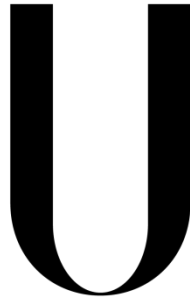


Universidade de Lisboa  
Instituto de Geografia e Ordenamento do Território



**LISBOA**

---

UNIVERSIDADE  
DE LISBOA

**Otimização da gestão territorial das florestas mistas de sobreiro e  
pinheiro-manso: ocupação potencial e produtividade**

**David Lloberas Lafuente**

Orientação de:

Doutor António Manuel Teixeira Monteiro

e

Doutora Alexandra Cristina Pires Correia

Mestrado em Sistemas de Informação Geográfica e Modelação Aplicados ao  
Ordenamento

2023

Universidade de Lisboa  
Instituto de Geografia e Ordenamento do Território



**Otimização da gestão territorial das florestas mistas de sobreiro e pinheiro-manso: ocupação potencial e produtividade**

**David Lloberas Lafuente**

Orientação de:

Doutor António Manuel Teixeira Monteiro

e

Doutora Alexandra Cristina Pires Correia

Júri:

Presidente: Professor Doutor Nuno Manuel Sessarego Marques da Costa, do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa

Vogais:

- Doutor Sven Mutke Regneri, Vicedirector Técnico do Instituto de Ciencias Forestales - ICIFOR/INIA, CSIC
- Doutora Paula Sofia Antunes Matos, Investigadora do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa
- Doutora Alexandra Cristina Pires, Investigadora Auxiliar do Instituto Nacional de Investigação Agrária e Veterinária - INIAV, I.P. (Orientadora)

2023



## AGRADECIMIENTOS

En primer lugar, me gustaría destacar el aprendizaje que ha representado formar parte del Proyecto CORKNUT, principalmente gracias al trabajo constante realizado con su coordinadora y codirectora de esta tesis, la investigadora Alexandra Correia, quien tiene la habilidad de saber compartir y poner en práctica sus conocimientos de manera didáctica y rigurosa. Agradecer también la atención y consejos recibidos por otros miembros del proyecto como las investigadoras Cathy Besson y Encarnação Marcelo, la profesora Margarida Tomé y otros miembros del Centro de Estudos Florestais (CEF).

En segundo lugar, agradezco al investigador del Centro de Estudos Geográficos (CEG), António Monteiro, quien ha codirigido esta tesis y organizó fantásticamente la *COST action* Training School Pen-CaFoRR en Manteigas, lo que me permitió entrar en contacto con grandes ecólogos, geógrafos y forestales a nivel europeo. Del mismo modo, agradezco a los miembros del jurado haberse prestado para esta función, de quién escucharé atentamente sugerencias y críticas el día de la presentación.

En tercer lugar, me gustaría destacar la amabilidad del investigador Pedro Soares — y su equipo del Instituto Dom Luiz de la FCUL— por facilitarnos la base de datos climáticos interpolados IBERIA01 (1970-2015). En este sentido, también destaco la buena disposición de la investigadora Conceição Gonçalves del INIAV por haber disponibilizado la interpolación de la base de datos edáficos INFOSOLO.

Así pues, hago mención de la Fundação para a Ciência e a Tecnologia (FCT) del Ministério de Ciência, Tecnologia e Ensino Superior de la República Portuguesa, quien ha financiado esta investigación a través del ISA-UL con el proyecto *PCIF/MOS/0012/2019 CORKNUT - “Florestas mistas de sobreiro e pinheiro manso: gestão para valorização dos produtos, promoção da biodiversidade e prevenção de fogos florestais”*.

Finalmente me gustaría destacar el empeño de personas como Alexandra Elbakyan, fundadora de Sci-Hub, sin el cual muchos trabajos académicos realizados por estudiantes de diferente nivel se hubieran encontrado con muchas puertas cerradas, por eso aquí está mi reconocimiento a su compromiso por una ciencia abierta y universalmente accesible.

## Resumen

Frente a los diferentes fenómenos ligados al cambio global se hace necesario explorar alternativas de resiliencia y sostenibilidad para los bosques mediterráneos, como lo pueden ser las masas mixtas de *Pinus pinea* y *Quercus suber*, una combinación con posibilidad de ampliar los servicios del ecosistema en comparación con la tradicional gestión forestal monoespecífica. Así pues, este trabajo consiste en la creación de modelos de distribución potencial, basados en GLM y *Random Forests*, para el pino piñonero y el alcornoque, lo que permitirá concretar qué áreas son aptas para este tipo de bosques gestionados. La calibración de los modelos se ha realizado en la mitad sur del país, para, una vez validados, ser aplicados al resto del país.

En este sentido, los modelos se centran solamente en la parte abiótica de la distribución y constatan la gran perturbación antrópica de la ocupación actual, sobre la que se basa la modelación de la distribución potencial. Los resultados muestran que, de las variables edafoclimáticas, una de las más significativas para ambas especies es la textura del suelo, concretamente el % de arena; mientras que la temperatura máxima media y la continentalidad tienen más relevancia en los modelos del pino piñonero, siendo la precipitación la que más tiene para el alcornoque. Existe poco consenso científico sobre los determinantes de la distribución de estas dos especies, lo que hace necesario conocer mejor el impacto de las variables sobre ambas para aumentar el poder explicativo de los modelos en nuestra región.

## Resumo introdutório alargado em português

### DISTRIBUIÇÃO POTENCIAL DE FLORESTAS MIXTAS DE *PINUS PINEA* E *QUERCUS SUBER*: CONSTRUÇÃO DE UM MODELO GEOGRÁFICO PARA PORTUGAL CONTIENTAL

**Palavras chave:** Modelação da distribuição de espécies (SDM), variáveis edafo-climáticas, antropização, *Generalized Linear Model* (GLM), *Random Forests*.

As florestas mistas de Pinheiro-manso e Sobreiro têm vindo a ganhar o interesse de gestores e proprietários florestais porque permitem aliar uma maior rentabilização da exploração florestal, através da produção mista de cortiça e pinha. Uma floresta mista será, em teoria, mais resiliente a um clima em mudança, aos desafios de uma economia global e com um papel importante na melhoria das funções do ecossistema de regulação hídrica, sequestro de carbono e manutenção da biodiversidade.

O investimento na floresta comporta algumas incertezas. O sucesso das plantações está estreitamente ligado à aptidão ecológica do local, o que é determinado pelas condições climáticas, edáficas e orográficas. Porém, no território, existem também constrangimentos económicos e políticos. Reconhecer estes locais é fundamental para garantir a manutenção das espécies e a sustentabilidade na produção. Este trabalho pretende responder a este desafio e apoiar o gestor florestal numa tomada de decisão consciente e informada.

Para tal criaram-se modelos probabilísticos de distribuição de espécies (*Species Distribution Model* - SDM, em inglês) que têm como objetivo, por um lado, permitir: i) identificar a influência de cada variável explicativa na distribuição das espécies; e, por outro, ii) estimar a aptidão do território continental de Portugal para a eventual presença das duas espécies em conjunto.

Os modelos de ocupação potencial têm uma resolução de 500m e basearam-se na distribuição geográfica de a) presenças / ausências atuais da variável dependente — a existência de povoamentos de Pinheiro-manso, Sobreiro ou mistos— e b) os valores das variáveis explicativas. Aplicaram-se dois algoritmos inovadores na literatura científica (Pechi et al., 2019; Santini et al. 2021), pertencentes a duas famílias diferentes, com a finalidade de explorar as suas potencialidades: *Generalized Linear Model* (GLM) e *Random Forests* (RF).

A informação geográfica utilizada para conhecer a atual distribuição de ambas as espécies, como variável dependente, provém do 6º Inventário Florestal Nacional (ICNF, 2015), enquanto que as variáveis explicativas foram extraídas de diferentes fontes:

- Os indicadores climáticos — alguns deles relativos à categorização BIOCLIM — foram calculados com dados de precipitação e temperaturas média, mínima e máxima dos últimos 30 anos provenientes da base de dados interpolada IBERIA01 para o período 1971-2015 (Herrera et al., 2019). As variáveis selecionadas foram: o número de dias de geada no ano (T. mín. < 0°C), a temperatura máxima média em °C, a continentalidade em °C (BIO 07) e a média das precipitações acumuladas anuais em mm (BIO 12).
- As propriedades físico-químicas dos solos, foram compiladas da base de dados INFOSOLO 1966-2014 (Ramos et al., 2017). As variáveis seleccionadas foram o pH e a textura dos solos, mais concretamente a percentagem de areia.
- A Carta de Ocupação de Solos 2018 (DGT) foi usada como variável de perturbação antrópica utilizando camadas *booleanas* de exclusão associadas a determinados usos humanos do território: construção, agricultura e pastagens. Criou-se, adicionalmente, outra camada de exclusão associada a reservatórios de água temporários e permanentes: albufeiras, rios e sapais.

Os modelos foram calibrados para a área de estudo, criada a partir das NUTS III da região sul do país, e incluem as regiões do Ribatejo, Alentejo e Algarve. Posteriormente, são validados com curvas ROC (*Receiver Operating Characteristics*), utilizando-se o valor AUC (*Area Under the Curve*) para comparar o rendimento de cada algoritmo. Contudo, apesar de RF obter melhores resultados do que GLM o processo de validação —ligeiramente para o pinheiro-manso e massas mixtas e bastante para o sobreiro—, decidiu-se a aplicação dos dois algoritmos para fora da área de estudo (resto área de Portugal continental), pelo facto dos modelos de aprendizagem automática poderem estar sobre-calibrados (Jiménez-Valverde et al., 2008).

No que diz respeito aos resultados estatísticos, as variáveis que mais contribuem para a modelação da ocupação de *Pinus pinea* foram a textura arenosa dos solos e a

temperatura máxima média; enquanto que para *Quercus suber*, as variáveis mais significativas foram a percentagem da areia e a precipitação anual. A relação nos três primeiros casos, respetivamente para cada espécie, é positiva, mas no último caso a relação é negativa, sendo o intervalo mais estável para precipitações entre os 600 e 700 mm.

No que se refere à aplicação dos modelos fora da área de estudo e no caso do pinheiro-manso, a representação cartográfica para RF sugere áreas de expansão que o modelo GLM descarta totalmente: regiões do alto Mondego e o Pinhal de Leiria. Já no caso dos modelos para o sobreiro mostram resultados muito parecidos entre si, sugerindo a idoneidade da expansão desta espécie nas regiões da Beira Alta e Trás-os-Montes. Finalmente, a aplicação dos modelos criados com as ocorrências de povoamentos mistos mostra probabilidades muito baixas de ocorrência utilizando ambos os algoritmos, não sendo possível definir áreas de distribuição potencial neste caso.

Consideramos que a forte influência antropogénica na distribuição atual dos povoamentos de pinheiro-manso e sobreiro constitui um forte constrangimento à distribuição potencial modelada deste tipo de floresta mista. Destaca-se também o pouco consenso científico em relação às variáveis abióticas que determinam a distribuição do *Pinus pinea* e de *Quercus suber*. Existem grandes discrepâncias nos modelos criados para estas duas espécies na atualidade, o que mostra a necessidade de continuar a trabalhar interdisciplinarmente de forma melhorar o valor explicativo dos modelos de distribuição de espécies na região mediterrânica. Só assim será possível obter resultados de modelação com a aplicação prática necessária para os diferentes agentes territoriais trabalharem com o mínimo de incerteza possível.



# Índice

Resumen.....	2
Resumo introdutório alargado em português.....	3
Índice.....	6
Índice de Figuras.....	8
Índice de Tablas.....	10
1. Introducción.....	11
1.1 Objetivos y estructura del trabajo.....	12
2. Contexto y aproximación a los bosques gestionados al sur del Tajo.....	15
2.1 ¿Por qué bosques mixtos? Oportunidades e incertezas.....	15
2.2 ¿Por qué pino piñonero y alcornoque? Un abordaje integral.....	17
3. Enfoques metodológicos para la modelación de la distribución potencial de especies: dos familias de algoritmos.....	21
3.1 ¿Qué es un Species Distribution Model (SDM)?.....	21
3.2 Modelos binomiales y de aprendizaje automatizado: <i>Generalized Linear Models</i> (GLM) y <i>Random Forests</i> (RF).....	22
4. Metodología: fuentes de datos, variables, área de estudio.....	24
4.1 Fuentes de datos.....	24
4.1.1 La variable dependiente.....	24
4.1.2 Las variables explicativas.....	25
4.2 Selección de las variables.....	28
4.2.1 Revisión bibliográfica sobre los factores asociados a la distribución del <i>Pinus pinea</i> y el <i>Quercus suber</i> .....	28
4.2.2 Evaluación de la colinearidad estadística entre variables.....	32
4.3 Delimitación del área de estudio.....	36
4.3.1 Definición del área.....	36
4.4 Modelación estadística con <i>GLM</i> y <i>Random Forests</i> .....	40
4.4.1 Calibración del modelo.....	40
5. Resultados.....	42
5.1 Capacidad predictiva de los modelos GLM y RF.....	42
5.1.1 GLM y RF para <i>Pinus pinea</i> .....	42
5.1.2 GLM y RF para <i>Quercus suber</i> .....	43

5.1.3 GLM y RF para mixtos .....	45
5.2 Importancia y contribución de las variables .....	46
5.3 Validación del modelo.....	47
5.4 Ilustración cartográfica de la aplicación de los modelos .....	48
<b>6. Discusión.....</b>	<b>52</b>
6.1 Comprobación de la información de entrada.....	52
6.2 Limitaciones de los modelos.....	54
6.3 Operacionalización de los bosques mixtos en las áreas de expansión.....	56
<b>7. Conclusiones.....</b>	<b>58</b>
<b>8. Bibliografía.....</b>	<b>61</b>
<b>Anexos .....</b>	<b>65</b>
Anexo 1: Test de autocorrelación .....	65
Anexo 2: Calibración del modelo .....	65
Anexo 3: Validación del modelo .....	67
Anexo 4: Transferibilidad del modelo.....	67

## Índice de Figuras

Figura 1. Bosque mixto de pinos y quercíneas cerca del embalse de Odivelas. Fotografía propia.....	16
Figura 2. Cosecha manual de piña cerca del embalse de Montargil. Fotografía propia .....	18
Figura 3. Extracción del corcho en el Vale do Sorraia. Fotografía propia. ....	18
Figura 4. . Evolución del precio medio del piñón blanco 2006-2021 en la Llotja de Reus. Datos divulgados por el Observatori Forestal Català (OFC) muestran un encarecimiento repentino en los últimos 3 años. Fuentes: 2006 – 2017: UNAC (2017); 2018 – 2021: OFC. ....	19
Figura 5 . Evolución del precio medio del corcho entre 2009 y 2018. Fuente UNAC, 2018. ....	19
Figura 6. Pila de piñas a la espera del proceso industrial que permitirá la extracción del piñón.....	20
Figura 7. Pila de planchas de corcho acabadas de extraer siendo transportadas hacia el secadero. ....	20
Figura 8. Sumario de variables preseleccionadas.....	33
Figura 9. Matriz de correlación entre las variables definitivamente seleccionadas. .	34
Figura 10. Número de días T <sub>min</sub> <0°C, Temperatura máxima media (°C), Continentalidad / BIO07 (°C), Precipitación / BIO12 (mm), Arena (%), pH. Fuentes: IBERIA01 & INFOSOLO. Elaboración propia.....	35
Figura 11. a) Distribución actual de Pinus pinea, b) Quercus suber y c) masas mixtas (6 <sup>o</sup> IFN), d) usos del suelo considerados como antropizantes (COS18) y e) elevación (MDE, Copernicus). Elaboración propia. ....	37
Figura 12. Extensiones agro-silvo-pastoriles alentejanas, finca situada en la Ribeira de Canha. Corknut. ....	38
Figura 13. Número de días T <sub>min</sub> <0°C, Temperatura máxima media (°C), Continentalidad / BIO07 (°C), Precipitación / BIO12 (mm), Arena (%), pH. Fuentes: IBERIA01 & INFOSOLO. Elaboración propia.....	39
Figura 14. Curvas de respuesta de los modelos GLM (rojo) y RF (azul) para el Pinus pinea. Por orden: Precipitación / BIO12 (mm), Número de días T <sub>min</sub> <0°C, Continentalidad / BIO07 (°C), Temperatura máxima media (°C), Arena (%) y pH....	43
Figura 15. Curvas de respuesta de los modelos GLM (rojo) y RF (azul) para el Quercus suber. Por orden: Precipitación / BIO12 (mm), Número de días T <sub>min</sub> <0°C, Continentalidad / BIO07 (°C), Temperatura máxima media (°C), Arena (%) y pH....	44
Figura 16. Curvas de respuesta de los modelos GLM (rojo) y RF (azul) para masas mixtas. Por orden: Precipitación / BIO12 (mm), Número de días T <sub>min</sub> <0°C, Continentalidad / BIO07 (°C), Temperatura máxima media (°C), Arena (%) y pH....	45
Figura 17. Curvas ROC para Pinus pinea (a), Quercus suber (b) y mixtos (c).....	47
Figura 18. Aplicación de los modelos GLM y RF para el Pinus pinea en Portugal continental, probabilidad de presencia entre 0 y 1. ....	49
Figura 19. Aplicación de los modelos GLM y RF para el Quercus suber en Portugal continental, probabilidad de presencia entre 0 y 1. ....	50

Figura 20. Aplicación de los modelos GLM y RF para masas mixtas de Pinus pinea y Quercus suber en Portugal continental, probabilidad de presencia entre 0 y 1. .... 51

Figura 21. Localización de las parcelas de monitorización permanente de los proyectos Corknut (mixtos), Fertipinea (pino piñonero) y Nutrisuber (alcornoque). . 53

## Índice de Tablas

Tabla 1. Tabla resumen de las fuentes de datos. ....	27
Tabla 2. Variables climáticas para Pinus pinea y Quercus suber. En verde las variables preliminarmente seleccionadas, de color más oscuro aquellas finalmente seleccionadas. ....	29
Tabla 3. Variables edáficas y orográficas para Pinus pinea y Quercus suber. En verde las variables preliminarmente seleccionadas, de color más oscuro aquellas finalmente seleccionadas. ....	30
Tabla 4. Sumario de variables preseleccionadas. ....	32
Tabla 5. Cálculo de la Variance Inflation Factor (VIF) de cada variable seleccionada. ....	33
Tabla 6. Significancia estadística de cada variable en el modelo GLM Pinus pinea. ....	42
Tabla 7. Significancia estadística de cada variable en el modelo GLM Quercus suber. ....	44
Tabla 8. Significancia estadística de cada variable en el modelo GLM de masas mixtas. ....	45
Tabla 9 Importancia y contribución de las variables. ....	46

## 1. Introducción

Los bosques mixtos de *Pinus pinea* y *Quercus suber* han despertado interés entre la comunidad científica para investigar su posible capacidad de hacer más resilientes las masas forestales al cambio global. En concreto, los bosques situados en la región mediterránea, especialmente vulnerable a la desertificación, donde los incendios se han incrementado y se tiende hacia una aridez cada vez mayor.

Analizar los actuales y futuros problemas territoriales a través del concepto de cambio global es importante para entender las múltiples causas geográficas de estos: cambio climático, éxodo rural, concentración de la propiedad, integración de mercados, etc. Las formaciones mixtas en áreas boscosas no son algo nuevo, al contrario, ya que fue precisamente esta diversidad lo que permitió a las comunidades, a lo largo de la historia, vivir de los variados productos que se ellas se extraían. Sin embargo, con la irrupción del modelo forestal monoespecífico, inspirado en los principios asentados por la silvicultura centroeuropea, se asentaron las dinámicas de especialización que actualmente rigen las masas forestales productivas.

En este sentido, actualmente existen pocos fenómenos estrictamente regidos por razones naturales por encontrarnos en un periodo que algunos autores designan como *Antropoceno* (Crutzen, 2006; Castree, 2015) o como *Urbanización planetaria* (Lefebvre, 1970; Brenner, 2015). La total antropización de los fenómenos terrestres y la expansión de las dinámicas urbanas más allá de lo construido ha hecho que sea casi imposible estudiar cualquier elemento de la realidad, incluso ambiental, sin tener en cuenta la huella de la Humanidad sobre el planeta.

Las masas forestales en forma de dehesa —“montado” en portugués— constituyen un ejemplo paradigmático de esta incidencia antrópica sobre el territorio, ya que se trata por sí misma de una formación boscosa creada por la acción humana para el aprovechamiento agro-silvo-pastoril del “saltus”. Precisamente es en este tipo de formación forestal gestionada que la mixtura entre pino piñonero y alcornoque puede ser una fórmula de éxito. Así pues, con este análisis geográfico se quiere incidir sobre el territorio mediante una planificación que aborde esos desafíos, y que se dirija a satisfacer las necesidades de las poblaciones.

