

Modelización de distribución de plantas alóctonas invasoras más problemáticas en la península Ibérica

Proyecto SPAINCLIM

Escenarios futuros de índices bioCLIMáticos en ESPAÑA y aplicación al estudio de especies invasoras



Autoría: Salvia García Álvarez, Joe David Zambrano Suárez, Joaquín Solana Gutiérrez, Ignacio García-Amorena Gómez del Moral

2020

Introducción	3
1 Metodología general	4
1.1 Definiciones y estructura de la modelización	4
1.2 Preparación y filtrado de las presencias de cada especie	5
1.3 Obtención de variables ambientales	6
1.4 Selección de las variables ambientales	9
1.5 Ajuste de los modelos de distribución de especies en el presente	10
Generación de pseudo-ausencias	10
Conjuntos de datos para calibración y evaluación	11
Evaluación de los modelos	12
1.5 Proyección a escenarios futuros	14
2 Modelizaciones realizadas	15
2.1 <i>Arundo donax</i>	15
2.1.1 Datos de entrada en el MDE	15
2.1.2 Ajuste del MDE	17
2.1.3 Modelización de la distribución potencial futura:	22
2.2 <i>Baccharis halimifolia</i>	26
2.2.1 Datos de entrada en el MDE	26
2.2.2 Ajuste del MDE	28
2.2.3 Modelización de la distribución potencial futura:	33
2.3 <i>Carpobrotus acinaciformis</i>	35
2.3.1 Datos de entrada en el MDE	35
2.3.2 Ajuste del MDE	37
2.3.3 Modelización de la distribución potencial futura	42
2.4 <i>Opuntia ficus-indica</i>	44
2.4.1 Datos de entrada en el MDE	44
2.4.2 Ajuste de los MDE	47
2.4.3 Modelización de la distribución potencial futura	61
Referencias	62
Referencias software utilizado	66

Introducción

Este documento contribuye al desarrollo del proyecto “Escenarios futuros de índices bioCLIMáticos en ESPAÑA y aplicación al estudio de especies invasoras (SPAINCLIM)”, a cargo de la Fundación para la Investigación del Clima (FIC) y la Universidad Politécnica de Madrid (UPM). Este proyecto se apoya en los trabajos previamente realizados en proyectos anteriores, como “Análisis de la vulnerabilidad de la BIODiversidad FORESTal para la defensa ante el Cambio Climático (BIOFOREST-CC)” y “Búsqueda de Indicadores Óptimos para la Defensa y Evaluación de la Biodiversidad Forestal ante el Cambio Climático (BIODEF-CC)”. El objetivo principal del proyecto SPAINCLIM es la generación de un conjunto de Índices Bioclimáticos para toda España asociados a las simulaciones por downscaling estadístico de un conjunto de simulaciones globales de modelos climáticos. Como ejemplo de aplicación de estos índices bioclimáticos a uno de los factores más preocupantes para la biodiversidad en España, se ha previsto un estudio de distribución potencial de especies vegetales invasoras de interés peninsular, modelizando su idoneidad en escenarios presentes y futuros. Por tanto, el segundo objetivo específico del proyecto SPAINCLIM es «aplicar los Índices Bioclimáticos generados (a presente y a futuro) al estudio de la distribución presente y futura de las especies vegetales invasoras más relevantes para el conjunto del territorio español peninsular». Las principales actividades diseñadas específicamente para realizar este estudio aplicado son:

- **Actividad 5. Selección de las especies vegetales invasoras más relevantes para el conjunto del territorio español peninsular.** Se recopilarán datos bibliográficos científicos y legislativos, en base a los que se seleccionarán las plantas invasoras que, por su potencial impacto, resulten a priori más preocupantes a nivel nacional.
- **Actividad 6. Ajuste, validación y proyección de los modelos de distribución de especies a las plantas invasoras seleccionadas.** Con la base de datos de presencias existentes de las especies seleccionadas (IFN, GBIF), y los Índices Climáticos desarrollados, se procederá al ajuste, validación y evaluación de distintos modelos (Maxent, GLM, GAM, etc.), para proyectar los más robustos a los diferentes escenarios de clima futuro generados anteriormente.

