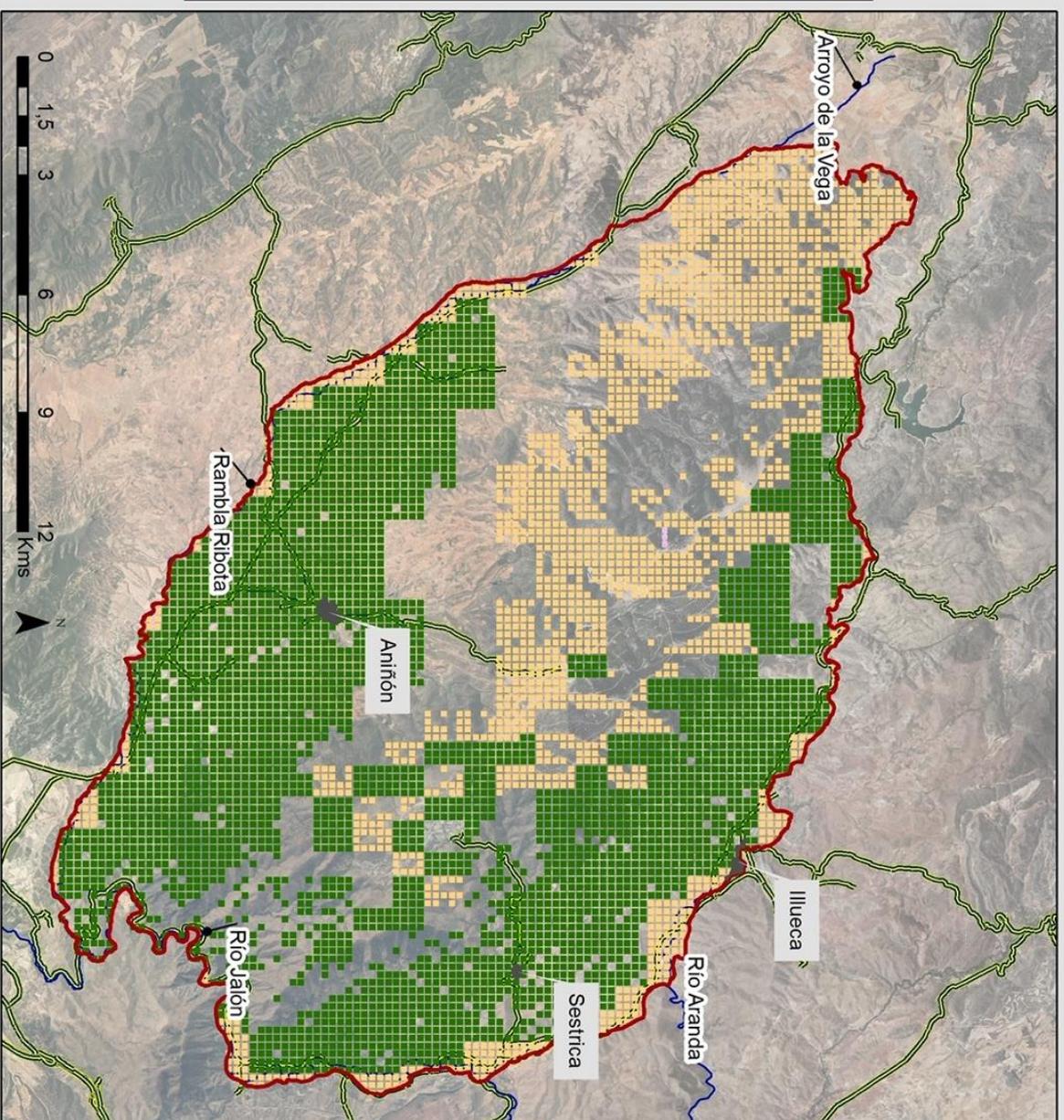
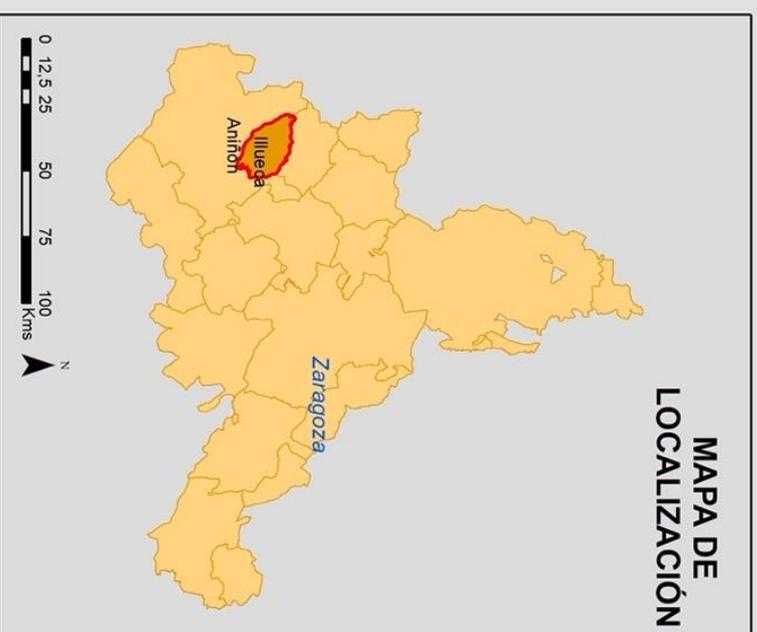
### Otros

- Área de estudio
- Cursos de agua
- Población
- Carreteras

Fuente: Ortofotografía 2018, IGN



# MAPA DE LOCALIZACIÓN FINAL



- ### LEYENDA
- Zonas con mayor probabilidad de albergar la especie:
- Cubierta de Q. suber al 30%
  - Cubierta de Q. suber al 70%
- Otros
- Área de estudio
  - Población
  - Carreteras
  - Río

Fuente: Ortofotografía año 2018, IGN



## 7.1 Tablas

Debido al tamaño de todas las tablas, estas se adjuntan en archivo a parte.

## 7.2 Otros