## 1.1 Objetivos y estructura del trabajo

Así pues, la presente tesis de máster pretende entender en qué condiciones los bosques mixtos de pino piñonero (*Pinus pinea*) y alcornoque (*Quercus suber*) pueden ser una alternativa adaptativa del territorio a las dinámicas de la globalización y los cambios multiescalares que esta lleva asociados. En concreto, se busca identificar la distribución geográfica potencial de las masas formadas por estas dos especies en Portugal continental. Su finalidad consiste en dotar a los agentes territoriales de información que permita a las comunidades obtener un mayor aprovechamiento de los servicios del ecosistema prestados por estos bosques, contribuyendo así para su desarrollo a escala local y regional.

El trabajo consiste fundamentalmente en un ejercicio práctico, la creación de tres modelos de distribución de especies (*Species Distribution Model* - SDM), para el pino piñonero, el alcornoque y las masas mixtas. Siendo así una aportación más al vasto campo de la creación de modelos explicativos para fenómenos de índole espacial, misión que la revolución teórico-cuantitativa encomendó a nuestra ciencia, la Geografía (Shaefer, 1953; Chorley & Hagget, 1967). En consecuencia, esta es una tarea que ha ocupado y ocupa numerosos geógrafos, ecólogos, forestales y otros académicos o planificadores, ya que se trata de un asunto con relevancia científica pero también aplicada.

Por lo tanto, se define la distribución potencial del bosque mixto de *Pinus pinea* y *Quercus suber*, y aquello que la condiciona, como el objeto de estudio sobre el que se desarrollará este trabajo. Para ello, se usa como caso específico la ocupación máxima que estas dos especies podrían tener en Portugal continental con los condicionantes ambientales y humanos actuales, existentes dentro de un área de estudio previamente delimitada.

Los modelos probabilísticos creados tienen como objetivo: i) identificar la influencia de cada variable sobre la distribución de las especies; ii) conocer qué abordaje de modelación tiene mayor robustez; y iii) estimar la adecuación del territorio para la eventual presencia de las dos especies estudiadas. La modelación se desarrolla en dos fases: la creación del modelo (calibración y validación) y su posterior aplicación.

Los modelos de distribución potencial tienen una resolución de 500m, y están basados en la distribución geográfica de a) las presencias / ausencias actuales de la variable dependiente y b) los valores de las variables explicativas. Se han aplicado dos algoritmos altamente populares en la literatura científica con la finalidad de explorar las posibilidades que plantea cada uno de ellos, *Generalized Linear Model* (GLM) y *Random Forests* (RF), pertenecientes a familias diferentes, el primero a los modelos logísticos y el segundo a la inteligencia artificial (Pecchi et al., 2019; Santini et al., 2020).

La información geográfica utilizada para conocer la actual distribución de ambas especies, como variable dependiente, proviene del 6º Inventário Florestal Nacional (ICNF, 2015). Mientras que la información relativa a las variables explicativas, de carácter abiótico, está constituida por diferentes datos procedentes de diversas fuentes.

En primer lugar, los indicadores climáticos —algunos de ellos pertenecientes a la categorización mundial BIOCLIM— están calculados con los datos de precipitación y temperaturas media, máxima y mínima, de la base de datos interpolada IBERIA01 1971-2015. En segundo lugar, las características fisicoquímicas de los suelos, las variables edáficas, provienen de la compilación de datos INFOSOLO 1966-2014 realizadas por el INIAV (Ramos, 2017).

Además, se han consultado factores orográficos como la altimetría, la orientación y la pendiente, extraídas del Modelo Digital de Elevación que disponibiliza el programa Copernicus. Otras variables como la categorización litológica y edáfica, extraídas del Atlas do Ambiente, no se han podido introducir en el modelo por su naturaleza cualitativa. Por último, la Carta de Ocupação de Solos 2018 (DGT) se ha usado para crear diferentes capas booleanas asociadas a usos antropizados del territorio: agricultura y pastos, medio construido; y otra asociada a las masas de agua: embalses, ríos, humedales y marismas.

En conclusión, esta es una tarea exploratoria alrededor de la distribución potencial que pueden tener estas dos especies forestales y los factores que la condicionan. Se parte del presupuesto de que el área de estudio está constituida por espacios que no escapan a dinámicas antropizantes actuales o pretéritas, lo que constituye una



perturbación que condiciona las distribuciones actuales. Se trata de un trabajo con vocación aplicada que no renuncia a la reflexión teórica científica.

Esta tesis está enmarcada en el proyecto Corknut, financiado por la Fundação para a Ciência e a Tecnologia (FCT), cuya misión es profundizar en todas las oportunidades que presentan los bosques mixtos de las especies ya mencionadas. El proyecto *PCIF/MOS/0012/2019 CORKNUT - “Florestas mistas de sobreiro e pinheiro manso: gestão para valorização dos produtos, promoção da biodiversidade e prevenção de fogos florestais”* se desarrolla en cooperación entre el Instituto Nacional de Investigação Agrária e Veterinária (INIAV, IP) y el Instituto Superior de Agronomia - Universidade de Lisboa (ISA-UL).

## **2. Contexto y aproximación a los bosques gestionados al sur del Tajo**

### **2.1 ¿Por qué bosques mixtos? Oportunidades e incertezas**

Las masas forestales mixtas constituyen sistemas complejos, ya que implican una mezcla de dos o más especies, con características morfológicas y fisiológicas diferentes, que interactúan entre sí mediante relaciones ecológicas de complementariedad, facilitación y competición todavía difíciles de comprender en la naturaleza (Pretsch et al. ed., 2017; Coll et al., 2018; del Río et al., 2021). Es a causa de esta complejidad, y de la incertidumbre que inevitablemente genera, que todavía no existe un cuerpo teórico y metodológico sólido que pueda apoyar a los agentes en la constitución de bosques mixtos. Actualmente apenas están establecidas unas bases sobre las que se lleva unas décadas avanzando (del Río et al., 2016).

Todavía la gestión forestal está muy vinculada a los principios básicos de la silvicultura centroeuropea de hace más de un siglo, principalmente enfocada hacia la producción maderera por parte de bosques monoespecíficos, mucho más fáciles de gestionar y estudiar que los mixtos. Sin embargo, múltiples elementos asociados al cambio global como el incremento de episodios climáticos extremos, la propagación de plagas, y la creciente vulnerabilidad a incendios forestales, están haciendo cambiar ese paradigma (Corknut, 2021) (*fig. 1*).

Se ha observado que son múltiples las cuestiones a tener en cuenta para evaluar la idoneidad de las especies que pueden constituir las masas mixtas, lo que a menudo está más relacionado con la selección de las especies que con su variedad (del Río et al., 2021). En Löff et al. (2018) se mencionan algunas interacciones que pueden ser consideradas como compatibles y otras como incompatibles que impactan sobre los procesos ecológicos.



*Figura 1. Bosque mixto de pinos y quercíneas cerca del embalse de Odivelas. Fotografía propia.*

Por un lado, el autor indica factores como la polinización y la dispersión de las semillas; y, por otro, están factores como la competición por recursos entre cubiertas en los momentos de recepción de luz solar o entre los sistemas radiculares por la captación de nutrientes o de recursos hídricos. En este sentido, otra de las claves —más allá de la proporción de capa especie— se encuentra en la densidad de las masas arbóreas y arbustivas, como ya lo es en las masas monoespecíficas (del Río et al., 2016).

En el caso de que las diferentes cuestiones planteadas se encuentren mayoritariamente en la vertiente de la compatibilidad y no de la incompatibilidad, las masas mixtas pueden ver mejorado el desempeño de algunas de sus funciones ecológicas. Ciertamente parecen existir muestras incipientes en relación con la mejora de ese desempeño, provocando así el aumento de productividad (Liang et al., 2016), incremento de la regeneración natural (Correia et al, 2018) y mayor resistencia a eventos climáticos —sequías, inundaciones e incendios— (Bauhus et al. in Pretzsch, 2017) y amenazas bióticas (Jactel et al. 2017; 2021).

## 2.2 ¿Por qué pino piñonero y alcornoque? Un abordaje integral

Tanto alcornocales como pinares tanto aislada como conjuntamente, además de ser muy interesantes desde el punto de vista económico, presentan una buena adaptación a condiciones climáticas adversas a las que la cuenca mediterránea se ve sometida durante los meses de verano. Así pues, el conocimiento ya adquirido a lo largo de generaciones y las altas productividades registradas en la mitad sur del país, hacen de estas dos especies una posibilidad suficientemente atractiva para los propietarios, como alternativa a las plantaciones altamente rentables de eucalipto, en un territorio en que la propiedad de los terrenos está cerca de ser totalmente privada.

La mixtura entre estas dos especies no es algo nuevo, puesto que ya coexisten espontáneamente por regeneración natural en terrenos abandonados o poco gestionados (Tomaz, 2013). Existe bibliografía sobre las relaciones de sucesión ecológica entre pinos y quercíneas, y el papel facilitador que tienen los primeros sobre los segundos para establecerse en *late-sucesional* áreas (Pausas et al., 2004; Urbietta et al., 2011). A medida que los bosques llegan a etapas adultas no existe consenso si las relaciones son de facilitación (Prieto et al. 2012) o de competición (Prévosto et al., 2016), dada la escasez de agua y nutrientes en la región mediterránea.

Por un lado, el pino piñonero y el alcornoque, como otras especies forestales, son importantes desde el punto de vista ecosistémico para proveer en estas regiones servicios del ecosistema regulatorios y de apoyo como: fijación de carbono (Nunes & Lopes, 2016), retención hídrica, enriquecimiento de los suelos (Richards et al., 2010; Pereira et al., 2011), abrigo a la biodiversidad, etc.

Por otro lado, estos bosques necesitan ser gestionados para poder aportar con éxito valor social al territorio en forma de servicios, tanto a nivel económico como cultural: cosechas anuales de piña (*fig. 2 & 6*), aprovisionamiento rotativo de corcho (*fig. 3 & 7*) —y madera en algunos casos—, bellotas y condiciones para el desarrollo de hierba para la alimentación de ganado y caza; al mismo tiempo se aporta un valor estético caracterizante del paisaje del Alentejo y Ribatejo. La dehesa constituye un sistema agro-silvo-pastoril con alto valor paisajístico.



*Figura 2. Cosecha manual de piña cerca del embalse de Montargil. Fotografía propia.*

En consecuencia, existe toda una economía local enfocada para tales efectos, con todo lo que ello comporta para un área con bajas densidades poblacionales que está relativamente cerca de Lisboa. Buena parte de Ribatejo y Alentejo constituye el hinterland de la capital, tanto a nivel funcional como también simbólico. Así pues, el desarrollo de las actividades de explotación de recursos conlleva asociadas una serie de contradicciones: urbano-rural, propiedad-trabajo, global-local, humanidad-naturaleza (Morcuende, 2021).



*Figura 3. Extracción del corcho en el Vale do Sorraia. Fotografía propia.*

A escala global, no es menor la relevancia de los productos alfa extraídos en estas regiones portuguesas del suroeste ibérico, ya que en 2010 representaban el 40% (Agri-ciência, 2014) y el 49,6% (APCOR, 2020) de la producción mundial de piñón y corcho respectivamente. En los últimos años el precio ha sido fluctuante, tendiendo a aumentar, situándose alrededor de los 45€/kg de media para el piñón blanco (fig. 4) y a poco menos de los 32€/@ en el caso del corcho (fig. 5).

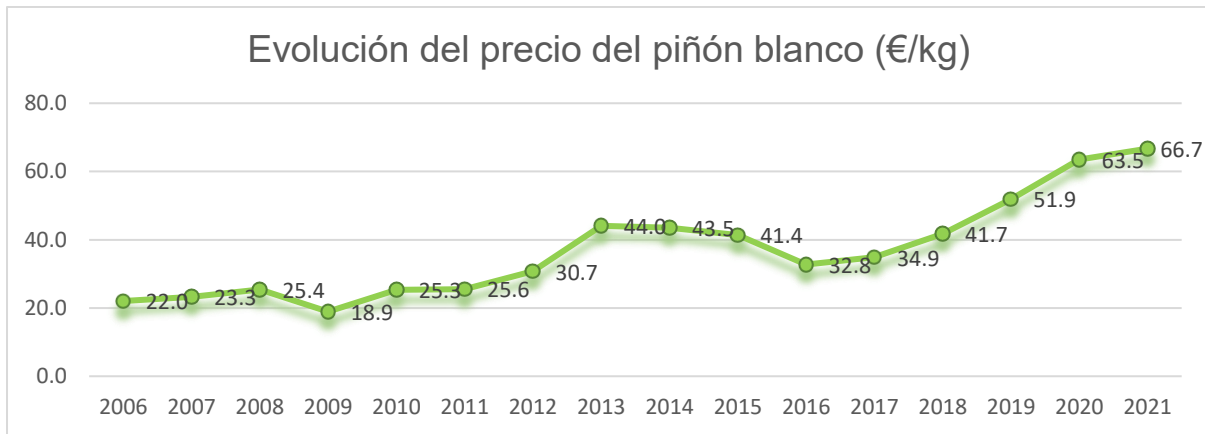


Figura 4. . Evolución del precio medio del piñón blanco 2006-2021 en la Llotja de Reus. Datos divulgados por el Observatori Forestal Català (OFC) muestran un encarecimiento repentino en los últimos 3 años. Fuentes: 2006 – 2017: UNAC (2017); 2018 – 2021: OFC.

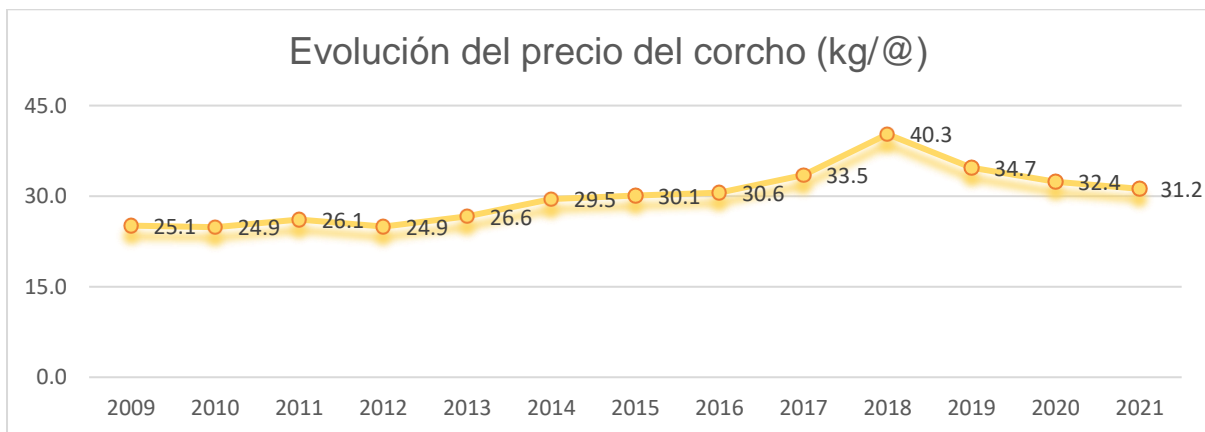


Figura 5. . Evolución del precio medio del corcho entre 2009 y 2018. Fuente UNAC, 2018.

Está claro que con estas características se trata de un sector estratégico cuyo resultado son productos diferenciados muy atractivos para su consumo o transformación en todo tipo de mercados. Por lo tanto, en un mundo cambiante es de gran interés conocer mejor cuales son y donde se dan las condiciones que facilitan o dificultan la existencia de los árboles que los producen.



*Figura 6. Pila de piñas a la espera del proceso industrial que permitirá la extracción del piñón.*



*Figura 7. Pila de planchas de corcho acabadas de extraer siendo transportadas hacia el secadero.*

### **3. Enfoques metodológicos para la modelación de la distribución potencial de especies: dos familias de algoritmos**

#### **3.1 ¿Qué es un Species Distribution Model (SDM)?**

La modelación de la distribución de especies — comúnmente representada en la bibliografía anglosajona como *Species Distribution Modeling*, bajo el acrónimo de SDM— consiste en el tratamiento estadístico de datos empíricos espacialmente explícitos. En la bibliografía, el mismo concepto a menudo aparece también denominado como *habitat suitability model*, *niche-based model*, *bioclimatic model* *environmental niche model* (Pecchi et al., 2019)

Guisan & Thuiller (2005) definen los SDM como “*empirical models relating field observations to environmental predictor variables*”. Al mismo tiempo, se trata, como indican Guisan & Zimmermann (2000), de modelos que en principio son estáticos y probabilísticos, ya que relacionan la distribución geográfica de las especies presentes con las condiciones ambientales. Así pues, estos modelos sirven, en última instancia, para conocer cuál es el nicho ecológico de una especie y ii) donde se dan las condiciones ambientales que definen ese nicho.

En otras palabras, Mateo et al. (2013) definen estos modelos como “representaciones cartográficas de la idoneidad de un espacio para la presencia de una especie en función de las variables empleadas para generar dicha representación”, entendiendo esta idoneidad como “la relación estadística entre la distribución real conocida y un conjunto de variables independientes que se usan como indicadores”.

Los precedentes de los SDM se encuentran en la creación de correlaciones entre presencias y factores climáticos durante el primer y segundo tercio del s. XX, y su origen se concretiza a mediados de la década de 1970, con los inicios de la computación para desarrollar los primeros modelos predictivos (Guisan & Thuiller, 2005). Estos autores consideran la obra de Nix et. al. (1977), en la que se generaban predicciones espaciales del nicho ecológico de especies arbustivas, como uno de los primeros SDM conocidos.



A lo largo de los últimos años la expansión de su uso ha sido exponencial y hoy en día son ampliamente usados para la ordenación del territorio, concretamente para la restauración de hábitats o la introducción de especies en determinadas áreas (Elith & Franklin, 2013). Por otro lado, actualmente, el mayor uso que se está dando a este tipo de modelación es el de prever potenciales impactos derivados directa o indirectamente del cambio climático sobre la distribución de las especies (Santini et al., 2021).

Este tipo de modelación basada en las presencias y ausencias de una especie presenta algunos aspectos que según Jiménez-Valverde et al. (2008) hay que tener en cuenta, concretamente i) la distinción entre la distribución potencial y la real, ii) los efectos de las muestras sobre la evaluación de los modelos y, por último, iii) las posibles imprecisiones de los modelos en su extrapolación y aplicación a áreas ajenas a la modelación.

### **3.2 Modelos binomiales y de aprendizaje automatizado: *Generalized Linear Models (GLM) y Random Forests (RF)***

Existen diferentes algoritmos aplicables a la construcción de un *SDM*, varios de ellos pertenecientes a diferentes familias estadísticas y de computación. Es de consenso general que no existe el algoritmo definitivo, y que la elección de uno u otro tiene que depender de las características y necesidades del objeto de estudio; sin embargo, es también consensual que la elección del algoritmo correcto puede reducir notablemente la incertidumbre del modelo (Pecchi et al., 2019).

Los criterios más comunes para clasificar los algoritmos son i) si funcionan mediante presencias-ausencias o solo presencias de las variables independientes y ii) si están basados en regresiones o en *machine learning*. Siguiendo este segundo criterio, en el primer grupo se encuentran: los *General Linear Models (GLM)*, *Generalized Additive Models (GAM)* y los *Multivariate Adaptive Regression Splines (MARS)*; mientras que en el segundo grupo están los: *Random Forests (RF)*, *Maximum Entropy (MaxEnt)*, *Boosting Regression Tree (BRT)*, *Classification and Regression Trees (CART)*, *Generic Algorithm for Rule-set Prediction (GARP)*, *Artificial Neural Networks (ANN)*.

Según diversas revisiones realizadas alrededor de la literatura científica, el algoritmo más popular a lo largo de esta última década ha sido MaxEnt, seguido de GLM y RF,

encabezando las familias de algoritmos lineales y los *machine learning* (Pecchi et al., 2019; Santini et al., 2020). RF destaca por su desempeño predictivo superior a otros algoritmos del mismo tipo (Fukuda et al. 2013), aunque haya otros estudios en que se critique su conservadurismo materializado en una baja sensibilidad y una alta especificidad, presentando algoritmos como MaxEnt y GLM como más precisos (Shabani et al., 2016). RF también ha recibido críticas por sus elevados costes computacionales en comparación con otros *machine learning* como MaxEnt, MARS y BRT (García-Callejas & Araújo, 2015).

Por un lado, los *Generalized Linear Models* (GLM) forman parte de la familia de los clásicos modelos regresivos, en los que se basan las funciones paramétricas para relacionar la variable dependiente con una combinación de variables explicativas (Shabani et al., 2016). Los tipos más comunes de GLM funcionan con regresiones lineales, logísticas o de Poisson, siendo este segundo tipo óptimo cuando la variable dependiente es booleana —presencia / ausencia—, mientras que el último se usa cuando esta es cuantitativa —abundancia— (Li & Wang, 2013). Shabani et al. (2016) subrayan la sensibilidad de este algoritmo a la significancia de las variables, por tanto, el estudio recomienda la eliminación de las variables poco significativas para asegurar para una correcta calibración.