Dentro de este marco, el presente documento se corresponde con la Actividad 6 “Ajuste, validación y proyección de los modelos de distribución de especies a las plantas invasoras seleccionadas”. Con la intención de que los resultados obtenidos sean comprensibles por un público no excesivamente especializado, en primer lugar se explican los conceptos básicos y la metodología general seguida en todas las modelizaciones realizadas. Después se detallan las particularidades metodológicas y las fuentes de datos para cada especie analizada, así como los resultados obtenidos en cada caso. Finalmente, en el último apartado, se discuten los resultados de forma conjunta para facilitar una visión más integral.

1 Metodología general

1.1 Definiciones y estructura de la modelización

Los modelos de distribución de especies -MDE- son algoritmos que establecen una relación matemática entre los puntos conocidos de la distribución de una especie y diferentes variables que describen las condiciones ambientales, extrapolando esta relación al resto del área de estudio. El resultado es un mapa que nos indica la idoneidad de presencia de la especie en el área de estudio. Estos modelos nos permiten ver cómo varía la distribución de las especies en el espacio y en el tiempo cuando cambiamos las condiciones ambientales que introducimos en el modelo.

Los MDE se utilizan ampliamente en ecología con diversos objetivos, como, por ejemplo, describir patrones y hacer predicciones de las distribuciones geográficas de especies (Elith y Leathwick, 2009), facilitar la gestión especies y territorios (Arribas et al., 2012, Schwartz, 2012), descubrir impactos derivados del cambio climático en especies y ecosistemas (Felicísimo et al., 2012), o estudiar el potencial invasor de especies invasoras (Ficetola et al., 2007 y Gallien et al., 2012). En el campo de las invasiones biológicas, los MDE proporcionan información objetiva de qué especies o regiones serán más susceptibles a la invasión por parte de otras especies y al cambio climático y, por tanto, constituyen una herramienta muy útil para identificar regiones de gran importancia para la conservación. Asimismo, estos MDE sirven para generar mapas que localicen zonas de gran idoneidad de las especies invasoras y que permitan detectar tempranamente posibles invasiones.

Es necesario advertir que los mapas de idoneidad resultantes nos informan sobre la potencialidad de la distribución, mientras que la distribución real depende también de otros factores. Para obtener predicciones más realistas se recomienda cruzar esta información de idoneidad ambiental (fundamentalmente climática) con otras fuentes de datos (Capdevila-Argüelles et al., 2011). Destacamos tres cuestiones que deberán contemplarse a la hora de enriquecer y complementar los resultados obtenidos en este estudio:

- La presión de propágulos, junto con estudios sobre las vías de entrada y dispersión de la especie estudiada. Si no llegan propágulos a una zona con alto potencial, no se establecerá allí la planta.
- Las interacciones bióticas. Como veremos ya se apunta más adelante en este trabajo, los ecosistemas receptores presentan diferentes niveles de vulnerabilidad. Y además esta vulnerabilidad puede cambiar con las condiciones climáticas futuras.
- Las actividades antrópicas, que pueden ser favorecedoras o mitigadoras de la expansión de la planta.

Para generar los MDE se sigue un proceso complejo que se puede descomponer en cuatro pasos principales: 1) la recopilación de los datos de presencia de la especie (variable respuesta), 2) obtención de los valores de las variables ambientales (variables predictoras), 3) el ajuste de un modelo de predicción de idoneidad de presencia a partir de las variables ambientales en los puntos de presencias reales y, por último, 4) proyección del modelo en las condiciones temporales (presente, pasado o futuro) y geográficas (área nativa o área de invasión) que nos interesen.

1.2 Preparación y filtrado de las presencias de cada especie

La primera parte del proceso consiste en la preparación de los datos sobre los que se va a trabajar. Este es uno de los puntos clave del proceso de modelización y en el que se debe invertir gran parte del tiempo debido a que se necesita una cantidad importante de datos de presencia de la especie, sin errores y con una georreferenciación precisa e información ambiental suficiente del área de estudio.

Antes de generar los modelos es imprescindible analizar los datos de presencias de las especies para eliminar los posibles errores existentes. Este proceso es especialmente importante para depurar los datos de presencias obtenidos de bases de datos globales como GBIF (<http://www.gbif.org/>). Se utilizó esta base de datos por ser la más completa y, por tanto, la más frecuente en este tipo de estudios. Sin embargo, sus datos no están exentos de sesgos derivados de las diferencias en la precisión en las coordenadas, la identificación taxonómica o los esfuerzos de muestreo no uniformes en los diferentes territorios (Beck et al., 2014 y Mateo et al., 2013).

Entre los procesos necesarios en el tratamiento de datos están: la correcta identificación de las especies, la eliminación de presencias duplicadas, la verificación de la correcta georreferenciación y la eliminación de presencias con altos sesgos de muestreo (Hijmans y Elith, 2013). Por tanto, debemos eliminar presencias con dudas en la identificación o aquellas mal identificadas, eliminar o corregir discrepancias entre el campo de coordenadas y el de localidad o país, y eliminar las presencias duplicadas. Asimismo, se deben eliminar las presencias con alto sesgo de muestreo o con baja precisión en las coordenadas, pues algunos pliegos de herbario se digitalizan y publican en bases de datos con precisión de coordenadas de más de 20 kilómetros, lo que no permite localizar el punto con precisión en el espacio geográfico ni relacionarlo con datos climáticos cuya resolución es de 1km². Cuando varias presencias se sitúan muy cerca unas de otras, en un área de menos de 1km², se consideraron como una sola, para evitar la introducción de redundancia y ruido en el modelo.

Otra consideración importante en cuanto a los datos de presencia de la especie es que estos datos deben representar adecuadamente la distribución actual del taxon a modelizar, recogiendo en el modelo todo el rango ambiental que ocupa, para minimizar los riesgos de extrapolación fuera de la zona de entrenamiento de las variables ambientales (Hijmans y Elith, 2013). Esto resulta mucho más sencillo en el caso de especies de espectro ambiental reducido, propias de ambientes muy concretos y definidos (Mateo et al., 2010). Aunque nuestro mayor interés se centra en la España peninsular, la distribución y el número de presencias de algunas de las plantas estudiadas requieren de territorios más amplios para construir los modelos, por lo que se definió el área de estudio de la **Figura 1**.