Por otro lado, *Random Forests* (RF) es un algoritmo basado en *machine learning*, que consiste en la recursiva partición de los datos. El algoritmo trabaja tanto con variables cualitativas con *classification trees* como con cuantitativas con *regression trees* (Valavi et al., 2021). RF es eficiente para reconocer patrones (Luan et al. 2020) y, como indican Liaw & Wiener (2002), funciona mediante el siguiente procedimiento: i) extracción de muestras del conjunto introducido para generar múltiples subconjuntos; (ii) creación árboles de regresión en crudo con las muestras de extraídas, se selecciona aleatoriamente un subconjunto de variables para definir la división, y se elige la mejor división; y (iii) las predicciones se realizan haciendo el promedio de las predicciones de los árboles de regresión.

## **4. Metodología: fuentes de datos, variables, área de estudio**

Con el objetivo de identificar la distribución geográfica potencial de ambas especies estudiadas —y, en consecuencia, de las masas mixtas—, la primera etapa de la modelación consiste en: i) identificar las fuentes disponibles de datos; ii) seleccionar las variables basándose, por un lado, en experimentos anteriores consultables en la selección bibliográfica, y por otro, en tests estadísticos que eviten colinearidad entre ellas; y iii) la definición del área de estudio en que se ha validado y calibrado los modelos estadísticos GLM y RF. Cada una de estas fases está descrita con detalle a continuación.

### **4.1 Fuentes de datos**

Los datos usados para la construcción del modelo provienen de diversas fuentes, cuyo origen se ha intentado que fuese lo más próximo posible al área de estudio. Haber dado preferencia a datos provenientes de agencias nacionales, o de proyectos a escala ibérica, ha permitido tener resoluciones bastante altas para todas las variables. Además, la vinculación de las fuentes con la tarea investigadora y planificadora —universidades e institutos públicos— garantiza un aceptable grado de fiabilidad de la información geográfica.

Al tratarse de una modelación para Portugal continental, y, por tanto, dentro de las fronteras administrativas del Estado, no ha sido necesario recorrer a muchas fuentes europeas, tan útiles para proyectos internacionales de escala regional subcontinental. En este caso, cuando se han usado datos provenientes de organismos europeos ha sido para complementar, o compararlos con, las fuentes a nivel nacional.

#### **4.1.1 La variable dependiente**

El Instituto de Conservação da Natureza e das Florestas (ICNF) realiza cada diez años el Inventário Florestal Nacional (IFN), del que ha usado su 6ª edición —que tiene 2015 como año de referencia— para conocer la ocupación actual de las dos especies estudiadas en cada *fotoponto*, una presencia principal o secundaria. La proyección utilizada es la ETRS89-PT/TM06.

Los datos se han obtenido mediante técnicas mixtas de fotointerpretación de ortofotos a 4 bandas y con la realización de trabajo de campo, para lo cual se inventariaron 11

649 parcelas de muestra (ICNF, 2015). Los datos están estructurados vectorialmente en una malla regular de puntos con una resolución de 500m que cubre toda la superficie del país, lo que se materializa en 360 000 de los llamados “*fotopontos*”.

Al ser una malla regular que alberga la información correspondiente a las presencias / ausencias, es decir, a la variable dependiente, se ha entendido oportuno extraer los valores de las variables dependientes. Además, se ha considerado su resolución de 500m como la adecuada para nuestra escala de trabajo, tratándose también de un término medio entre las resoluciones de las posibles variables independientes. Por tanto, se concretiza esta malla de puntos como estructura de datos seleccionada para realizar la modelación de la distribución potencial.

#### **4.1.2 Las variables explicativas**

##### **a) *Climáticas***

Las variables climáticas se han obtenido de IBERIA01, una base de datos climáticos interpolados para una matriz regular de 0'1° de la Península Ibérica, una resolución de 9'967 km en nuestras latitudes (Herrera et al., 2019). Esta base de datos consta de valores diarios de precipitación, evapotranspiración potencial y temperaturas media, mínima y máximas del periodo 1971-2015; aunque para este trabajo se haya usado el intervalo 1986-2015. IBERIA01 es heredera de interpolaciones de precipitación a resoluciones menores, a 0'2°, como Spain02 (Herrera et al., [2010] 2012) y PT02 (Belo-Pereira et al., 2011), ambas fusionadas en IB02, que no constituía una malla homogénea por tratarse de la unión de dos conjuntos de datos diferentes. Se trata, pues, de una de las bases de datos más detalladas para la región.

La interpolación de IBERIA01 se ha desarrollado con la consulta de una densa red de 3486 estaciones con observaciones válidas para la precipitación y 275 para temperatura en toda la península, datos proporcionados por la Agencia Estatal de Meteorología (AEMET) y el Instituto Português do Mar e da Atmosfera (IPMA) (Herrera et al., 2019, p. 1949). Se usa un método de interpolación en dos pasos basado en un krigeaje ordinario y un cálculo de superficie tridimensional, es decir, se introduce la altitud como covariable de la interpolación (Herrera et al., 2019, p. 1951). La exactitud de esta se ha comparado con la versión más reciente de E-OBS (Cornes

et al., 2018), la v17e, matriz climática de referencia para datos diarios de precipitación y temperatura a nivel europeo.

### **b) Edáficas**

Los datos relativos a las propiedades fisicoquímicas de los suelos se han obtenido del repositorio INFOSOLO, herramienta desarrollada recientemente desde el INIAV. Su cometido es el de dar acceso a una recopilación de muestras provenientes de 3461 perfiles edáficos, tomadas entre 1966 y 2014 por diferentes organismos oficiales, como: la Direção Regional de Agricultura do Norte (36%), Direção Geral de Agricultura (21%), INIAV (18%) o el Joint Research Center (13%), entre otros (Ramos et al., 2017, p. 392). Muestras todas ellas tomadas a lo largo de los últimos 50 años con objetivos e instituciones muy heterogéneas, y con una distribución espacial todavía más. Los perfiles de este último organismo, el JRC, corresponden las 465 muestras del programa LUCAS Topsoil.

INFOSOLO disponibiliza una interpolación en estructura matricial a una resolución de 1km, en proyección Lisboa Hayford Gauss del IGeoE, realizada con 2990 muestras en superficie “topsoil” (<25cm) de las ya mencionadas 3461 totales (Ramos et al., 2017, p. 400). Se trata de variables todas ellas cuantitativas, entre las que hay: pH, carbono orgánico y capacidad de intercambio catiónico, por un lado; y, por otro, valores de textura (%): porcentaje de arena, limo y arcilla. Para interpolar la textura se ha usado un krigeaje simple, mientras que para los primeros (pH, CO, CIC) se ha usado un krigeaje incorporando los segundos, entre otros, como covariables a través de un modelo lineal mixto (Ramos et al., 2017, p. 397). En consecuencia, se dará preferencia a la textura sobre las variables químicas durante la calibración del modelo.

La calidad de la interpolación de las variables de INFOSOLO anteriormente mencionadas se ha comparado con aquella realizada con la base de datos LUCAS - Topsoil, para la que se ha seguido el mismo método que con la primera. LUCAS - Topsoil es un programa paneuropeo que pretende estudiar los suelos a escala continental a través de 20 000 perfiles de superficie (Tóth et al., 2013), de los cuales 465 están homogéneamente distribuidos por Portugal continental. A pesar de que la distribución de los puntos de INFOSOLO no está diseñada expresamente para la interpolación de los datos extraídos, el trabajo de Ramos et al. (2017, p. 409) concluye que esta es igualmente rigurosa a la desarrollada con datos de LUCAS - Topsoil, y

que puede tener una exactitud mayor en determinados niveles regionales en que haya una mayor densidad de muestras (norte de Portugal).

### **c) Orográficas**

La única fuente europea que se ha usado ha sido para la altimetría, en concreto el modelo digital de elevación (DEM) creado en el marco del programa europeo de observación de la tierra Copernicus. El modelo está proyectado en ETRS89 LAEA (Lambert Azimutal Equal Area), la estructura es matricial y su resolución espacial es de 1 arc-second  $00^{\circ} 00.001'$  ( $0.00028^{\circ}$  decimales), lo que son 25m en nuestras latitudes. El EU-DEM cubre todos los estados del Espacio Económico Europeo (EEA39, en inglés), cuyos datos se obtuvieron entre 2003 y 2009 desde el satélite ICESat mediante el Geoscience Laser Altimeter System (GLAS) (Tøttrup & Sørensen, 2014).

### **d) Usos del suelo**

Por último, se ha usado una fuente complementaria, la Carta de Uso e de Ocupação do Solo (COS18), a modo de máscara booleana para definir las perturbaciones antrópicas que afectan la distribución basada en las variables abióticas. Actualmente la COS de referencia es la edición que publicó en 2018 la Direção-Geral do Território (DGT). Sus precedentes anteriores desde su primera edición en 1990 son las versiones de 1995, 2007, 2010, 2015 (Caetano et al. 2017). La información geográfica se ha obtenido mediante fotointerpretación de ortofoto RGB e infrarrojo próximo, producto de vuelos a 3200m y 4300m de altitud realizados en conjunto por de la DGT, el IFAP y el ICNF. A nivel de escala, la unidad mínima cartográfica es de 1ha y la distancia mínima entre polígonos es de 20m; la proyección es ETRS89/TM06; y su estructura de datos vectorial (DGT, 2019).

*Tabla 1. Tabla resumen de las fuentes de datos.*

Producto	Fuente	Año	Escala	Proyección	Estructura	Tipo	Ámbito
6°IFN	ICNF	2015	500m	ETRS89-PT/TM06	Vectorial de punt.	Cualitativa	Portugal
IBERIA01	CSIC	2019	10 000m	WGS84	Matricial	Cuantitativa	Iberia
INFOSOLO	INIAV	2017	1000m	Lisboa Hayford Gauss del IGeoE	Matricial	Cuantitativa	Portugal
MDE	Copernicus	2003-09	25m	ETRS89 LAEA	Matricial	Cuantitativa	EEA-39
6°COS	DGT	2018	1ha	ETRS89-TM06	Vectorial de políg.	Categorica	Portugal

## **4.2 Selección de las variables**

Delante la diversidad de fuentes y materiales, para determinar aquellas variables independientes más útiles con las que poder generar nuestro modelo, se ha realizado a un proceso de selección en dos fases. La primera fase es de carácter bibliográfico, basada en artículos científicos que relatan la construcción de modelos parecidos al nuestro, y que ha consistido en la recopilación de las variables con resultados más satisfactorios en experiencias anteriores. La segunda fase, como consecuencia de la primera, ha consistido en reducir esa lista recopilatoria para evitar multicolinealidad entre variables aplicando tests de autocorrelación entre ellas, lo que ha permitido seleccionar aquellas que están menos correlacionadas entre sí.

### **4.2.1 Revisión bibliográfica sobre los factores asociados a la distribución del *Pinus pinea* y el *Quercus suber***

Una parte de la bibliografía disponible sobre creación de modelos de distribución tanto para *Pinus pinea* como para *Quercus suber* está bastante enfocada hacia el uso de exclusivo de variables climáticas (López-Tirado & Hidalgo, 2016a; Akyol & Örucü, 2019; Akyol et al. 2020; Laaribya et al., 2021), aunque ciertamente la mayoría de los autores apunten hacia la idoneidad de combinarlas con variables de carácter edáfico, litológico o topográfico.

Así pues, en uno u otro caso, las fuentes usadas por los autores son diversas, lo que conduce a una amplia gama de variables utilizados por ellos, y, en consecuencia, también de aquellos que muestran un buen desempeño. Por tanto, para representar de manera más visual las variables destacadas en cada artículo, se ha considerado oportuno de sintetizarlos en tablas para cada una de las especies (*tab. 2 & 3*).

La revisión bibliográfica se ha sistematizado a través de la cuantificación de veces que cada variable se ha usado con éxito en los artículos analizados, visible en el campo “suma”. Además, se ha realizado también un abordaje cualitativo, ya que se han ponderado aquellas variables con mejores rendimientos, lo que se puede apreciar en el campo “n”.

Tabla 2. Variables climáticas para *Pinus pinea* y *Quercus suber*. En verde las variables preliminarmente seleccionadas, de color más oscuro aquellas finalmente seleccionadas.

<i>Pinus pinea</i>	Forma	Escala	Unidad	BioClim	Fuente	Rouget2001	Montero2004	Tirado2016a	Akyol2019	Akyol2020	n	Suma
<b>Variable Climática</b>												
Precipitación	suma	anual	mm	bio12	prec	0	1	2	1	0	1	4
PET	suma	anual	mm		pet	0	0	2	0	1	1	3
Precipitación INV	suma	estacional	mm		prec	2	1	0	0	0	1	3
T med	media	anual	°C	bio1	tas	0	1	2	0	0	1	3
Precipitación VER	suma	estacional	mm		prec	2	1	0	0	0	1	3
Precipitación PRI	suma	estacional	mm		prec	2	1	0	0	0	1	3
Continentalidad anual (Tmx warm-tmin cold)	resta	anual	°C	bio7	tmax&tmin	0	0	2	1	0	1	3
T min	med de las max	mensual	°C	bio6	tmin&tas	0	0	2	1	0	1	3
Índice de Aridez (P/E)	ratio	anual	ratio		prec&pet	0	0	2	0	0	1	2
T min del mes con mayor T°C	med das min diarias	mensual	°C	bio5	tmin&tas	0	0	2	0	0	1	2
Índice Ombrotérmico	ratio	anual	ratio		prec&tas	0	0	2	0	0	1	2
Precipitação Q+humido	suma	estacional	mm	bio16	prec&prec	0	0	0	1	1	0	2
T med del mes con mayor T°C	média	mensual	°C	bio10	tas&tas	0	1	0	0	1	0	2
T med del mês com menor T°C	media	mensual	mm	bio11	tas&tas	0	1	0	1	0	0	2
Precipitación OTO	suma	estacional	mm		prec	0	1	0	0	0	0	1
Prec-PET	resta	anual	mm		precpet	0	0	1	0	0	0	1
Precipitação M+humido	suma	estacional	mm	bio13	prec&prec	0	0	0	1	0	0	1
Precipitação Q+frio	suma	estacional	mm	bio19	prec&tas	0	0	0	1	0	0	1
Precipitação Q+seco	suma	estacional	mm	bio17	prec&prec	0	0	0	0	1	0	1
Temp. Seasonality (st dv *100)	index	anual	índice	bio4	tas	0	0	0	1	0	0	1
N de días Tmin<0°C	suma	anual	días		tmin	0	0	0	0	0	0	0
N de días Tmax>35°C	count	anual	días		tmax	0	0	0	0	0	0	0
Temperatura VER	média	estacional	°C		tas	0	0	0	0	0	0	0
Temperatura INV	média	estacional	°C		tas	0	0	0	0	0	0	0
Precipitation seasonality	index	anual	índice	bio15	prec	0	0	0	0	0	0	0
T max	med das max diárias	anual	°C		max	0	0	0	0	0	0	0
T min	med das min diárias	anual	°C		tmin	0	0	0	0	0	0	0
Meses con 2T>P	count	anual	N meses		prec&tas	0	0	0	0	0	0	0
Temperatura PRI	media	estacional	°C		tas	0	0	0	0	0	0	0
Temperatura OTO	media	estacional	°C		tas	0	0	0	0	0	0	0
N de días com Prec	suma	anual	días		prec	0	0	0	0	0	0	0

<i>Quercus suber</i>	Forma	Escala	Unidad	BioClim	Fuente	Hidalgo2008	Lacamb2010	Paulo2015	Tirado2016	Duque2018	Laarbya2021	n	Suma
<b>Variable Climática</b>													
PET	suma	anual	mm		pet	2	2	2	1	1	0	3	8
Precipitación	suma	anual	mm	bio12	prec	2	2	1	0	1	0	2	6
Precipitación INV	suma	estacional	mm		prec	2	2	0	1	0	0	2	5
Índice de Aridez (P/E)	ratio	anual	ratio		prec&pet	0	0	0	2	2	0	2	4
N de días Tmin<0°C	suma	anual	días		tmin	0	0	2	0	2	0	2	4
T med	media	anual	°C	bio1	tas	1	2	1	0	1	1	1	6
Precipitación VER	suma	estacional	mm		prec	2	1	0	1	0	0	1	4
Precipitación PRI	suma	estacional	mm		prec	1	2	0	1	0	0	1	4
T min del mes con mayor T°C	med das min diarias	mensual	°C	bio5	tmin&tas	1	2	0	0	1	0	1	4
Precipitación OTO	suma	estacional	mm		prec	1	2	0	1	0	0	1	4
Continentalidad anual (Tmx warm-tmin cold)	resta	anual	°C	bio7	tmax&tmin	0	0	0	2	0	1	1	3
Índice Ombrotérmico	ratio	anual	ratio		prec&tas	0	0	1	2	0	0	1	3
Precipitação Q+humido	suma	estacional	mm	bio16	prec&prec	0	0	0	0	0	2	1	2
N de días Tmax>35°C	count	anual	días		tmax	0	0	0	0	2	0	1	2
Temperatura VER	média	estacional	°C		tas	0	0	0	2	0	0	1	2
Temperatura INV	média	estacional	°C		tas	0	0	0	2	0	0	1	2
Precipitation seasonality	index	anual	índice	bio15	prec	0	0	0	0	0	2	1	2
T min	med de las max	mensual	°C	bio6	tmin&tas	1	1	0	0	1	0	0	3



T med del mes com mayor T°C	média	mensal	°C	bio10	tas&tas	1	0	0	0	1	0	0	2
Prec-PET	resta	anual	mm		precpet	0	0	0	1	1	0	0	2
T max	med das max diárias	anual	°C		max	0	0	1	0	1	0	0	2
T min	med das min diárias	anual	°C		tmin	0	0	1	0	1	0	0	2
Meses con 2T>P	count	anual	N meses		prec&tas	0	1	0	0	1	0	0	2
T med del mês com menor T°C	media	mensal	mm	bio11	tas&tas	0	0	0	0	1	0	0	1
Precipitação M+humido	suma	estacional	mm	bio13	prec&prec	0	0	0	0	0	1	0	1
Precipitação Q+frio	suma	estacional	mm	bio19	prec&tas	0	0	0	0	0	1	0	1
Temperatura PRI	media	estacional	°C		tas	0	0	0	1	0	0	0	1
Temperatura OTO	media	estacional	°C		tas	0	0	0	1	0	0	0	1
N de dias com Prec	suma	anual	dias		prec	0	0	1	0	0	0	0	1
Precipitação Q+seco	suma	estacional	mm	bio17	prec&prec	0	0	0	0	0	0	0	0
Temp. Seasonality (st dv *100)	index	anual	indice	bio4	tas	0	0	0	0	0	0	0	0

Por un lado, en algunos artículos se ha optado por usar las 19 variables BIOCLIM para representar sus variables climáticas (Akyol & Örucü, 2019; Akyol et al. 2020; Laaribya et al., 2021). Al tratarse de variables estándar para el estudio del clima aplicado a la ecología, su uso facilita el cálculo de las variables con series climáticas extraídas de fuentes diferentes.

Por otro lado, el resto de los autores ha optado por indicadores tradicionales, que en algún caso pueden llegar a coincidir con alguno de las variables BIOCLIM (tab. 2). Estos están principalmente basados en: medias y sumatorios de una sola variable (precipitación y temperatura), índices compuestos por dos o más variables, la contabilización de días con determinadas características, o basados en cualquier de los anteriores pero medidos de manera estacional.

Además, algunos estudios han optado por dar notable relevancia a variables edáficas y litológicas (tab. 3) aspecto en que existía más divergencia de criterios y diversidad de fuentes. Buena parte de la bibliografía se inclina por usar variables simples como pH o la textura, por falta de datos (Hidalgo et al., 2008; Jovellar-Lacambra et al., 2010) o por convicción (López-Tirado & Hidalgo, 2016b; Montero ed. et al., 2004), este último sin constituir un ejercicio de modelación.

Tabla 3. Variables edáficas y orográficas para *Pinus pinea* y *Quercus suber*. En verde las variables preliminarmente seleccionadas, de color más oscuro aquellas finalmente seleccionadas.

	Unidad	BioClim	Fuente	Rouget2001	Montero2004	Tirado2016a	Akyol2019	Akyol2020	n	SUMA
Espécie				<i>pineae</i>	<i>pineae</i>	<i>pineae</i>	<i>pineae</i>	<i>pineae</i>		
<b>Variable edáfica</b>									0	0
pH	-		Infosolo	0	1	2	0	0	1	3
Profundidad	cm		n	0	1	0	0	0	0	1

Arena / Sand	%	Infosolo	0	1	0	0	0	0	0	1
Arcilla / Clay	%	Infosolo	0	1	0	0	0	0	0	1
Materia orgánica	%	Infosolo	0	1	0	0	0	0	0	1
Limo / Slit	%	Infosolo	0	1	0	0	0	0	0	1
Soil group WRB	wrb2006	n	0	0	0	0	0	0	0	0
Lithology	cual	Atlas Amb.	0	0	0	0	0	0	0	0
N content	%	n	0	0	0	0	0	0	0	0
PBS	%	n	0	0	0	0	0	0	0	0
CEC	meq/100g	Infosolo	0	0	0	0	0	0	0	1

<b>Variable orográfica</b>									0	0
Altitud	n	n	MDE	2	1	0	0	0	1	3
Radiación anual	jls/m2	n	MDE	0	0	2	0	0	1	2
Orientación	n	n	MDE	0	1	0	0	0	0	1
Pendiente	n	n	MDE	0	1	0	0	0	0	1

	Unidad	Fuente	Hidalgo2008	Lacamb2010	Paulo2015	Tirado2016	Duque2018	Laarbya2021	n	SUMA
<i>Quercus suber</i>										
<b>Variable edáfica</b>									0	0
pH	-	Infosolo	0	2	0	2	2	0	3	6
Profundidad	cm	n	0	0	2	0	2	0	2	4
Soil group WRB	wrb2006	n	0	0	2	0	2	0	2	4
Lithology	cual	Atlas Amb.	2	0	2	0	0	0	2	4
Arcilla / Clay	%	Infosolo	0	0	0	0	2	0	1	2
N content	%	n	0	0	0	0	2	0	1	2
PBS	%	n	0	0	0	0	2	0	1	2
CEC	meq/100g	Infosolo	0	0	0	0	2	0	1	2
Materia orgánica	%	Infosolo	0	0	0	0	1	0	0	1
Limo / Slit	%	Infosolo	0	0	0	0	1	0	0	1
Arena / Sand	%	Infosolo	0	0	0	0	1	0	0	1

<b>Variable orográfica</b>									0	0
Altitud	n	MDE	2	0	0	0	1	2	2	5
Orientación	n	MDE	1	0	1	0	1	0	0	3
Pendiente	n	MDE	1	0	1	0	1	0	0	3
Radiación anual	jls/m2	MDE	0	0	0	0	1	0	0	1

Sin embargo, otras obras prefieren combinar esas y otras variables cuantitativas — como el contenido de materia orgánica, la capacidad de intercambio catiónico, el contenido de nitrógeno o la profundidad de los suelos— con la clasificación edáfica y litológica cualitativa —clasificación litológica de Silva (1983) y clasificación WRB de la FAO— (Paulo et al., 2015; Duque-Lazo et al., 2018).

Finalmente, algunos autores han decidido incluir variables orográficas en sus ejercicios de modelación. Muy abundantes en el caso de Duque-Lazo et al. (2018), en que se usan variables más allá de la altura como: la radiación anual y horas de

luz solar por estación, la acumulación de flujos, la orientación o la pendiente. Estas dos últimas variables también son usadas en Paulo et al. (2015) y Montero ed. et al. (2004). Rouget et al. (2001) determina la altitud como uno de los factores más influyentes para diferentes especies de *Pinus*.

Tabla 4. Sumario de variables preseleccionadas.

<i>Tipo de variable</i>	<i>BioClim</i>	<i>Unidad</i>	<i>Dataset</i>	<i>Fuente</i>
<b>Climática</b>				
Precipitación	bio12	mm	prec	IBERIA01
Índice de continentalidad	bio07	°C	Tmax & Tmin	IBERIA01
N de días Tmin<0°C	-	días	Tmin	IBERIA01
T max	-	°C	Tmax	IBERIA01
T. Seasonality (st dv *100)	bio04	índice	Tmed	IBERIA01
Precipitation seasonality	bio15	índice	prec	IBERIA01
<b>Edáfica</b>				
pH		-	Infosolo	INFOSOLO
Materia orgánica		%	Infosolo	INFOSOLO
Arena / Sand		%	Infosolo	INFOSOLO
Limo / Slit		%	Infosolo	INFOSOLO
Arcilla / Clay		%	Infosolo	INFOSOLO
CEC		meq/100g	Infosolo	INFOSOLO
<b>Orográfica</b>				
Altitud		m.s.n.m	MDE	COPERINUCS
Pendiente		°	MDE	COPERNICUS

Aunque la bibliografía relacionada con la modelación de cada especie presente un orden diferente sobre las variables más influyentes, sí que se ha conseguido encontrar puntos en común que se han reflejado en la preselección (tab. 4). Además, hay que subrayar que cuesta encontrar un consenso científico alrededor de cuáles son los variables más adecuados para modelar cada especie, en consecuencia, la aproximación a la bibliografía que se ha optado por hacer en esta investigación es de tipo orientativo.

#### 4.2.2 Evaluación de la colinealidad estadística entre variables

Con el objetivo evitar redundancia entre las variables que forman el modelo se han realizado múltiples tests de autocorrelación en R (Wei & Simko, 2021) (anexo 1). Las variables han sido recortadas en base al área de estudio, el sur de Portugal continental, y se han excluido las áreas con usos del suelo considerados como antropizantes —urbano, logístico, agricultura y pastos—, además de embalses, ríos, humedales y marismas.

Primero se ha testado la autocorrelación entre las variables dentro de su propio género: climáticas, edáficas y orográficas. En el momento de realizar esta selección preliminar, se han elegido las variables menos correlacionadas dentro de cada género, usando como criterio discriminatorio las consideradas como más convenientes en la bibliografía (tab. 2 y 3). Esta primera selección consiste en 6 variables climáticas, 6 edáficas y 2 orográficas, valores de las cuales se encuentran resumidos para poder conocerlas mejor (tab. 4).



Figura 8. Sumario de variables preseleccionadas.

Con esta primera selección de variables se ha procedido a la calibración del modelo, viendo la significancia estadística que podía presentar cada uno de ellos. Una vez conocida la significancia de cada variable, se ha testado su correlación con cada uno de ellos (fig. 8) para poder realizar la selección final en base a estos dos elementos anteriores. Los principios fundamentales han sido: i) maximizar el rendimiento de cada variable, ii) minimizar la autocorrelación entre variables y iii) reducir el número de ellas para no sobredimensionar el modelo.

Tabla 5. Cálculo de la Variance Inflation Factor (VIF) de cada variable seleccionada.

	Precipitación	dTmi<0°C	Continentalidad	Temp. Máx.	Arena	pH
GLM pinea	2	2,33	6,51	7,86	2,20	2,24
GLM suber	3,20	2,38	2,44	4,45	1,80	2,05
GLM mixto	2,51	2,27	4,52	5,87	1,99	2,13

De los criterios anteriores surge la selección definitiva, compuesta por 4 variables climáticas y 2 variables edáficas (*fig. 9 & 10*); quedando fuera 2 variables climáticas —BIO04 y BIO15—, 4 edáficas —CEC, materia orgánica y las texturas de limo y arcilla (%)— y las 2 variables orográficas que había en la selección preliminar —altitud (m) y pendiente (°)—. Por un lado, de las variables edáficas, dos de textura han quedado excluidas por ser correlativas al porcentaje de arena, una variable más significativa para estas especies. Por otro lado, las variables orográficas también se han excluido por ser correlativas con la temperatura, ya que, como se ha comentado en el punto 4.1.2a, esta última está interpolada en base a la altura (Herrera et al., 2019, p. 1951).

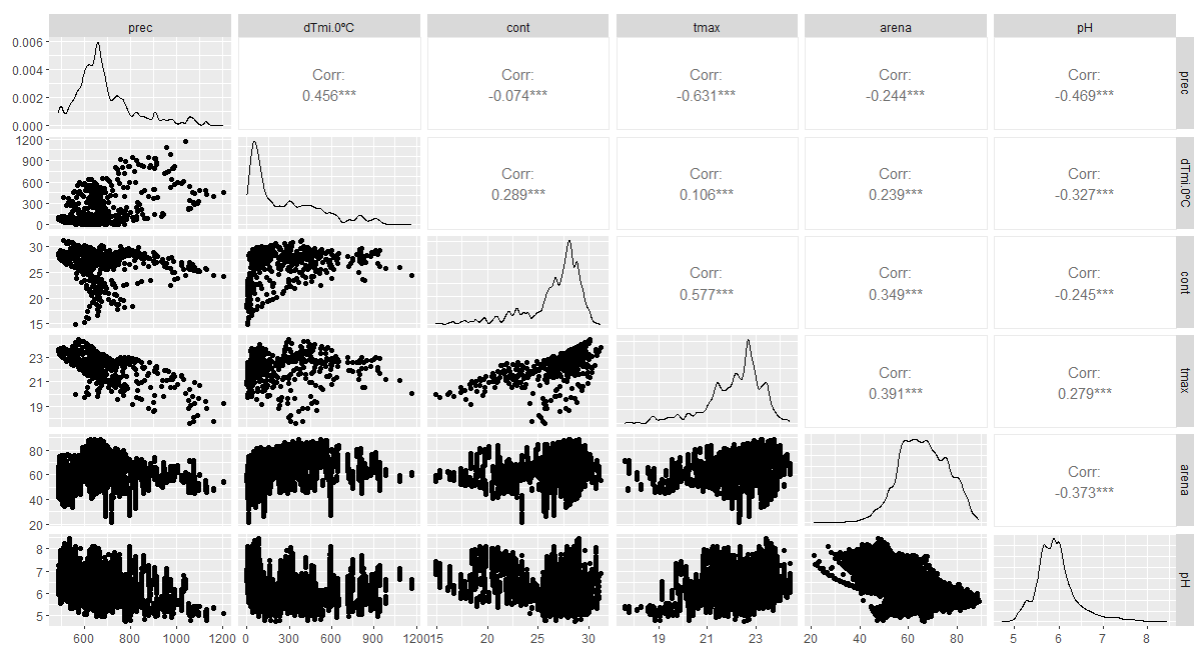


Figura 9. Matriz de correlación entre las variables definitivamente seleccionadas.

Por un lado, las variables climáticas seleccionadas son: i) el número de días de escarcha [Días con T.min. < 0°C], ii) la temperatura máxima media (°C), iii) la continentalidad (BIO07) [temperatura máxima media (°C) del mes más cálido - temperatura mínima media (°C) del mes más frío], iv) media de las precipitaciones (mm) anuales acumuladas (BIO12). Por otro lado, las variables edáficas seleccionadas definitivamente consisten en: i) uno de los niveles granulométricos de la textura de los suelos, la arena (%); y ii) la acidez de los suelos (pH).

Todos ellos se encuentran por debajo del umbral del |0.65| en el coeficiente de correlación (*fig. 9*), lo que indica que no se dan situaciones problemáticas de colinealidad. Con todo, el cálculo del *Variance Inflation Factor* (VIF) muestra que

Continentalidad y Temperatura máxima media presentan algunas limitaciones, principalmente para el modelo del pino piñonero, en el que ambas variables sobrepasan el poco recomendable umbral del VIF=6 (tab. 5).

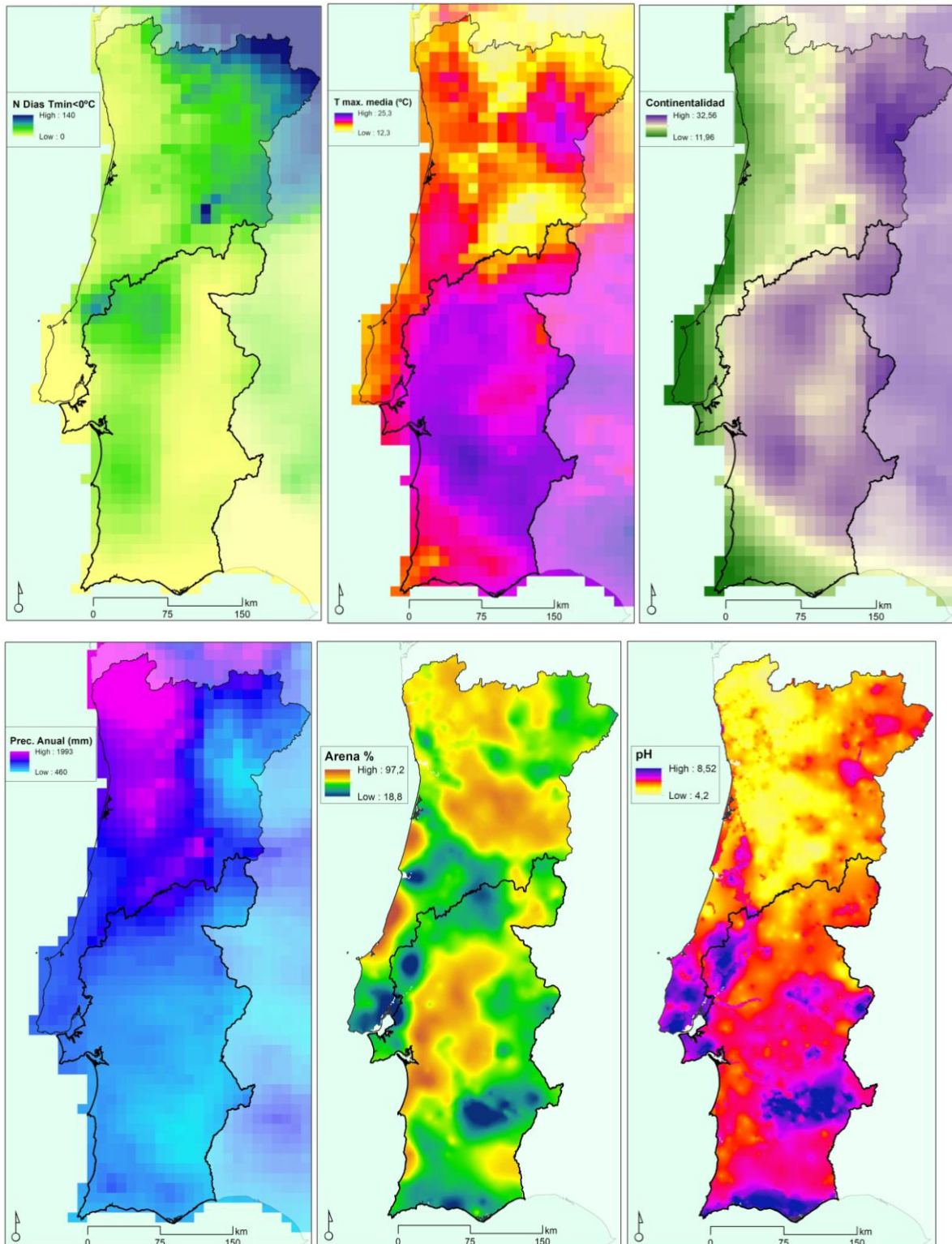


Figura 10. Número de días  $T_{min} < 0^{\circ}\text{C}$ , Temperatura máxima media ( $^{\circ}\text{C}$ ), Continentalidad / BIO07 ( $^{\circ}\text{C}$ ), Precipitación / BIO12 (mm), Arena (%), pH. Fuentes: IBERIA01 & INFOSOLO. Elaboración propia.

### 4.3 Delimitación del área de estudio

La elección de un área de estudio, lo que en Geografía constituye el estudio de caso, es siempre una elección delicada, por riesgo de caer en la arbitrariedad. El caso de estudio debería de contener en sí mismo una serie de características suficientemente diversas para calibrar correctamente el modelo (Lobo et al. 2010). Una vez calibrado y validado, el modelo debe permitir a la investigación extrapolar las conclusiones a otros territorios que se encuentren dentro del envelope edafoclimático en que el modelo fue creado.

#### 4.3.1 Definición del área

Para este caso, el área seleccionada alberga el 95.8%, 96% y 100% de las presencias totales de pino piñonero, alcornoque y masas mixtas, respectivamente, en el país (*fig. 11a, b & c*). En este sentido, según el 6ºIFN (ICNF, 2015) el 20% de las presencias de *Pinus pinea* se encuentra en masas mixtas con *Quercus suber*, mientras que a la inversa se trata tan solo del 5,5%.

El área de estudio no está solamente delimitada por las NUTS III que se encuentran al sur del Tajo, sino que se han excluido de ella aquellas partes con usos del suelo considerados antropizantes —urbano, logístico, agrícola y pastos— (*fig. 11d*). Además, las masas de agua —embalses, ríos, humedales y marismas— se han apartado tanto de la creación del modelo como de su posterior extrapolación (Jiménez-Valverde et al., 2008). El área de estudio está constituida por 54 924 km<sup>2</sup> de los 89 015 km<sup>2</sup> totales de Portugal continental.

Así pues, el área en que el modelo se ha calibrado es aquella delimitada en los mapas menos la que tiene usos considerados como más antropizantes o relacionados con masas de agua, mencionados en el párrafo anterior (*fig. 11d*). Posteriormente el modelo se ha replicado para fuera del área de estudio, pero tan solo para el territorio peninsular de Portugal que se encuentre dentro del envelope edafoclimático, siguiendo las buenas prácticas propuestas por Santini et al. (2021).

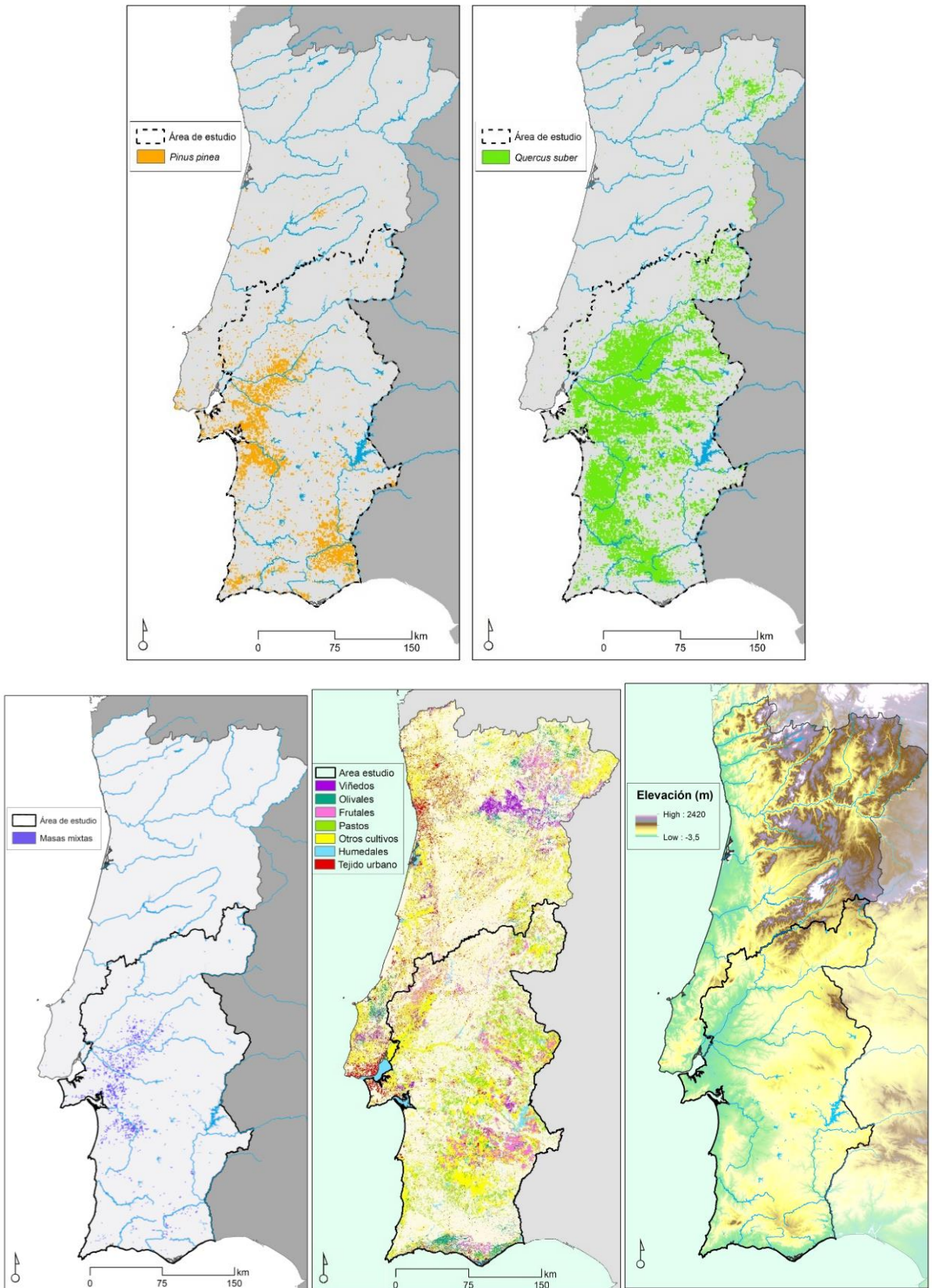


Figura 11. a) Distribución actual de *Pinus pinea*, b) *Quercus suber* y c) masas mixtas (6<sup>o</sup>IFN), d) usos del suelo considerados como antropizantes (COS18) y e) elevación (MDE, Copernicus). Elaboración propia.