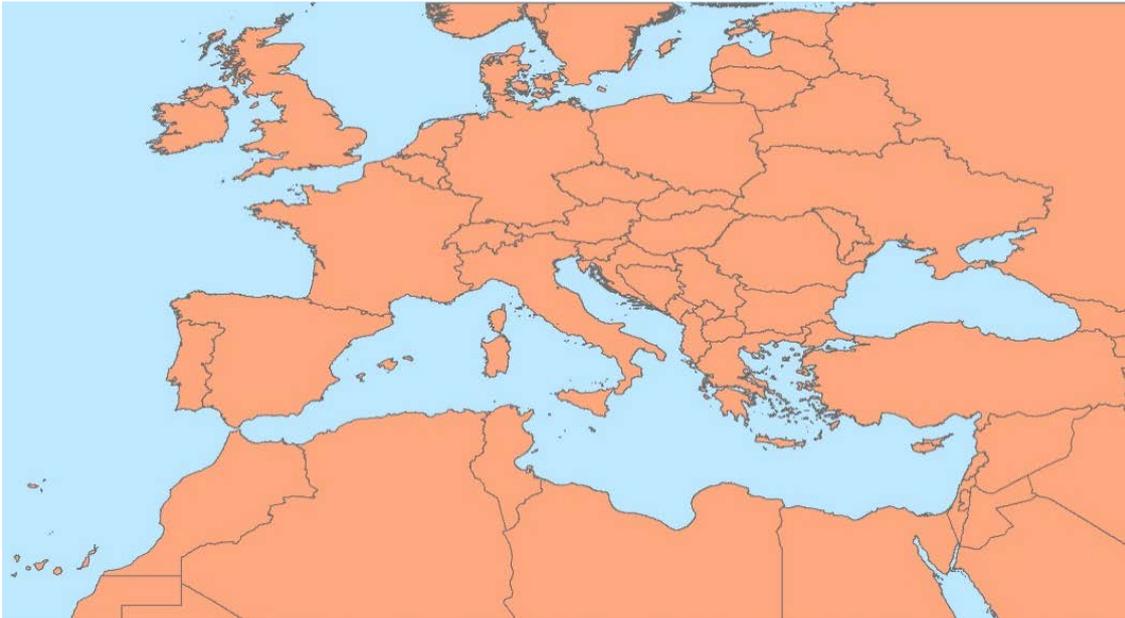


Figura 1. Zona de estudio definida.

Establecer el número mínimo de presencias para obtener un modelo fiable no es tarea fácil. Los estudios consultados aportan cifras diferentes, pero coinciden en que la calidad aumenta con el tamaño muestral (Stockwell & Peterson, 2002; Wisz et al., 2008; Mateo et al., 2010; van Proosdij et al., 2016; Warren et al., 2019; Santini et al., 2020). Uno de los factores importantes es la prevalencia de la especie, definida como la fracción del área de estudio ocupada por dicha especie, de forma que se necesita un tamaño muestral mayor para especies ampliamente distribuidas que para especies con una baja prevalencia (van Proosdij et al., 2016). En las modelizaciones de este estudio hemos trabajado siempre con un mínimo de 150 presencias para cada especie.

1.3 Obtención de variables ambientales

Para realizar los MDE, además de registros de presencia de la especie, es necesario disponer de información ambiental (temperatura, precipitación, topografía, tipo de suelo, etc.) del área de estudio. Aunque se pueden introducir todo tipo de variables, el cuerpo más potente es casi siempre climático. Como en la mayoría de los estudios de este tipo, se utilizaron los índices bioclimáticos recogidos en WorldClim (**Tabla 1**), derivados de las temperaturas y precipitaciones mensuales obtenidas mediante la interpolación de datos climáticos de estaciones meteorológicas. Presentan un gran sentido biológico como factores limitantes a la hora de explicar la distribución de los organismos, ya que no sólo muestran tendencias mensuales, sino que también incluyen tendencias trimestrales importantes para la vegetación (Kumar y Stohlgren, 2009).

Tabla 1. Variables bioclimáticas usadas en el estudio (Hijmans et al., 2005).

Código	Descripción
BIO 01	Temperatura media anual
BIO 02	Rango diario de temperaturas (media mensual (max temp - min temp))
BIO 03	Isotermalidad: Relación entre el rango de temperaturas entre el día y la noche y entre el verano e invierno (BIO02/BIO07) (* 100)
BIO 04	Estacionalidad en la temperatura (desviación estándar*100)
BIO 05	Temperatura máxima del mes más cálido
BIO 06	Temperatura mínima del mes más frío
BIO 07	Rango de temperatura anual (BIO05-BIO06)
BIO 08	Temperatura media del trimestre más húmedo
BIO 09	Temperatura media del trimestre más seco
BIO 10	Temperatura media del trimestre más cálido
BIO 11	Temperatura media del trimestre más frío
BIO 12	Precipitación anual
BIO 13	Precipitación del mes más húmedo
BIO 14	Precipitación del mes más seco
BIO 15	Estacionalidad en la precipitación (Coeficiente de variación)
BIO 16	Precipitación del trimestre más húmedo
BIO 17	Precipitación del trimestre más seco
BIO 18	Precipitación del trimestre más cálido
BIO 19	Precipitación del trimestre más frío