### 4.3.2 Caracterización y envelope edafoclimático

Un factor relevante que caracteriza el área es, precisamente, la ocupación humana del territorio, basada en la explotación latifundiaria agrícola, ganadera y / o silvícola (*fig. 11d & 12*). El poblamiento es disperso y existen grandes extensiones dedicadas a una misma actividad, lo que, cuando esta es agrícola, genera una perturbación sobre la distribución basada en el nicho ecológico de especies forestales. Por este último motivo, se ha decidido excluir del área de estudio el uso agrícola del suelo.



*Figura 12. Extensiones agro-silvo-pastoriles alentejanas, finca situada en la Ribeira de Canha. Corknut.*

El área de estudio tiene bastante heterogeneidad en relación con la precipitación (*fig. 13*), a pesar de que se trate de una región de orografía suave con altitudes que no exceden los 250m sobre el nivel del mar (*fig. 11e*). Algo menos heterogéneas son las variables relacionadas con la temperatura, destacando el índice de continentalidad como el que presenta mayor variabilidad (*fig. 13*). Son los factores edáficos los que mayor homogeneidad presentan, principalmente, la variable seleccionada para representar la textura: la arena (*fig. 13*).

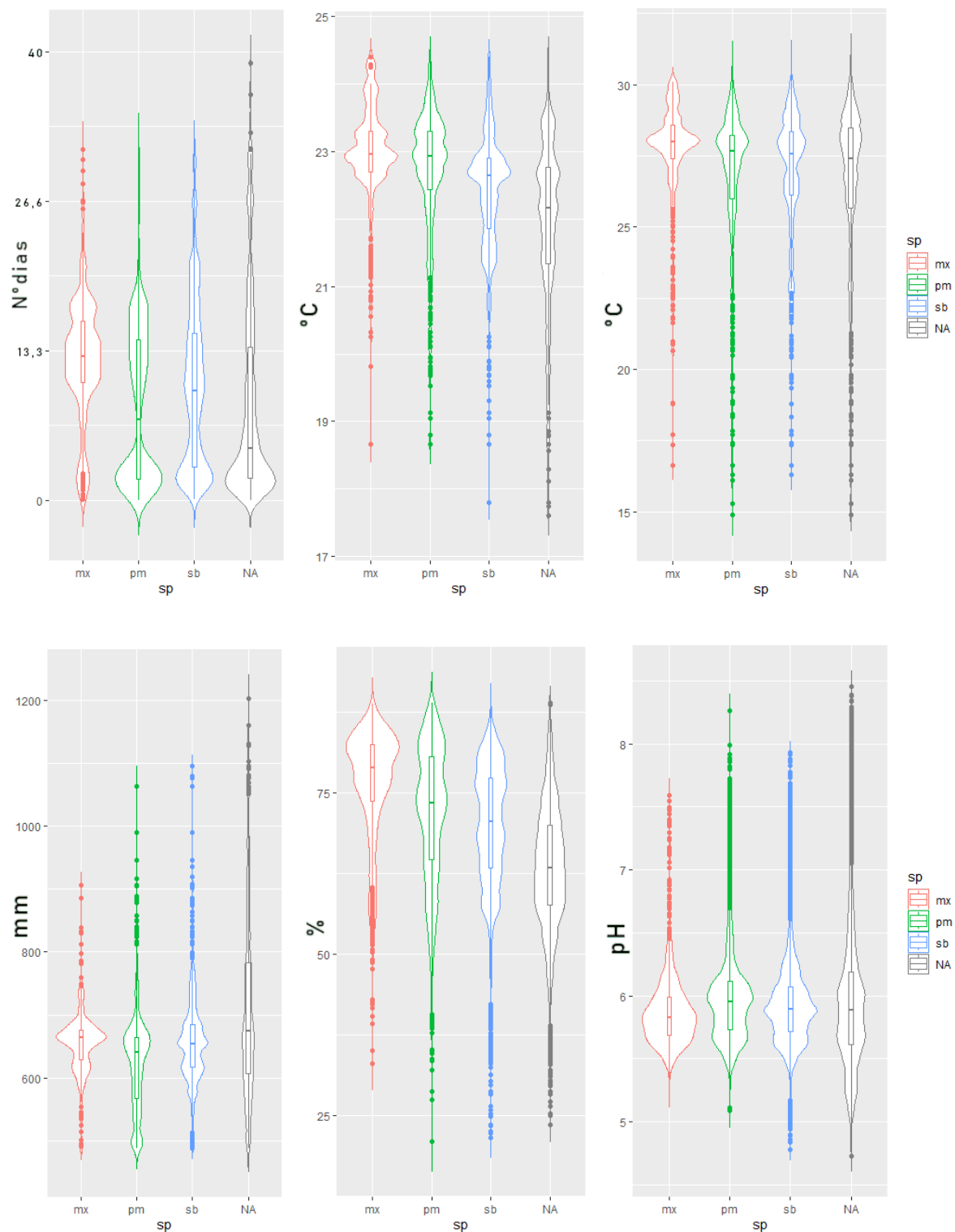


Figura 13. Número de días  $T_{min} < 0^{\circ}\text{C}$ , Temperatura máxima media ( $^{\circ}\text{C}$ ), Continentalidad / BIO07 ( $^{\circ}\text{C}$ ), Precipitación / BIO12 (mm), Arena (%), pH. Fuentes: IBERIA01 & INFOSOLO. Elaboración propia.

El denominado envoltorio edafoclimático constituye el rango de valores de las variables independientes dentro del área de estudio, en la que se encuentran las observaciones mediante las cuales se procede a la calibración del modelo. Así pues,

en la aplicación del modelo solo se tendrán en cuenta áreas que se sitúen dentro de estos rangos de valores: precipitación (mm) [487, 1204], número días con  $T_{min} < 0^{\circ}C$  [0, 39], continentalidad ( $^{\circ}C$ ) [14.9, 31.2], temperatura máxima media ( $^{\circ}C$ ) [17.6, 24.4], arena (%) [21, 89] y pH [4.7, 8.5].

#### **4.4 Modelación estadística con *GLM* y *Random Forests***

Una vez definidos todos los requisitos previos para la modelación, como son el que y el dónde —es decir, decidir qué indicadores conforman la lista de variables independientes y cuál es el área de estudio—, hay que poner el modelo a funcionar. Para ello, en las siguientes líneas, se expondrán paso a paso todas las especificaciones técnicas utilizadas durante los ejercicios de modelación.

##### **4.4.1 Calibración del modelo**

En primer lugar, se realiza una partición de los datos de entrada correspondientes al área de estudio (*anexo 2*) —recordamos: con delimitación territorial y con exclusión de los usos del suelo considerados como más antropizantes—. La partición aleatoria de estos datos se hace con una relación  $\frac{1}{4}$ : el 75% de las observaciones se usa para calibrar el modelo (*train*) y el 25% restante será usado para su validación (*test*) —82 386 y 27 463 registros respectivamente—.

El área de estudio contiene un total de 109 849 registros —correspondientes a la red uniforme de puntos con resolución de 500m del 6º Inventario Florestal Nacional—, de los cuales 8434 (7,7%) corresponden a la presencia de pino piñonero, 31 664 (28,8%) al alcornoque, y tan solo 1644 (1,5%) a parcelas mixtas de las dos especies; 71 395 (65%) corresponden a la ausencia de ambos.

En segundo lugar, con el subset de calibración (75%) se hacen funcionar dos algoritmos —GLM y RF— para generar modelos de distribución para masas de *Pinus pinea*, *Quercus suber* y mixtas dentro del área de estudio (*anexo 2*). Por un lado, el modelo GLM consiste en una regresión binomial; por otro, el modelo RF se ha creado con 500 árboles y un mínimo de 10 observaciones por cada nodo terminal.

En tercer lugar, se piden las importancias de cada variable para los modelos creados con *Random Forests*, lo que el algoritmo GLM ya hace automáticamente. Así pues, a

GLM también se le pide el tipo de correlación que genera cada variable, algo que no se puede saber con *Random Forests* porque se trata de un algoritmo de caja negra.

#### 4.4.2 Validación y aplicación del modelo

A continuación, una vez calibrado el modelo con el subset de calibración (75%), hay que proceder a su validación con el subset de validación (*anexo 3*), el 25% de puntos que no ha intervenido en el proceso de modelación. Para ello se contrasta la predicción del modelo para esa cuarta parte de los puntos aleatoriamente excluidos, con la presencia o ausencia real de la especie en estos puntos. De ahí resultan una serie de métricas, pero en concreto, la que se ha usado de referencia en este trabajo es el área debajo de la curva (*Area Under the Curve*) ROC (*Receiver Operating Characteristics*) —ampliamente utilizada en modelación (Pecchi et al., 2019)—, mediante los paquetes de R ROCR (Sing et al., 2005) y pROC (Robin et al., 2011).

Finalmente, el modelo calibrado y validado está listo para ser extrapolado a otros espacios o tiempos (Guisan & Thuiller, 2005; Elith & Franklin, 2013). En este caso, el modelo se aplica para todo del país (*anexo 4*) —excluyendo las áreas construidas y los humedales—, conceptualizando la información geográfica disponible como perteneciente al presente. La extrapolación incluye el área de estudio, pero el interés se focaliza a aquellas áreas que no han formado parte de la modelación: áreas con usos agrícolas y pastos, por un lado; y por otro, todo el resto del país que se encontraba fuera del área de estudio. En el territorio externo al área de estudio no se extrapola el modelo en aquellos puntos que se encuentren fuera del envelope edafoclimático sobre el que se ha construido el modelo (*fig. 13*).

## 5. Resultados

En este capítulo se presentan los resultados obtenidos de la modelación, en concreto se muestra la contribución de las variables, la capacidad predictiva de cada algoritmo y, de modo ilustrativo, la distribución geográfica potencial determinada por la aplicación de los modelos. La baja capacidad explicativa que han demostrado algunos modelos, principalmente aquellos creados con presencias / ausencias de masas mixtas, conduce a que la representación cartográfica de la distribución potencial tenga un valor relativo.

### 5.1 Capacidad predictiva de los modelos GLM y RF

La lectura de los resultados correspondientes al exponente de los coeficientes — representados en la columna OR, media del intervalo de confianza—, consiste en considerar que por cada variación en una unidad de la variable independiente (°C, mm, %, etc.) la probabilidad de presencia de la variable dependiente crece ( $x > 1$ ) o decrece ( $x < 1$ ).

#### 5.1.1 GLM y RF para *Pinus pinea*

En el modelo GLM para el pino piñonero (*tab. 6*) existen solamente dos variables explicativas con una relación positiva con la variable dependiente: la temperatura máxima media (°C) y el porcentaje de arena; mientras que el resto de las variables explicativas presentan relaciones negativas, siendo la más acentuada la continentalidad, algo coincidente con López-Tirado & Hidalgo (2016a).

Tabla 6. Significancia estadística de cada variable en el modelo GLM *Pinus pinea*.

					OR	>dev.explained
<b>GLM pinea</b>	estimate	std. Error	z value	pr(> z )		0.1710
inintercept	-31,64	825,10	-38,35	< 2e-16	2.317e-14	***
prec	0,00	0,30	-7,57	< 2e-16	0.997	***
dTmi<0°C	0,00	0,10	-5,66	< 2e-16	0.999	***
cont	-0,56	12,67	-44,00	0.12	0.575	--
tmax	1,80	45,12	39,94	< 2e-16	5.769	***
arena	0,08	2,17	38,45	< 2e-16	1.091	***
pH	-0,12	54,16	-2,14	9.79e-05	0.985	***

Ambos modelos muestran el incremento de probabilidad para la presencia *Pinus pinea* con temperaturas máximas medias a partir de los 22,5°C y porcentajes de arena alrededor del 75% (fig. 14), en la línea de lo observado en Montero ed. et al. (2004).

Otras variables se muestran discrepantes según el algoritmo usado, por ejemplo, RF indica una brusca caída de la probabilidad de presencia a partir de los 600mm de precipitación anual, mientras que GLM muestra un suave descenso de esta. Algo parecido pasa con el pH, el modelo GLM muestra un suave descenso de la probabilidad con la disminución de la acidez, mientras que RF muestra un claro aumento de la probabilidad a partir del pH 6,5 hasta suelos más básicos (fig. 14), sobrepasando el umbral indicado en Montero ed. et al. (2004).

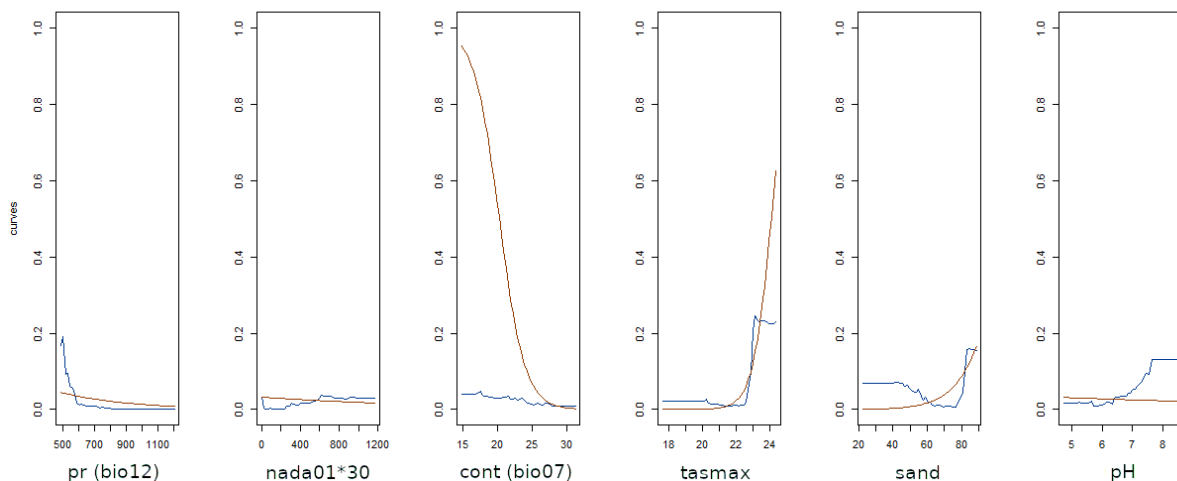


Figura 14. Curvas de respuesta de los modelos GLM (rojo) y RF (azul) para el *Pinus pinea*. Por orden: Precipitación / BIO12 (mm), Número de días  $T_{min} < 0^{\circ}\text{C}$ , Continentalidad / BIO07 ( $^{\circ}\text{C}$ ), Temperatura máxima media ( $^{\circ}\text{C}$ ), Arena (%) y pH.

### 5.1.2 GLM y RF para *Quercus suber*

En el modelo GLM para el alcornoque (tab. 7), según se muestra en la columna OR, 3 variables explicativas guardan una relación positiva con la variable dependiente: el porcentaje de arena y el pH, y en menor medida el número de días con  $T < 0^{\circ}\text{C}$ . Mientras, el resto de las variables muestran una relación negativa, principalmente la temperatura máxima media ( $^{\circ}\text{C}$ ) y la precipitación (mm), y en menor medida el índice de continentalidad.

Tabla 7. Significancia estadística de cada variable en el modelo GLM *Quercus suber*.

GLM suber	estimate	std. Error	z value	pr(> z )	OR	> dev.explained
						0.0991
inintercept	4,86	382,40	12,71	<2e-16	129.280	***
prec	-0,01	0,15	-45,05	<2e-16	0.993	***
dTmi<0°C	0,00	0,05	39,39	0.4247	1.002	--
cont	-0,01	4,79	-1,56	<2e-16	0.994	***
tmax	-0,26	17,47	-15,12	<2e-16	0.773	***
arena	0,06	1,20	45,86	<2e-16	1.056	***
pH	0,11	28,45	3,90	0.0014	1.095	**

Así pues, coinciden con las relaciones negativas mostradas por López-Tirado (2016b) la continentalidad y la precipitación. Sin embargo, la relación con el pH mostrada por el modelo GLM difiere totalmente, de la mostrada por el autor, quien entiende que esta tendría que ser totalmente negativa —como se puede observar en el modelo RF—, es decir, que el alcornoque preferiría los suelos tendencialmente más ácidos.

Las curvas de respuesta (*fig. 15*) muestran algunas relaciones semejantes entre los modelos GLM y RF, como la disminución de la probabilidad de presencia por encima de los 750mm —considerándose como óptimo el intervalo entre los 600 y 700mm—, o el incremento de esta con la existencia de algunas heladas anuales. Sin embargo, se observan notables contradicciones en variables como el pH, la continentalidad o la temperatura máxima media (°C). El porcentaje de arena tiene una lectura ambigua, con la que uno se puede inclinar a pensar que a partir del 60% hay un aumento de la probabilidad.

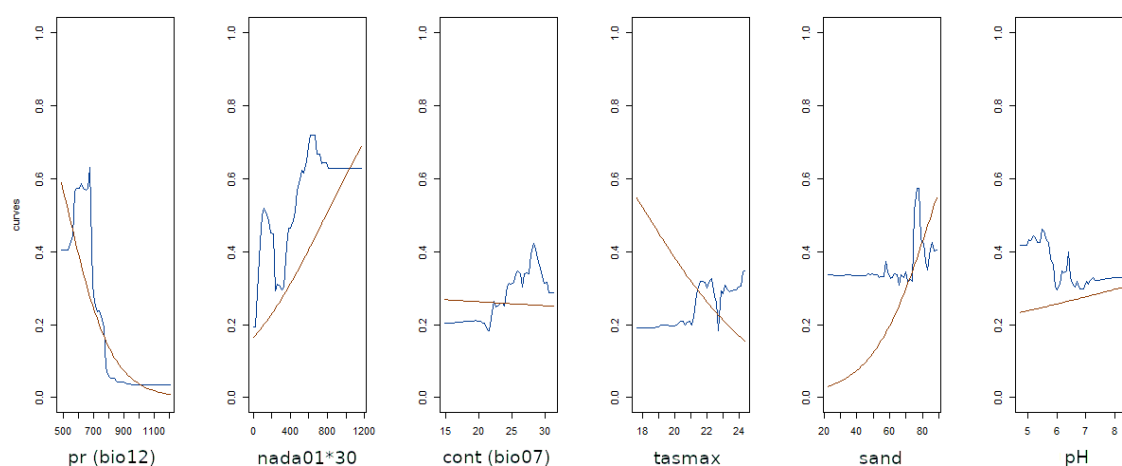


Figura 15. Curvas de respuesta de los modelos GLM (rojo) y RF (azul) para el *Quercus suber*. Por orden: Precipitación / BIO12 (mm), Número de días Tmin <0°C, Continentalidad / BIO07 (°C), Temperatura máxima media (°C), Arena (%) y pH.

### 5.1.3 GLM y RF para mixtos

En el modelo GLM para masas mixtas (*tab. 8*) se pueden apreciar algunas dinámicas parecidas al modelo para el pino piñonero, principalmente en la relevancia de factores relacionados con la temperatura como la continentalidad ( $^{\circ}\text{C}$ ) o la temperatura máxima media ( $^{\circ}\text{C}$ ), con relaciones también negativa y positiva respectivamente. El porcentaje de arena también muestra una relación positiva, algo común con los otros modelos. Para GLM existe una relación positiva con la precipitación (mm), contrario a lo que indican los modelos creados para *Pinus pinea* y *Quercus suber*, algo que RF descarta por completo. Precisamente RF muestra porcentajes de probabilidad muy bajos para todas las variables, las curvas de respuesta solo parecen ser mínimamente sensitivas a lo que la temperatura máxima media ( $^{\circ}\text{C}$ ) y arena se refiere (*fig. 16*).

Tabla 8. Significancia estadística de cada variable en el modelo GLM de masas mixtas.

GLM mixto	estimate	std. Error	z value	pr(> z )	OR	> dev.explained
inintercept	-50,46	2,01	-25,13	<2e-16	1.214e-22	***
prec	0,01	0,00	8,46	<2e-16	1.006	***
dTmi<0°C	0,00	0,00	-0,80	0.4247	0.999	--
cont	-0,46	0,03	-14,75	<2e-16	0.631	***
tmax	1,97	0,10	19,41	<2e-16	7.193	***
arena	0,11	0,00	22,22	<2e-16	1.115	***
pH	0,41	0,13	3,20	0.0014	1.508	**

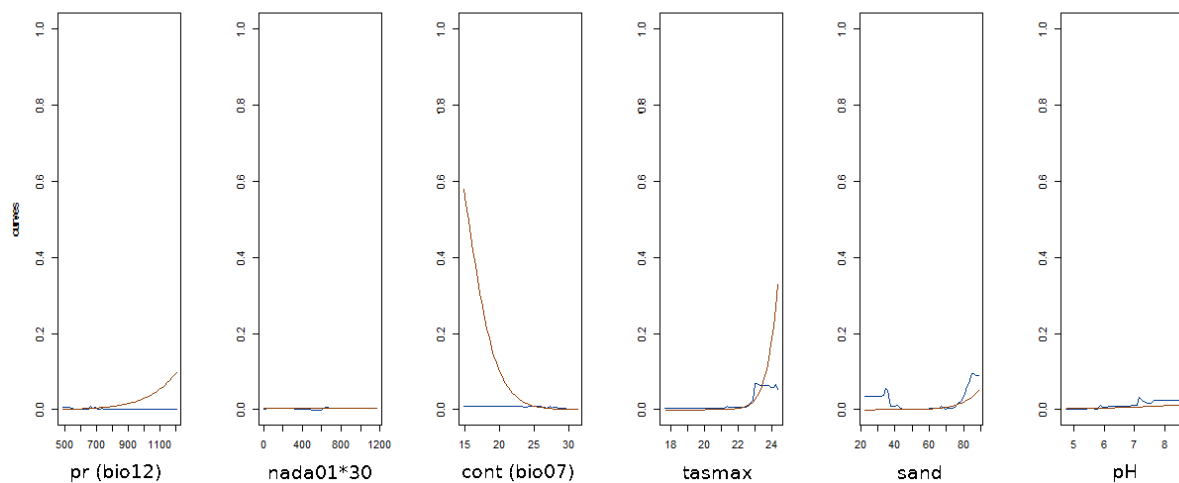


Figura 16. Curvas de respuesta de los modelos GLM (rojo) y RF (azul) para masas mixtas. Por orden: Precipitación / BIO12 (mm), Número de días Tmin <0°C, Continentalidad / BIO07 ( $^{\circ}\text{C}$ ), Temperatura máxima media ( $^{\circ}\text{C}$ ), Arena (%) y pH.



## 5.2 Importancia y contribución de las variables

Para considerar la contribución de cada variable en la modelación de ambas especies se han usado dos métricas, por un lado, *var.imp*, obtenida con el paquete de R *caret* (Kuhn, 2008), compatible con ambos algoritmos; y, por otro lado, *Mean Decrease Gini*, únicamente para RF.

Como indican Shabani et al. (2016) la elección del algoritmo de modelación puede generar correlaciones, importancias y predicciones notablemente diferentes. Sin embargo, Li & Wang (2013) destacan las posibilidades de extraer información complementaria usando diferentes algoritmos, ya que su grado de transparencia es variable. En esta investigación se han usado GLM y *Random Forests*, cuyas métricas se podrá apreciar que difieren considerablemente.

Tabla 9 Importancia y contribución de las variables.

<b><i>Pinus pinea</i></b>	prec	dTmi<0°C	cont	tmax	arena	pH
(GLM) var.imp	7,45	5,66	<b>43,97</b>	<b>39,57</b>	<b>38,85</b>	1,05
(RF) var.imp	39,14	40,49	36,38	41,15	51,39	45,41
(RF) Mean Decr. Gini	541,36	433,29	393,16	653,45	<b>1697,35</b>	<b>1279,13</b>
<b><i>Quercus suber</i></b>						
(GLM) var.imp	<b>44,52</b>	<b>38,71</b>	1,16	14,47	<b>45,99</b>	3,46
(RF) var.imp	76,07	78,87	71,69	50,31	70,7	86,66
(RF) Mean Decr. Gini	<b>3520,71</b>	2762,88	1825,66	1790,01	<b>4488,32</b>	<b>3353,16</b>
<b>Mixtos</b>						
(GLM) var.imp	9,22	1,34	<b>15,64</b>	<b>19,85</b>	<b>22,58</b>	3,35
(RF) var.imp	13,07	16,12	12,32	18,75	16,09	14,37
(RF) Mean Decr. Gini	53,56	50,05	38,72	66,60	<b>311,08</b>	<b>267,73</b>

Para el caso del *Pinus pinea*, aquellas variables consideradas como más importantes para el modelo creado con GLM son la continentalidad, la temperatura máxima media (°C) y el porcentaje de arena; mientras que tan solo esta última variable y el pH, son aquellas con mayor importancia para RF (tab. 9).

Con relación a los modelos creados para *Quercus suber*, las variables que más contribuyen en la creación del modelo GLM son el porcentaje da arena, la precipitación y el número de días con helada, este último destacado en Duque-Lazo et al. (2018). Estas dos primeras variables también se consideran importantes en la creación del modelo RF, además de también el pH (tab. 9). Con el uso de ambos

algoritmos se da una total divergencia con la selección que hacen Akyol & Örucü (2019) y Laaribya et al. (2021) de las variables con más peso, en su caso tan solo climáticas.

Para el caso de las masas mixtas —como ya pasa en las curvas de respuesta (fig. 16)—, tanto GLM como RF muestran un resultado muy parecido al que presentan los modelos GLM y RF para *Pinus pinea*, muy relacionadas con la temperatura y los suelos (tab. 9). Se trata de similitudes que no son fruto de la casualidad, ya que la distribución de masas mixtas se asemeja a la distribución del pino piñonero, por lo tanto es lógico que dada la superposición de observaciones comentada en el apartado 4.3.1 conduzca a la generación de resultados parecidos.

### 5.3 Validación del modelo

El rendimiento de los modelos construidos con GLM ha sido muy diferente según la especie que se tratase, ROC AUC = 0,81, 0,71 y 0,85 para el pino piñonero, el alcornoque y mixtos respectivamente. Mientras que los construidos con *Random Forests* presentan rendimientos similares en los diferentes casos: ROC AUC = 0,88 para pino piñonero y alcornoque y 0,86 para mixtos (fig. 17). Cuanto más se acerque un modelo al valor de 1, más fuerte será su capacidad predictiva.

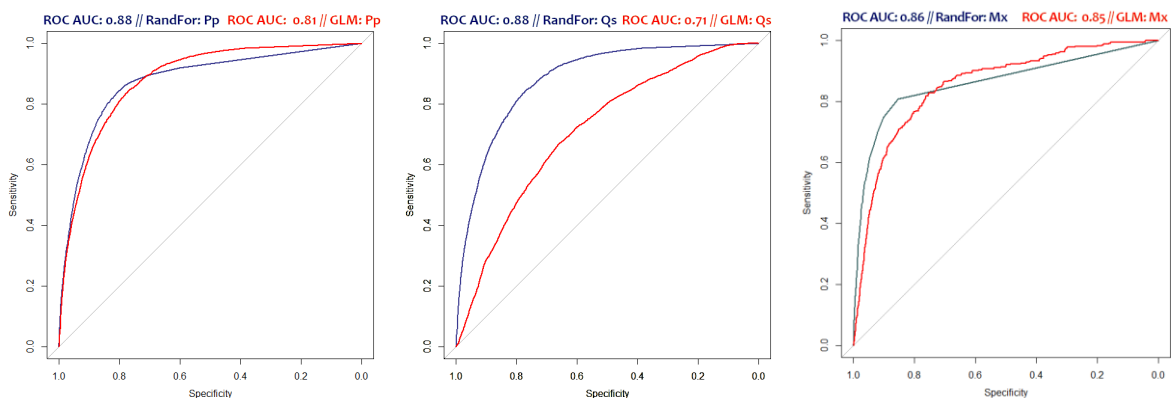


Figura 17. Curvas ROC para *Pinus pinea* (a), *Quercus suber* (b) y mixtos (c).

## 5.4 Ilustración cartográfica de la aplicación de los modelos

Dado que los modelos creados con GLM y RF presentan resultados positivos en la etapa de validación (*fig. 17a*) para la distribución del pino piñonero, se ha decidido extrapolar ambos (*fig. 15*), evitando así, tener que discriminar entre cualquiera de los dos. Aunque la validación presenta resultados peores para el modelo creado con GLM para la distribución del alcornoque (*fig. 17b*), por coherencia, y para poder establecer paralelismos, se ha decidido extrapolarlo igualmente junto con el modelo creado con *Random Forests* (*fig. 18*). Distintamente, los modelos creados con masas mixtas, a pesar de presentar métricas consideradas como exitosas, una vez aplicados se muestran incapaces de alcanzar ni siquiera propósitos ilustrativos.

Por un lado, la aplicación de los modelos para la distribución del *Pinus pinea* a nivel nacional refleja que la actual no difiere mucho de la que sería la potencial (*fig. 18*). Con la aplicación del modelo *GLM* puede observarse que las posibles áreas de expansión están recluidas en la propia área de estudio: Costa Vicentina y sotavento algarvío. Mientras que la aplicación del modelo *RF*, siendo muy conservador dentro del área de estudio, muestra algunas posibilidades fuera de esta, determinadas esencialmente por los suelos arenosos en: el Pinhal de Mira, el Pinhal de Leiria, la costa este, y entre los distritos de Viseu y Guarda, alrededor de la villa de Trancoso y los cursos altos del Mondego y Vouga.

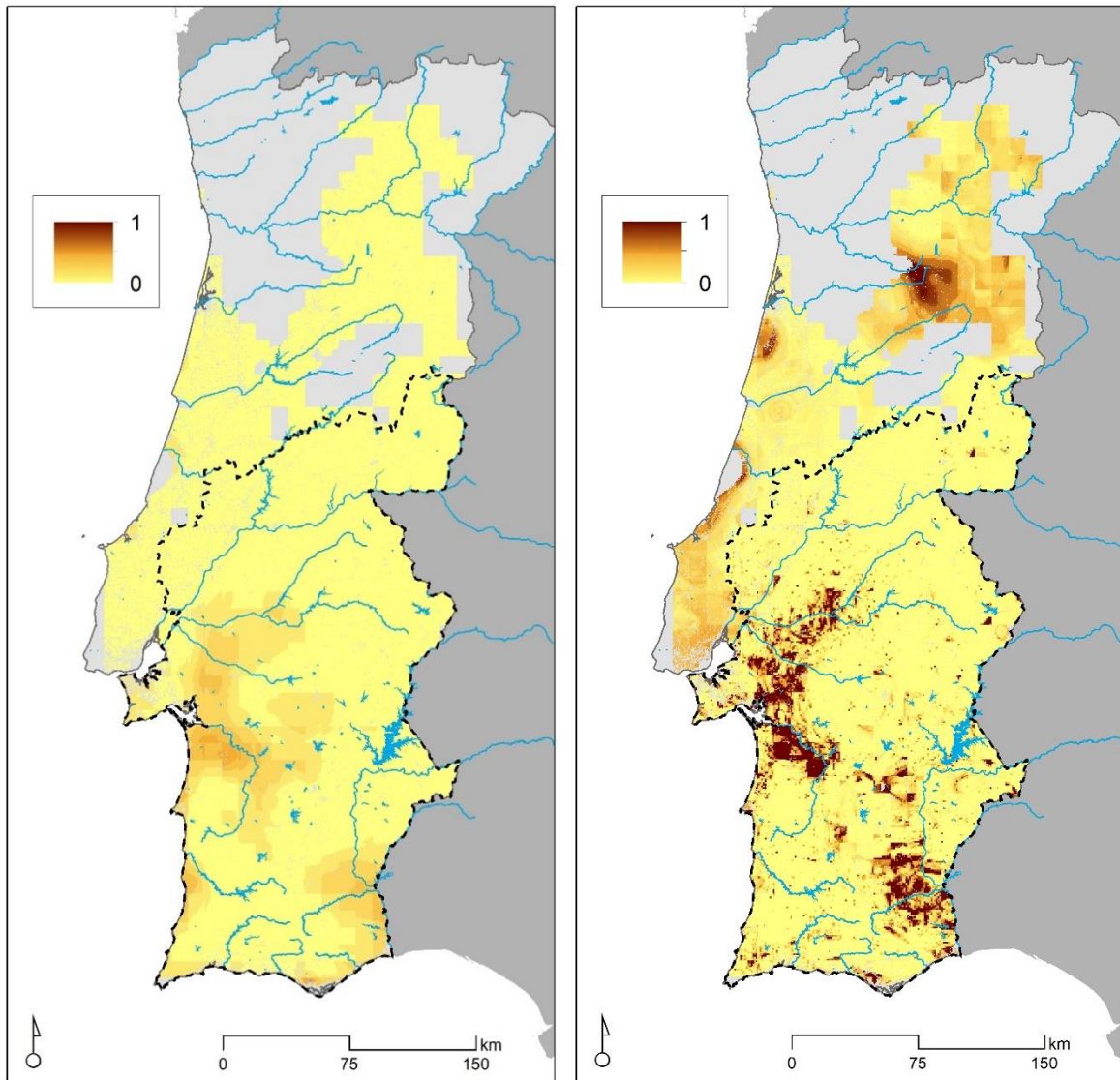


Figura 18. Aplicación de los modelos GLM y RF para el *Pinus pinea* en Portugal continental, probabilidad de presencia entre 0 y 1.

Por otro lado, la aplicación de los modelos para la distribución del *Quercus suber* puede considerarse más diversa a lo que posibles áreas de expansión se refiere (fig. 19). El modelo creado con GLM muestra una adecuación casi generalizada del Alentejo litoral, y puede verse como, además, prevé de manera poco precisa probabilidades bastante altas de presencia en la totalidad de la región. Asimismo, el modelo RF determina de manera más conservadora la adecuación de buena parte del Alentejo, exceptuando la cuenca del Río Guadiana y el estuario del Río Sado — este último considerado como muy adecuado por GLM—. Sin embargo, la coincidencia más fuerte entre modelos se da precisamente fuera del área de estudio, en determinar las regiones de la Beira Alta, Alto Douro y Trás-os-Montes como áreas de potencial expansión.

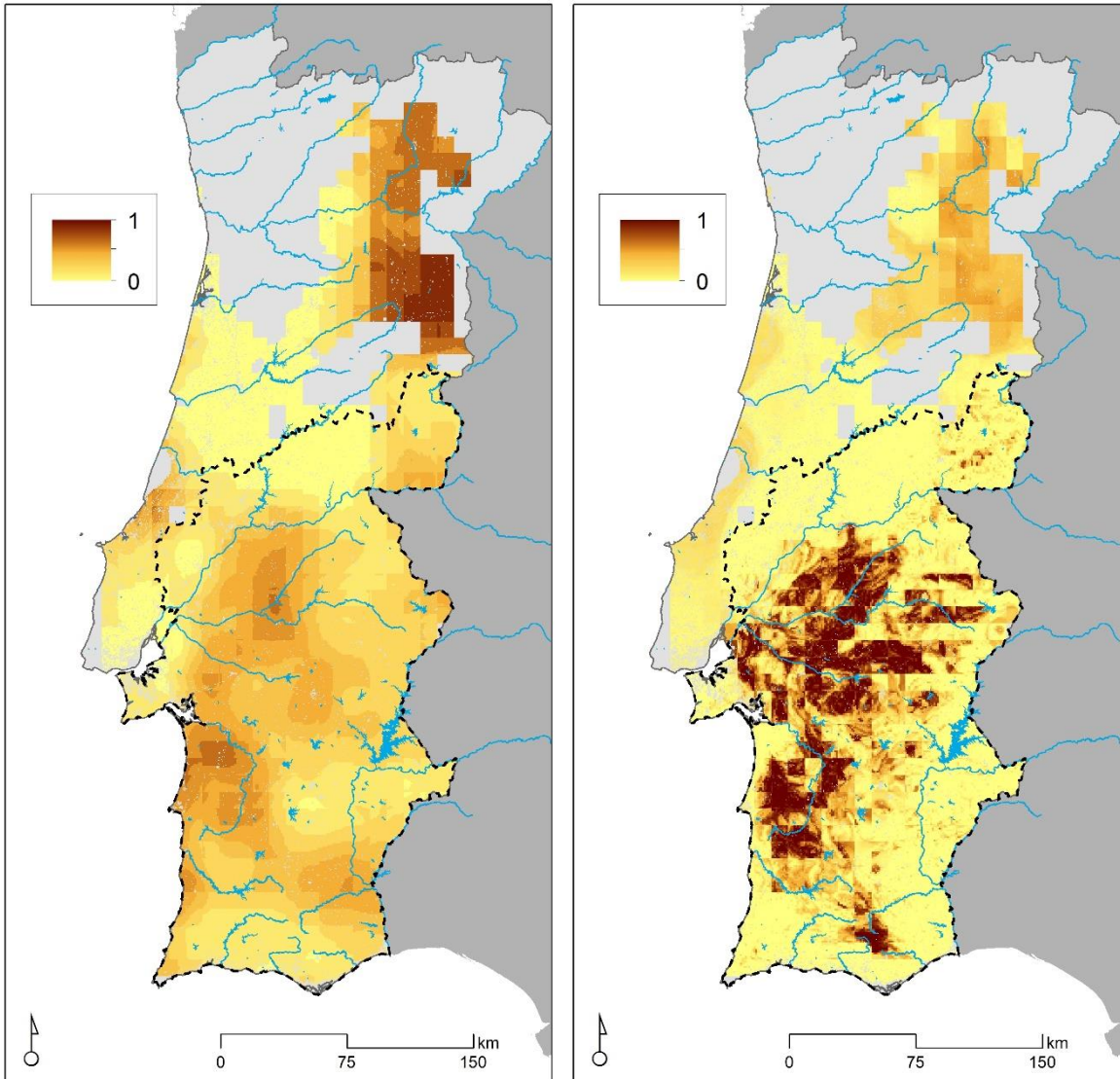


Figura 19. Aplicación de los modelos GLM y RF para el *Quercus suber* en Portugal continental, probabilidad de presencia entre 0 y 1.

Finalmente, la aplicación de los modelos para las masas mixtas constituye un fracaso, ya que ambos algoritmos muestran probabilidades muy bajas, incluso en áreas donde este tipo de bosques ya se encuentran presentes, principalmente en el valle del Sado (fig. 20). En la extrapolación, la única posible área de expansión sería el Pinhal de Mira, al sur de las marismas de Aveiro. En este sentido, la aplicación de los modelos para masas mixtas de pino piñonero y alcornoque sigue la tónica mostrada en las etapas de calibración y validación, la de tener comportamientos muy parecidos a los modelos creados para *Pinus pinea*, cuya causa está explicitada en el último párrafo punto 5.2.

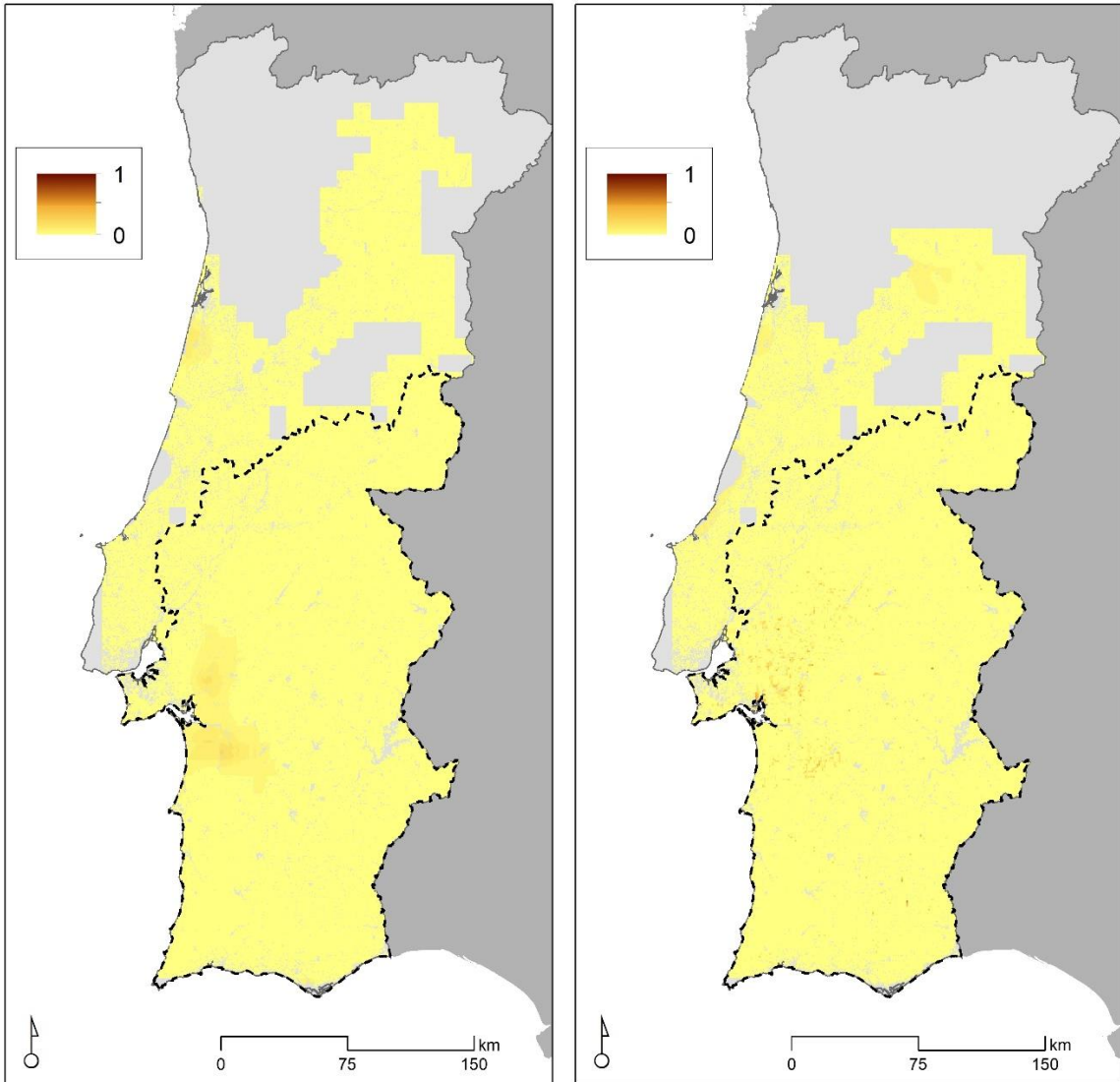


Figura 20. Aplicación de los modelos GLM y RF para masas mixtas de *Pinus pinea* y *Quercus suber* en Portugal continental, probabilidad de presencia entre 0 y 1.