La base de datos climáticos elegida depende del área de estudio definida en cada caso. En las modelizaciones realizadas en los proyectos previos BIOFOREST-CC y BIODEF-CC, la base de datos climáticos consistió en datos observados de temperatura y precipitación de la red de estaciones meteorológicas de la Agencia Nacional de Meteorología (AEMET; serie 1951–2006), solo disponible para España, con 1830 estaciones de temperatura y 5053 estaciones de precipitación. Es el caso de *Acacia dealbata*, *Ailanthus altissima*, *Arundo donax*, *Cortaderia selloana* y *Robinia pseudoacacia*. Sin embargo, para *Baccharis halimifolia*, *Carpobrotus acinaciformis* y *Opuntia ficus-indica*, las presencias ibéricas no parecen recoger la variabilidad ambiental de toda su distribución actual en la zona de estudio (**Figura 2**), por lo que se incorporaron datos de varios países europeos. Para estos taxones se utilizó la base de datos de WORLDCLIM porque, aunque disponga de menor densidad de estaciones en el territorio español, está disponible a nivel mundial (**Figura 3**).

Además de los 19 índices bioclimáticos ya explicados, se introdujo la pendiente topográfica con una resolución de píxeles de 1 km² como otra variable ambiental más. Los valores de pendiente obtuvieron a partir de un modelo digital del terreno que tiene sus unidades de latitud y longitud en grados sexagesimales y sobre el que se realizó la conversión a píxeles de 1 km² para toda el área de estudio de la **Figura 1**.

Los datos de WORLDCLIM se pueden descargar ya en formato ráster con una precisión de 1 km² (Fick & Hijmans, 2017). Sin embargo, en caso de utilizar datos de la AEMET como en proyectos anteriores, se pueden interpolar en una cuadrícula con una resolución de 1 km para la España continental siguiendo la metodología de interpolación Thin Plate Splines (Hutchinson, 1991), y el uso de la altitud como covariable para la temperatura y la precipitación. Finalmente, las 19 variables bioclimáticas para la actualidad se pueden obtener a través del paquete "dismo" R (Hijmans et al., 2015).

Las condiciones climáticas futuras en la España continental fueron generadas por la FIC para las proyecciones del 5º Informe del Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC5) para los periodos temporales 2050 (2041-2060) y 2070 (2061-2080) y dos trayectorias de concentración representativas (RCPs: RCP4.5 un escenario de mitigación medio y RCP8.5 un escenario de altas emisiones) y usando siete de los modelos climáticos más recientes (GCM- modelos de circulación general) (Tabla 2). Siguiendo la metodología de interpolación descrita anteriormente, se obtuvieron los 19 índices bioclimáticos futuros para la Península Ibérica a 1 km con la mediana de los siete modelos climáticos.