## 6. Discusión

### 6.1 Comprobación de la información de entrada

Como ya se ha referido en el apartado 4.1 sobre las fuentes de datos, la información de entrada consiste en una caracterización de las condiciones edafoclimáticas actuales, en i) el área de estudio para construir el modelo, y ii) el área de aplicación para la replicación de este. La selección de las variables definitivas ha consistido en determinar cuáles de las variables referenciadas en la bibliografía sintetizadas en una preselección tienen un mayor potencial explicativo y generan menor redundancia sobre el modelo. Es decir, realizar una selección de variables que ya se hayan usado en otros experimentos parecidos, que presenten una buena adaptabilidad a las áreas de estudio y aplicación, siguiendo siempre el principio de parsimonia —generar modelos lo más simples posible—.

Por un lado, las variables explicativas, representadas por las 6 variables edafoclimáticas, están extraídas de dos repositorios de datos interpolados a resoluciones diferentes —IBERIA 01 e INFOSOLO, ya referidos también en el apartado 4.1.2—, lo que lleva asociado diferentes grados de incertidumbre y la posibilidad de que este se herede en el modelo. Por otro lado, la variable dependiente, concretizada en la presencia o ausencia de las especies forestales estudiadas, está indicada por la red de “fotopuntos” del 6º IFN, que como está explicado en el punto 4.1.1, parte de un abordaje mixto por parte del ICNF: trabajo de campo en parcelas y fotointerpretación de imágenes satélite.

Así pues, en relación con las variables explicativas existen pocas posibilidades de comprobación de la calidad de la información, y nos basamos en la reputación de las fuentes y sus procesos internos de validación (Ramos et al., 2017; Cornes et al., 2018) para confiar en su precisión a escala nacional. Además, se trata de elementos relativamente estables a lo largo del tiempo como lo son el clima —precipitación & temperatura— y los suelos —propiedades fisicoquímicas—. Sin embargo, sobre la variable dependiente impactan una serie de factores relacionados con la composición de los bosques, principalmente relacionados con la gestión silvícola pero también determinados por las relaciones interespecíficas, que hacen que sea un elemento mucho más dinámico a nivel cronológico. De este modo, es viable y además deseable

realizar comprobaciones en relación con el 6ºIFN sobre de la precisión de las presencias de *Pinus pinea* y *Quercus suber*, aprovechando el trabajo de campo desarrollado en el ámbito del Proyecto Corknut, y anteriormente en Fertipinea y Nutrisuber.

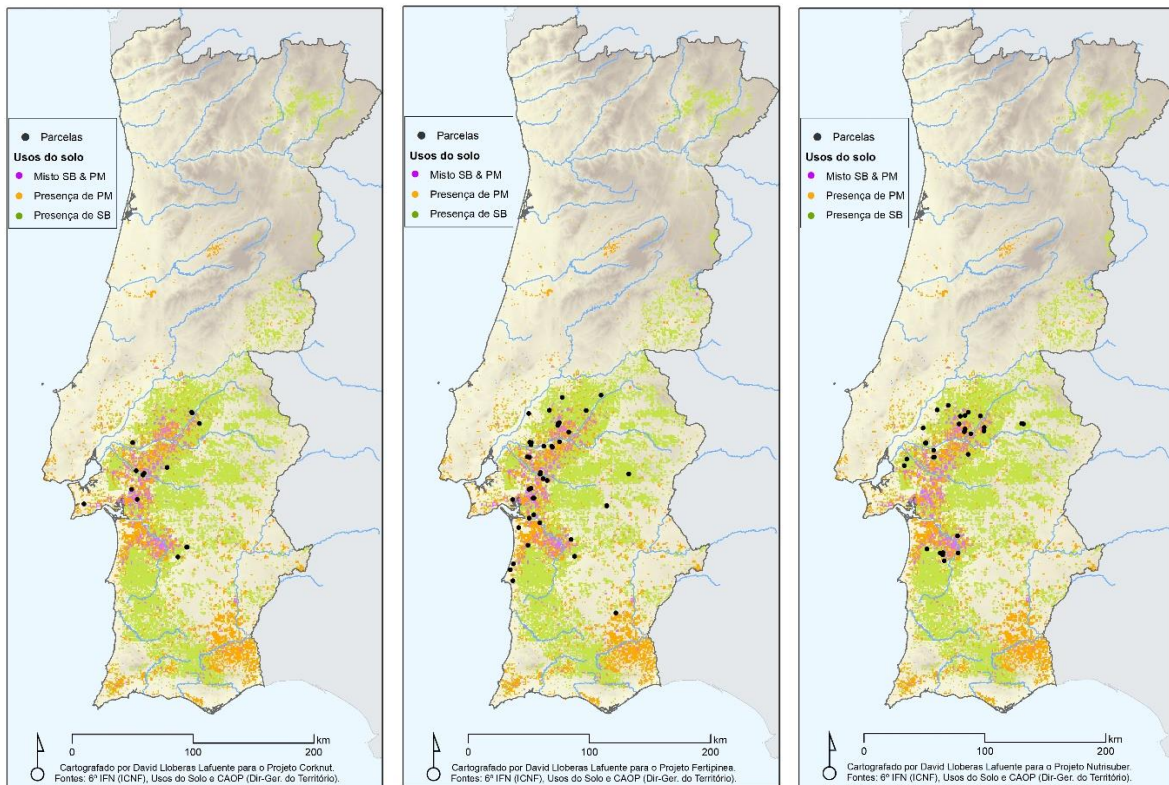


Figura 21. Localización de las parcelas de monitorización permanente de los proyectos Corknut (mixtos), Fertipinea (pino piñonero) y Nutrisuber (alcornoque).

El trabajo de campo realizado en las parcelas de monitorización permanente del Proyecto Corknut verifica la presencia de bosques mixtos en dichos puntos, mientras que el realizado en la red de parcelas de Fertipinea y Nutribusber sirve para confirmar la presencia de pino piñonero y alcornoque respectivamente (fig. 21). Detallado en el apartado sobre las fuentes 4.1.1, el 6ºIFN desarrolla una clasificación híbrida en la que se señala la especie que constituye la ocupación primaria y secundaria en cada “fotopunto”. Así pues, aquellas áreas clasificadas como mixtas lo son porque el binomio ocupación primaria - secundaria está compuesto por *Pinus pinea* y *Quercus suber*, o viceversa.

Las presencias de ambas especies se reflejan correctamente en el 6ºIFN cuanto a ocupación monoespecífica se refiere, sin embargo, se han detectado algunas ausencias metodológicas, es decir erróneas, en relación con áreas con mixtitud de



especies. Muchas de estas parcelas se encuentran en el límite de las áreas de concentración de presencias de bosques mixtos de este tipo, lo que podría explicar la inexactitud de algunas de las informaciones.

## **6.2 Limitaciones de los modelos**

A pesar de que las métricas de validación sean favorables en la mayoría de los modelos desarrollados, los resultados indican ciertas limitaciones que estos presentan. Ciertamente es posible que una parte de las limitaciones se deba a la a algunas inexactitudes de las informaciones de entrada —variables independientes y dependiente—, pues, aunque se haya comprobado su calidad, en modelación un pequeño error puede acabar multiplicándose cuando se usan esos mismos datos para crear nuevos modelos. No obstante, lo más probable es que el principal factor limitante para la calidad de los modelos sea la propia realidad, que cada vez más dificulta desarrollar modelos desde un enfoque exclusivamente ambiental.

Precisamente, fenómenos como la propia gestión forestal realizada y comandada por diferentes agentes territoriales, impacta en la composición de las especies arbóreas que forman el bosque.

Así pues, en este último caso es la decisión de un agente territorial la que determina la presencia o ausencia de una especie forestal. Evidentemente, esa decisión solo puede tener lugar en el caso de que determinada área se encuentre dentro del nicho ecológico. Por ejemplo, en una finca situada dentro de la zona ecológica óptima para la presencia de 3 especies forestales, que bajo condiciones naturales coexistirían unas con otras o por el contrario la mejor adaptada se impondría al resto, es el propietario forestal en tomar la decisión de cuál de ellas va a potenciar. Los criterios para esta selección responden a diversas dinámicas territoriales, que actualmente estando todas ellas relacionadas con el cambio global, en este caso, acaban por ser más de naturaleza económica que ecológica.

Por consiguiente, tratar la distribución actual como referencia para determinar la potencial —siendo que la primera es en parte resultante de procesos sociales— constituye una importante fuente de imprecisión en este tipo de modelos probabilísticos. Si bien es cierto que las presencias acostumbran a responder a la existencia de condicionantes ambientales favorables —con excepciones muy

costosas para el gestor como la riega, la fertilización o la fertirrigación—, la ausencia puede darse por diferentes motivos, de raíz ambiental o social anteriormente ejemplificados.

En este sentido, Lobo et al. (2010) subraya el impacto de las ausencias sobre modelos de distribución de especies, clasificándolas en 3 categorías: ambientales, contingentes y metodológicas. Las ambientales consisten en aquellas ausencias que se encuentran fuera del nicho ecológico de la especie, las contingentes se encuentran dentro del nicho ecológico a pesar de que alguna circunstancia antrópica o ambiental haya evitado la presencia de la especie, y finalmente, las metodológicas, son las observaciones incorrectamente atribuidas a una ausencia. Así pues, con la comprobación de la información de entrada realizada y las métricas de validación resultantes de los modelos, se puede afirmar que de los dos tipos de ausencias que pueden producir desvíos en la modelación —las que lo son a pesar de encontrarse dentro del nicho ecológico— una parte importante de ellas son contingentes, mientras que pueden existir algunas de tipo metodológico.

En este trabajo se ha buscado poder incluir algunos de los motivos socio-económicos que se concretizan en esas ausencias contingentes, pero exclusivamente se han podido sistematizar las razones más evidentes, los usos del suelo explícitamente antrópicos reflejados en la COS18: urbano, logístico, agrícola y ganadero. No obstante, no se han podido sistematizar los motivos más sutiles, igualmente relacionados con procurar el mayor rendimiento económico por hectárea, como es la potenciación de una especie forestal en una determinada área forestal donde otras especies también podrían existir.

Los modelos generados para esta investigación han tenido en cuenta una parte de la perturbación antrópica sobre la distribución —materializada en ausencias contingentes—, siguiendo las indicaciones de Jiménez-Valverde et al. (2008) de excluir de la calibración las ausencias causadas por los usos marcadamente antrópicos: urbano, logístico, agrícola y pastos. Es decir, se han incluido los motivos socio-económicos más explícitos que desencadenan una parte de las ausencias contingentes. No obstante, no se han podido sistematizar los causantes más sutiles, igualmente relacionados con procurar el mayor rendimiento económico por hectárea,

como es la potenciación de una especie forestal en una determinada área forestal donde otras especies menos rentables también podrían existir.

### **6.3 Operacionalización de los bosques mixtos en las áreas de expansión**

Para determinar las posibles áreas de expansión, se ha descartado el modelo creado con las observaciones de masas mixtas y se ha optado por sobreponer las superficies de probabilidad de presencia de los modelos para pino piñonero y alcornoque. Esta superposición, muestra un potencial de crecimiento de las masas mixtas en dos núcleos: uno litoral —Pinhal de Leiria y Pinhal de Mira—; y otro interior, la Beira Alta, cuando son usados los modelos RF. Sin embargo, cuando se usan los modelos creados con GLM, estos coinciden solamente en clasificar como área de expansión la franja litoral del Parque Natural do Sudoeste Alentejano e Costa Vicentina.

Los resultados muestran territorios donde existen condiciones en que *Pinus pinea* y *Quercus suber* pueden llegar a existir con éxito; aun así, y volviendo a la pregunta de investigación, ¿en estos territorios los bosques mixtos pueden ser una alternativa adaptativa del territorio a las dinámicas de la Globalización, permitiendo a los diferentes agentes un aprovechamiento multiescalar de sus servicios del ecosistema?

En base a la distribución actual, se puede deducir que de momento sí que lo son en aquellas áreas que además de encontrarse dentro del nicho ecológico de ambas especies consiguen cumplir algunos requisitos territoriales. Son estos requisitos territoriales los que hacen adquirir importancia a la valoración cualitativa del territorio. ¿Cuáles son los elementos que permiten la operacionalización de masas mixtas de *Pinus pinea* y *Quercus suber* en territorios edafoclimáticamente adecuados? ¿Basta con solamente ser ambientalmente aptos?

La cuestión es compleja, y probablemente concierne a investigaciones posteriores, a pesar de ello, se pueden discernir algunos aspectos para tener en cuenta en estas áreas de expansión potencial. Por un lado, el grado de fragmentación de la propiedad, en caso de ser elevado, plantea dificultades para la creación de economías de escala, causadas por la dispersión de la producción silvícola. Por otro lado, hay que tener en cuenta que las masas forestales en áreas aptas para su presencia no necesariamente

tienen porque ser productivas. Así pues, la logística que toda actividad lleva asociada y la capacidad productiva son factores económicos claves para la operacionalización.

## 7. Conclusiones

El desarrollo de este trabajo ha planteado una serie de cuestiones a nivel conceptual y metodológico, pero también empírico, que merecen ser consideradas para la creación de modelos de distribución de especies. El aspecto central es el elevado grado de antropización de la región mediterránea, cuyo enraizamiento en el tiempo es algo profundo (Lépart & Debussche, 1992), lo que lleva a que la distribución actual —en la que se basa el modelo para determinar la potencial— responda a variables geográficas difíciles de cuantificar.

Este primer elemento está ampliamente discutido, y es algo que está presente de una manera u otra en la modelación de especies mediterráneas (Rouget et al., 2001), lo que no quiere decir que esté resuelto. Así pues, se hace necesario afirmar lo que es una obviedad: las perturbaciones han constituido la principal fuente de incertidumbre durante el proceso de modelación. Proceso que, en este caso, parte de un abordaje principalmente abiótico de la distribución de las especies, es decir, sin tener en cuenta las relaciones interespecíficas —perturbaciones que pueden tener un papel facilitador o limitante (Guisan & Thuiller, 2005; Elith & Franklin, 2013) pero que se estudiarán en fases posteriores del proyecto Corknut—.

En consecuencia, la única manera que se ha encontrado de minimizar la incertidumbre derivada de la perturbación antrópica ha sido la exclusión del área de estudio de aquellos fragmentos del territorio considerados como totalmente bajo influencia humana: medio construido, áreas agrícolas y pastos. Sin embargo, hay que subrayar que existen diferentes grados antropización en áreas fuera de la clasificación anterior, o el origen antrópico de muchas de las relaciones interespecíficas, que, sin la introducción de ciertas especies forestales o el abandono de muchos terrenos agrícolas, no se podrían llegar a dar.

Una vez más, los conceptos de Antropoceno (Crutzen, 2006; Castree, 2015) o Urbanización planetaria (Lefebvre, 1970; Brenner, 2015), ya mencionados en la introducción, permiten entender de manera más precisa las actuales dinámicas territoriales. Aunque se encuentre alejado de los centros de toma de decisiones, el territorio dedicado a actividades económicas del sector primario no está exento de su integración en la globalización, en forma de verticalidades (Santos, 1994, 2000), y sus

posibilidades productivas presentes y pasadas lo han alejado de cualquier tipo de distribución naturalmente generada.

En segundo lugar, hay que destacar las diferencias entre algoritmos, algo visible en: la valoración que se hace de cada variable, los resultados de validación o en las distribuciones potenciales estimadas. *Random Forests* ha demostrado tener en ambas especies mejores métricas de validación que GLM, posiblemente por el hecho de ser considerablemente más conservador en su aplicación dentro del área de estudio. Este conservadurismo puede visualizarse en como la previsión con RF de potenciales presencias dentro del área de estudio parece puntual, en contraposición con el gradiente más continuo que se obtiene con GLM. En este sentido, Jiménez-Valverde et al. (2008) dudan que conservadurismo sea sinónimo de precisión, aduciendo que los algoritmos más complejos pueden estar sobrecalibrados (*overtrained*) para determinar presencias.

En relación con la aplicación de los modelos fuera del área de estudio, en el caso del pino piñonero, RF sugiere áreas de expansión que GLM descarta totalmente —el curso alto del Mondego y Pinhal de Leiria—. Mientras que para el alcornoque ambos algoritmos acaban por mostrar resultados muy parecidos, sugiriendo la idoneidad de la expansión de esta especie en la Beira Alta y Trás-os-Montes. Ciertamente ante la dualidad de resultados obtenidos por la aplicación de dos familias de algoritmos diferentes —los binomiales y los de aprendizaje automatizado— y la aparente dificultad en elegir “el mejor”, hay autores que han optado por la predicción en conjunto o *ensemble forecasting* (Araújo & New, 2007), una interesante vía para continuar la investigación.

En tercer lugar, es importante destacar el poco consenso científico en relación con las variables edafoclimáticas que determinan la distribución de *Pinus pinea* y *Quercus suber*, existen grandes divergencias en la bibliografía relacionada con la modelación de estas especies. La selección utilizada en este trabajo es una propuesta que ha resultado ser aceptable para las circunstancias del territorio analizado, pero que todavía tiene margen de mejora si se pretende aplicarlo a otros espacios geográficos.

Así pues, ambos modelos del pino piñonero permiten vislumbrar la importancia de los suelos arenosos para determinar la probabilidad de su presencia, lo que sucede concretamente a partir de porcentajes altos, superiores al 80% de arena en el suelo.

Al mismo tiempo, *RF* indica la importancia del pH, y sitúa la idoneidad del *Pinus pinea* en áreas tendencialmente más básicas. A nivel climático, *GLM* muestra una clara preferencia de esta especie para áreas con baja continentalidad y altas temperaturas máximas medias.

Por otro lado, los modelos del alcornoque coinciden en apuntar hacia la influencia de la precipitación y la textura del suelo, para la primera variable indican un gran descenso en la probabilidad a partir de los 750mm anuales, mientras que se considera como óptimo el intervalo 600-700mm de lluvia anual. Para la segunda variable mencionada, el porcentaje de arena, la correlación es positiva; al contrario de lo que las curvas de respuesta de *RF* muestran con el pH, que la relación es negativa y que se prefieren suelos ácidos. Asimismo, solo *GLM* destaca la importancia de la cantidad de días con heladas, pareciendo mostrar una correlación positiva.

Finalmente, hay que añadir que a lo largo de la realización de esta tesis han ido ganando peso dos consideraciones que pueden permitir obtener mejores resultados a investigaciones futuras en este ámbito. La primera es remarcar el necesario enfoque funcional en la regionalización de cualquier estudio geográfico (Schaefer, 1953), que en este caso corresponde a la Península Ibérica, y que no se ha podido llevar a cabo por falta de interoperabilidad de los datos a nivel supra estatal. Y la segunda consiste en insistir en la importancia del acceso abierto a bases de datos —con las que esta investigación ha podido trabajar— que sirvan para compartir el conocimiento y que permitan a la comunidad científica ponerlo a disposición de la Humanidad.

## 8. Bibliografía

Agri-ciência Consultores (2014). *Estudo sobre a cadeia de valor da fileira Pinha/Pinhão (Pinus pinea L.)*. UNAC - União da Floresta Mediterrânica.