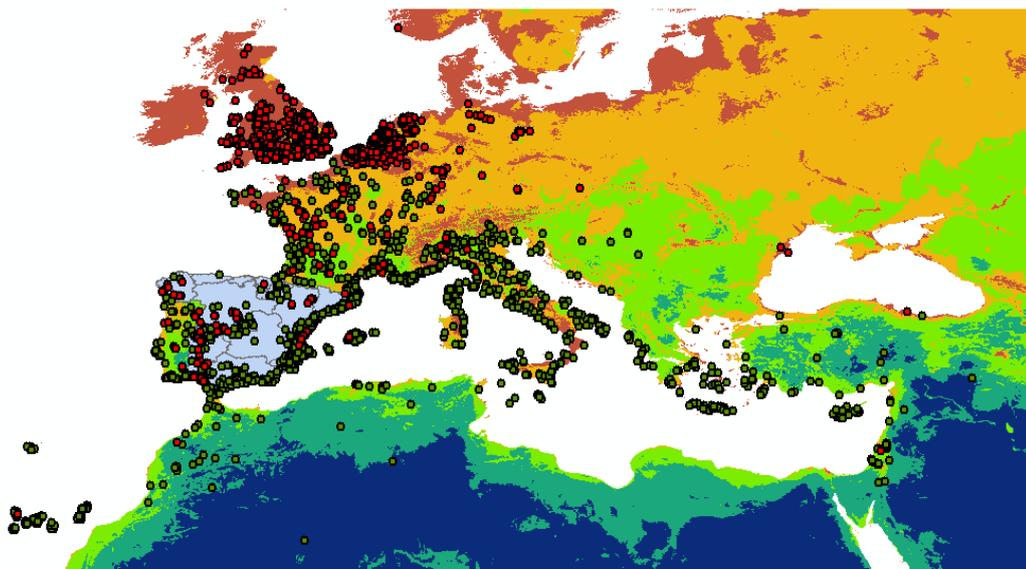


Figura 2. Zonas de estudio para las plantas exóticas invasoras analizadas. En algunos casos se toman presencias de España continental, mientras que en otros se incluyen datos de varios países de la Europa mediterránea. Fuente: elaboración propia a partir de datos obtenidos de GBIF.org.

Tabla 2. Modelos climáticos usados en el estudio para simular las condiciones de clima futuro en la Península Ibérica (CMIP5). <http://cmip-pcmdi.llnl.gov/cmip5/>

Modelo	Código	Institución	País	Resolución espacial (lon x lat)
BCC-CSM1-1	BC	Beijing Climate Center (BCC), China Meteorological Administration	China	2.7905° × 2.8125°
CNRM-CM5	CN	Centre National de Recherches Meteorologiques/Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique (CNRM-CERFACS)	France	1.4° × 1.4°
HADGEM2-CC	HG	Met Office Hadley Centre (MOHC)	UK	1.87° × 1.25°
MIROC-ESM-CHEM	MI	Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Atmosphere and Ocean Research Institute (AORI), and National Institute for Environmental Studies (NIES)	Japan	2.7905° × 2.8125°
MPI-ESM-MR	MP	Max Planck Institute for Meteorology (MPI-M)	Germany	1.8 × 1.8°
MRI-CGCM3	MG	Meteorological Research Institute (MRI)	Japan	1° × 0.5°
NorESM1-M	NO	Norwegian Climate Centre (NCC)	Norway	2.5° × 1.9°