Akyol, A., & Örucü, Ö. K. (2019). Investigation and evaluation of stone pine (*Pinus pinea* L.) current and future potential distribution under climate change in Turkey. *Cerne*, 25(4), 415-423. doi.org/10.1590/01047760201925042643

Akyol, A., Örucü, Ö. K., & Arslan, E. S. (2020). Habitat suitability mapping of stone pine (*Pinus pinea* L.) under the effects of climate change. *Biologia*, 75, 2175-2187. doi.org/10.2478/s11756-020-00594-9

APCOR – Associação Portuguesa da Cortiça (2020). *CORK 2020: Anuário de cortiça APCOR*.

Belo-Pereira, M., Dutra, E., & Viterbo, P. (2011). Evaluation of global precipitation data sets over the Iberian Peninsula. *Journal of Geophysical Research Atmospheres*, 116(20). doi.org/10.1029/2010JD015481

Brenner, N., & Schmid, C. (2015). Towards a new epistemology of the urban? *City*, 19(2-3), 151-182. doi.org/10.1080/13604813.2015.1014712

Caetano, M., Igreja, C., Marcelino, F., & Costa, H. (1995). Estatísticas e dinâmicas territoriais multiescala de Portugal com base na Carta de Uso e Ocupação do Solo (COS).

Castree, N. (2015). The Anthropocene: A primer for geographers. *Geography*, 100, 66-75. doi.org/10.1080/00167487.2015.12093958

Chorley, R., & Hagget, P. (ed.) (1967). *Models in Geography*. Methuen: London.

Corknut. (2020). *Relatório da candidatura à FCT do Projecto Corknut*.

Cornes, R. C., Schrier, G. van der, Besselaar, E. J. M. van den, & Jones, P. D. (2018). An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391-9409. doi.org/10.1029/2017JD028200

Correia, A. C., Galla, A., Nunes, A., & Pereira, J. S. (2018). Ecological interactions between cork oak (*Quercus suber* L.) and stone pine (*Pinus pinea* L.): Results from a pot experiment. *Forests*, 9(9). doi.org/10.3390/f9090534

DGT – Direção-Geral do Território (2019). *Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 2018*.

Duque-Lazo, J., Navarro-Cerrillo, R. M., & Ruíz-Gómez, F. J. (2018). Assessment of the future stability of cork oak (*Quercus suber* L.) afforestation under climate change scenarios in Southwest Spain. *Forest Ecology and Management*, 409, 444-456. doi.org/10.1016/j.foreco.2017.11.042

Ecco, U. (1977). *Come si fa una tesi di laurea*. Bompiani: Milano.

Fukuda, S., Baets, B. D., Waegeman, W., Verwaeren, J., & Mouton, A. M. (2013). Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environmental Modelling and Software*, 47, 1-6. doi.org/10.1016/j.envsoft.2013.04.005

García-Callejas, D., & Araújo, M. B. (2016). Of model and data complexity on predictions from species distributions models. *Ecological Modelling*, 326, 4-12. doi.org/10.1016/j.ecolmodel.2015.06.002



- González, G. M., Plaza, J. A. C., & Navarro, A. R. (2004). *El pino piñonero (Pinus pinea L.) en Andalucía*. Dirección General de Gestión del Medio Ambiente, Consejería de Medio Ambiente. Junta de Andalucía.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.
- Herrera, S., Cardoso, R. M., Soares, P. M., Espírito-Santo, F., Viterbo, P., & Gutiérrez, J. M. (2019). Iberia01: A new gridded dataset of daily precipitation and temperatures over Iberia. *Earth System Science Data*, 11(4), 1947-1956. doi.org/10.5194/essd-11-1947-2019
- Herrera, S., Gutiérrez, J. M., Ancell, R., Pons, M. R., Frías, M. D., & Fernández, J. (2012). Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). *International Journal of Climatology*, 32(1), 74-85. doi.org/10.1002/joc.2256
- Hidalgo, P. J., Marín, J. M., Quijada, J., & Moreira, J. M. (2008). A spatial distribution model of cork oak (*Quercus suber*) in southwestern Spain: A suitable tool for reforestation. *Forest Ecology and Management*, 255(1), 25-34. doi.org/10.1016/j.foreco.2007.07.012
- ICNF – Instituto de Conservação da Natureza e das Florestas (2015). *IFN 6: Relatório Final*.
- Jactel, H., Bauhus, J., Boberg, J., Bonal, D., Castagnérol, B., Gardiner, B., Gonzalez-Olabarria, J., Koricheva, J., Meurisse, N., Brockerhoff, E. (2017). Tree Diversity Drives Forest Stand Resistance to Natural Disturbances. *Current Forestry Reports*, 3(3), 223-243.
- Jactel, H., Moreira, X., & Castagnérol, B. (2020). Tree Diversity and Forest Resistance to Insect Pests: Patterns, Mechanisms and Prospects. *Annual Review of Entomology*. 66(14). doi.org/10.1146/annurev-ento-041720
- Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: The importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6), 885-890. doi.org/10.1111/j.1472-4642.2008.00496.x
- Jovellar-Lacambra, L. C., Andray, A. B., & Francés, F. S. (2010). Influence of the soil water holding capacity on the potential distribution of forest species. A case study: The potential distribution of cork oak (*Quercus suber* L.) in central-western Spain. *European Journal of Forest Research*, 129(1), 111-117. doi.org/10.1007/s10342-008-0251-5
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*.
- Laaribya, S., Alaoui, A., Ayan, S., Benabou, A., Labbaci, A., Ouhaddou, H., & Bijou, M. (2021). Prediction by maximum entropy of potential habitat of the cork oak (*Quercus suber* L.) in Maamora Forest, Morocco. *Forestist*, 71(2), 63-69. doi.org/10.5152/forestist.2021.20059
- Lefebvre, H. (1970). *La révolution urbaine*. Gallimard.
- Liang, J., Crowther, T. W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E. D., McGuire, A. D., Bozzato, F., Pretzsch, H., De-Miguel, S., Paquette, A., Hérault, B., Scherer-Lorenzen, M., Barrett, C. B., Glick, H. B., Hengeveld, G. M., Nabuurs, G. J., Pfautsch, S., Reich, P. B. (2016). Positive biodiversity-productivity relationship predominant in global forests. *Science*, 354(6309). doi.org/10.1126/science.aaf8957
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R news*, 2(3).

- Lobo, J., Jiménez-Valverde, A., Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modeling. *Ecography*, 33(1), 103-114.
- López-Tirado, J., & Hidalgo, P. J. (2016a). Ecological niche modelling of three Mediterranean pine species in the south of Spain: A tool for afforestation/reforestation programs in the twenty-first century. *New Forests*, 47(3), 411-429. doi.org/10.1007/s11056-015-9523-3
- López-Tirado, J., & Hidalgo, P. J. (2016b). Predictive modelling of climax oak trees in southern Spain: Insights in a scenario of global change. *Plant Ecology*, 217(4), 451-463. doi.org/10.1007/s11258-016-0589-6
- Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research*, 227. doi.org/10.1016/j.fishres.2020.105534
- Mateo, R. G., Felicísimo, Á. M., & Muñoz, J. (2011). Species distributions models: A synthetic revision. *Revista Chilena de Historia Natural*, 84, 217-240.
- Morcuende, A. (2021). Diferenciación y fragmentación socioespacial: La contradicción campo-ciudad como teoría y como método. *GeoUSP*, 25(2), 177986. doi.org/10.11606/issn.2179-0892
- Nunes, L., & Lopes, D. (2009). Quantificação de carbono fixado em povoamentos mistos no norte de Portugal. *Millenium - Journal of Education, Technologies, and Health*, 37(14). hdl.handle.net/10400.19/335
- Paulo, J. A., Palma, J. H. N., Gomes, A. A., Faias, S. P., Tomé, J., & Tomé, M. (2015). Predicting site index from climate and soil variables for cork oak (*Quercus suber* L.) stands in Portugal. *New Forests*, 46(2), 293-307. doi.org/10.1007/s11056-014-9462-4
- Pausas, J. G., Bladé, C., Valdecantos, A., Seva, J. P., Fuentes, D., Alloza, J. A., Vilagrosa, A., Bautista, S., Cortina, J., & Vallejo, R. (2004). Pines and oaks in the restoration of Mediterranean landscapes of Spain: New perspectives for an old practice-a review. *Plant Ecology*, 171, 209-220.
- Pereira, E. L., Santos, S. A. P., Arrobas, M., & Patricio, M. S. (2011). La biomasa microbiana y la mineralización de N en plantaciones mixtas de frondosas y especies fijadoras de nitrógeno. *Investigacion Agraria Sistemas y Recursos Forestales*, 20(3), 516-524. doi.org/10.5424/fs/20112003-11458
- Pretzsch, H., Forrester, D. I., & Bauhus, J. (ed) (2017). *Mixed-Species Forests*. Ecology and Management, Springer. doi.org/10.1007/978-3-662-54553-9
- Prévosto, B., Gavinet, J., Monnier, Y., Corbani, A., & Fernandez, C. (2016). Influence of neighbouring woody treatments on Mediterranean oak development in an experimental plantation: Better form but weaker growth. *Forest Ecology and Management*, 362, 89-98. doi.org/10.1016/j.foreco.2015.11.046
- Ramos, T. B., Horta, A., Gonçalves, M. C., Pires, F. P., Duffy, D., & Martins, J. C. (2017). The INFOSOLO database as a first step towards the development of a soil information system in Portugal. *Catena*, 158, 390-412. doi.org/10.1016/j.catena.2017.07.020
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). PROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12. doi.org/10.1186/1471-2105-12-77
- Rouget, M., Richardson, D., Lavorel, S., Vayreda, J., Gracia, C. & Milton, S. (2001). Determinants of distribution of six Pinus species in Catalonia. *Journal of Vegetation Science*, 12, 491-502.

- Santini, L., Benítez-López, A., Maiorano, L., Čengić, M., & Huijbregts, M. A. J. (2021). Assessing the reliability of species distribution projections in climate change research. *Diversity and Distributions*, 27(6), 1035-1050. doi.org/10.1111/ddi.13252
- Santos, M. (1996). *A Natureza do Espaço: Técnica e Tempo, Razão e Emoção* (4ª ed. [2002]). EdUSP: São Paulo. ISBN 978-85-314-0713-0.
- Santos, M. (2000). *Por uma outra globalização: Do pensamento único à consciência universal*. Record: Rio de Janeiro. ISBN 978-85-01-05878-2.
- Schaefer, F. (1953). Excepcionalism in Geography. *Annals of the Association of American Geographers*, 43(3), 226-249.
- Shabani, F., Kumar, L., & Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and Evolution*, 6(16), 5973-5986. doi.org/10.1002/ece3.2332
- Sing, T., Sander, O., Beerewinkel, N., & Lengauer, T. (2005). ROCr: Visualizing classifier performance in R. *Bioinformatics*, 21(20), 3940-3941. doi.org/10.1093/bioinformatics/bti623
- Tomaz, C., Alegria, C., Monteiro, J. M., & Teixeira, M. C. (2013). Land cover change and afforestation of marginal and abandoned agricultural land: A 10-year analysis in a Mediterranean region. *Forest Ecology and Management*, 308, 40-49. doi.org/10.1016/j.foreco.2013.07.044
- Tóth, G., Jones, A., & Montanarella, L. (2013). The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environmental Monitoring and Assessment*, 185(9), 7409-7425. doi.org/10.1007/s10661-013-3109-3
- Tøttrup, C., & Sørensen, M. K. (2014). EU-DEM Statistical Validation.
- UNAC – União da Floresta Mediterrânica (2017). *Boletim do mercado da pinha: Campanha de 2017/2018*.
- UNAC – União da Floresta Mediterrânica (2018). *Boletim do mercado da cortiça: Campanha de 2018*.
- Urbieto, I. R., García, L. V., Zavala, M. A., & Marañón, T. (2011). Mediterranean pine and oak distribution in southern Spain: Is there a mismatch between regeneration and adult distribution? *Journal of Vegetation Science*, 22(1), 18-31. doi.org/10.1111/j.1654-1103.2010.01222.x
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, 44(12), 1731-1742. doi.org/10.1111/ecog.05615

## Anexos

### Anexo 1: Test de autocorrelación

```
#Test de autocorrelación
library("corrplot")
wr = stack(dataset$var1, dataset$var2, dataset$var3, ... n)

correl = rcorr(as.matrix(wr), type="spearman")
wr_pvalue = correl$P
wr_rvalue = correl$r

matrix = cor(wr_rvalue)
corrplot.mixed(matrix, order = 'AOE')
```

### Anexo 2: Calibración del modelo

#### **#CARGAR LOS DATOS DEL ÁREA DE ESTUDIO Y FORMATO**

```
ar_est <- read_csv("inf15_ar_est_sul_plus_plus2_.txt")
View(ar_est)
df1=df[ar_est$cos18_agr_ == 0 & df$urb == 0 & df$aqua == 0,]

df2 = transform(df1, OBJECTID=as.factor(OBJECTID), pinus2=as.factor(pinus2),
suber2=as.factor(suber2), nada01=as.numeric(nada01), pr=as.numeric(pr_),
bio04=as.numeric(bio4a_), tmax=as.numeric(tasmax1),
bio15=as.numeric(bio15d_), bio07=as.numeric(cont_2extr),
cos18_agr_=as.factor(cos18_agr_), aqua=as.factor(aqua), urb=as.factor(urb),
mdt=as.numeric(mdtetrs89t), slope25=as.numeric(slope25),
sand=as.numeric(sand_prj), slit=as.numeric(slit_prj),
clay=as.numeric(clay_prj), ph=as.numeric(ph_prj), oc=as.numeric(oc_prj),
cec=as.numeric(cec_prj))

datafr <- data.frame(df2)

df3=subset(df2, select=c(OBJECTID, pinus2, suber2, mx, pr_, nada01,
cont_2extr, tasmax1, sand_prj, ph_prj))

df3 <- df3 %>%
  rename("pm" = "pinus2",
        "sb" = "suber2",
        "prec" = "pr_",
```

```

"dTmi<0°C" = "nada01",
"cont" = "cont_2extr",
"tmax" = "tasmax1",
"arena" = "sand_prj",
"pH" = "ph_prj")

```

#### **#PARTICIÓN DE LOS PUNTOS**

```

library("caTools")
sample = sample.split(datafr$OBJECTID, SplitRatio = 0.75)

train = subset(datafr, sample == TRUE)
test = subset(datafr, sample == FALSE)
dim(train)
dim(test)

```

#### **#GLM MODEL**

```

pm_mod_glm <- glm(pm ~ prec + dTmi.0°C + cont + tmax + arena
+ pH, family='binomial', data=train)
summary(pm_mod_glm)

sb_mod_glm <- glm(sb ~ prec + dTmi.0°C + cont + tmax + arena
+ pH, family='binomial', data=train)
summary(sb_mod_glm)

mx_mod_glm <- glm(mx ~ prec + dTmi.0°C + cont + tmax + arena
+ pH, family='binomial', data=train)
summary(mx_mod_glm)

```

#### **#EXPONENTES DE LOS COEFICIENTES**

```

exp_pm=exp(cbind(OR = coef(pm_mod_glm), confint(pm_mod_glm)))
exp_sb=exp(cbind(OR = coef(sb_mod_glm), confint(sb_mod_glm)))
exp_mx=exp(cbind(OR = coef(mx_mod_glm), confint(mx_mod_glm)))

```

#### **#PODER EXPLICATORIO**

```

dev.null <- pm_mod_glm$null.deviance
dev.resid <- pm_mod_glm$deviance

dev.null <- sb_mod_glm$null.deviance
dev.resid <- sb_mod_glm$deviance

dev.null <- mx_mod_glm$null.deviance
dev.resid <- mx_mod_glm$deviance

dev.explained <- (dev.null-dev.resid)/dev.null
dev.explained

```

### ***#RANDOM FORESTS MODEL***

```
randfor_pm = randomForest(pm ~ prec + dTmi.0°C + cont + tmax + arena
+ pH, data = train, ntree = 500, nodesize=10, importance =T)
print(randfor_pm)

randfor_sb = randomForest(sb ~ prec + dTmi.0°C + cont + tmax + arena
+ pH, data = train, ntree = 500, nodesize=10, importance =T)
print(randfor_sb)

randfor_mx = randomForest(mx ~ prec + dTmi.0°C + cont + tmax + arena
+ pH, data = train, ntree = 500, nodesize=10, importance =T)
print(randfor_mx)
```

## **Anexo 3: Validación del modelo**

### ***#VALIDACIÓN ROC AUC***

```
library(ROCR)
library(pROC)

pred_glm_pm= as.numeric(predict(pm_mod_glm, test, type="response"))
pred_glm_sb= as.numeric(predict(sb_mod_glm, test, type="response"))

pred_rf_pm= (predict(randfor_pm, test, type="prob"))
pr3d_rf_pm=as.numeric(pred_pm[,2])
pred_rf_sb= (predict(randfor_sb, test, type="prob"))
pr3d_rf_sb=as.numeric(pred_sb[,2])

ROC_log = roc(test$pinus2, pred_glm_pm)
auc(ROC_log)
plot(ROC_log, col = "red", main = "ROC AUC: 0.81 // GLM: Pm")
ROC_log = roc(test$sub2, pred_glm_sb)
auc(ROC_log)
plot(ROC_log, col = "red", main = "ROC AUC: 0.71 // GLM: Sb")

ROC_log = roc(test$pinus2, pr3d_rf_pm)
auc(ROC_log)
plot(ROC_log, col = "red", main = "ROC AUC: 0.88 // RandFor: Pm")
ROC_log = roc(test$sub2, pr3d_rf_sb)
auc(ROC_log)
plot(ROC_log, col = "red", main = "ROC AUC: 0.89 // RandFor: Sb")
```

## **Anexo 4: Transferibilidad del modelo**

### ***#CARGAR LOS DATOS DE PORTUGAL Y FORMATO***

```
pontos_portugal <- read_excel("ifn15_PORTUGAL.xlsx")
View(pontos_portugal)
```

```

df2_pt = transform(df_pt, OBJECTID=as.factor(OBJECTID), pm=as.factor(pm),
sb=as.factor(sb), dTmi.0°C=as.numeric(dTmi.0°C), prec=as.numeric(prec),
tmax=as.numeric(tasmax_), cont=as.numeric(cont), codlito=as.factor(codlito),
codsolo=as.factor(codsolo),
  cos18_agr_=as.factor(cos18_agr_), aqua=as.factor(cos18_aqua),
urb=as.factor(cos18_urb_), mdtetrs89t=as.numeric(mdtetrs89t),
arena=as.numeric(arena), slit_prj=as.numeric(slit_prj),
clay_prj=as.numeric(clay_prj),
  pH=as.numeric(pH), oc_prj=as.numeric(oc_prj), cec_prj=as.numeric(cec_prj))

```

***#DEFINIR EL ENVELOPE EDAFOCLIMÁTICO Y LAS ÁREAS DE EXCLUSIÓN***

```

df4pt=df3_pt[df3_pt$urb == 0 & df3_pt$aqua == 0,]

df4_pt=df4pt[df4pt$prec >= 487.4 & df4pt$prec <= 1203.9 &
df4pt$dTmi.0°C >= 0 & df4pt$dTmi.0°C <= 1170 &
df4pt$cont >= 14.91 & df4pt$cont <= 31.19 &
df4pt$tmax >= 17.61 & df4pt$tmax <= 24.40 &
df4pt$arena >= 21.06 & df4pt$arena <= 88.84 &
df4pt$pH >= 4.727 & df4pt$pH <= 8.457,]

```

***#EXTRAPOLAR EL MODELO PARA LOS PUNTOS DE PORTUGAL CONTINENTAL QUE SE ENCUENTREN DENTRO DEL ENVELOPE CLIMÁTICO***

```

modelo_pt_glm_pm <- predict(pm_mod_glm, df4_pt, type="response")
df4_pt$glm_Pm=modelo_pt_glm_pm

modelo_pt_glm_sb <- predict(sb_mod_glm, df4_pt, type="response")
df4_pt$glm_Sb=modelo_pt_glm_sb

modelo_pt_glm_mx <- predict(mx_mod_glm, df4_pt, type="response")
df4_pt$glm_Mx=modelo_pt_glm_mx

modelo_pt_rf_pm <- predict(randfor_pm, df4_pt, type="prob")
df4_pt$rfr_Pm=modelo_pt_rf_pm

modelo_pt_rf_sb <- predict(randfor_sb, df4_pt, type="prob")
df4_pt$rfr_Sb=modelo_pt_rf_sb

modelo_pt_rf_mx <- predict(randfor_mx, df4_pt, type="prob")
df4_pt$rfr_Mx=modelo_pt_rf_mx

```

***#EXPORTAR PARA CAROGRAFIAR EN EL SIG***

```

setwd("E:/modelo/results_MODEL_NOV")

write.csv(df4_pt, file = "model_RF.csv", row.names = TRUE)

```