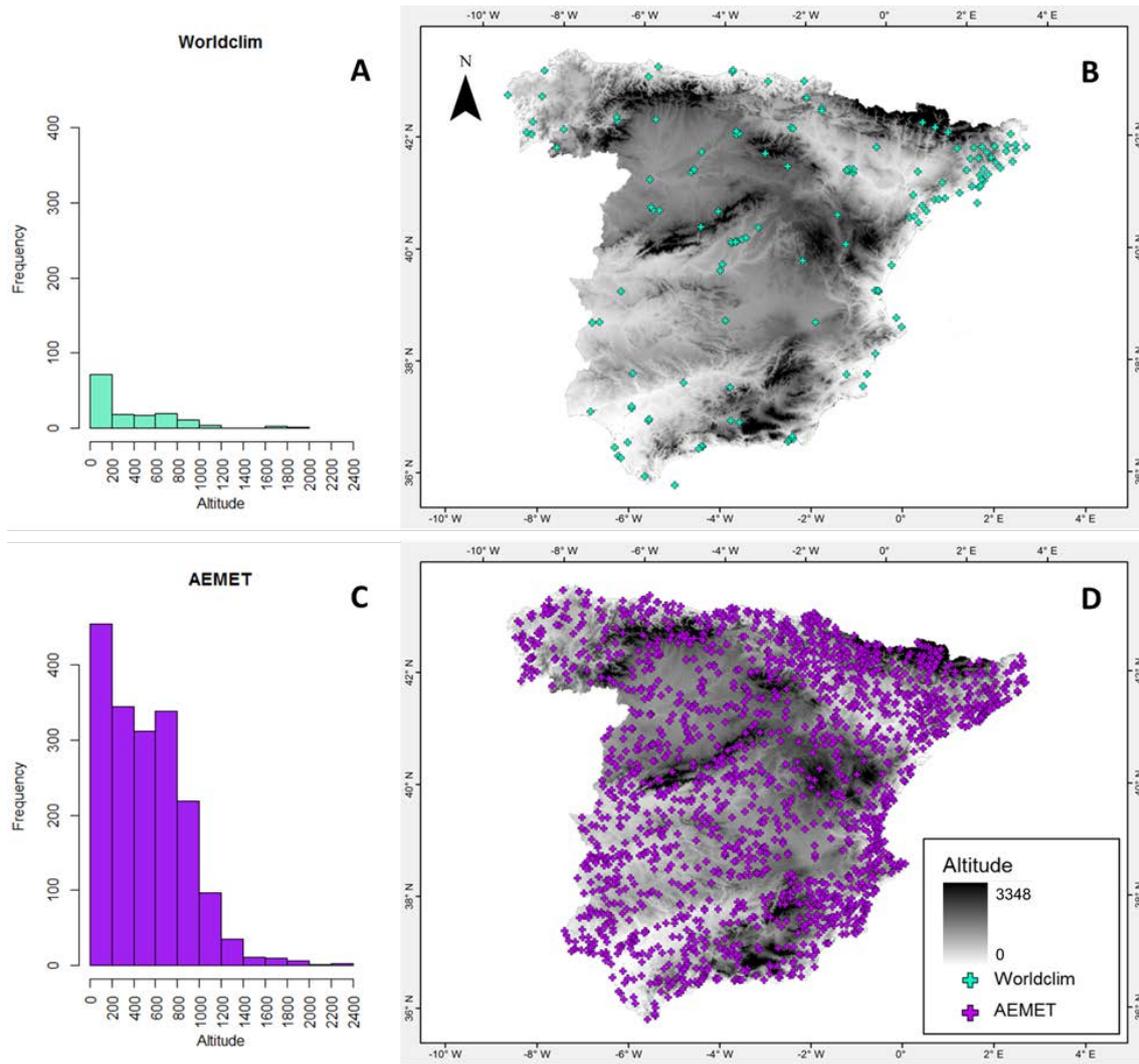


Figura 3. Estaciones meteorológicas usadas para generar la base de datos de clima usada en el presente estudio (C – distribución altitudinal de las estaciones y D- distribución espacial) y las estaciones usadas en la base de datos de WORLDCLIM para la Península Ibérica (A – distribución altitudinal de las estaciones y B- distribución espacial). Fuente: Moreno-Amat (2017).

1.4 Selección de las variables ambientales

La elección de las variables ambientales es otra parte importante y compleja del proceso de generación de los modelos, que influye especialmente en las proyecciones al futuro (Harris et al., 2013). Es necesario analizar previamente la correlación existente entre las variables para evitar incluir variables correlacionadas en el mismo modelo, lo que aportaría información redundante afectando negativamente a la capacidad predictiva del modelo. La reducción de variables se puede realizar mediante un análisis de correlación de las variables ambientales en los puntos de presencia de la especie o en puntos al azar del área de estudio, calculando el coeficiente de correlación de Pearson. Las variables se comparan dos a dos para descartar la variable con menor sentido biológico y con mayor correlación con el resto de variables cuando

el coeficiente de correlación es superior a cierto valor. Asimismo se puede realizar un análisis de factor de inflación de varianza VIF (*Variance Inflation Factor*) de las variables climáticas para eliminar variables correlacionadas y que puedan presentar problemas de multicolinealidad.

En este estudio, las variables seleccionadas para modelizar cada especie fueron elegidas a través de un proceso de selección basado en la ecología de la especie (Prentice et al., 1992), y la correlación (Pearson $r \leq 0.75$) y colinealidad de las variables [VIF <5; a través del paquete "HH" R (Heiberger, 2015)]. En primer lugar, se observó la colinealidad (VIF) de todas las variables y se propusieron para eliminación las de mayor valor VIF. Paralelamente, se observaron los valores de correlación dos a dos (Pearson r) y el dendrograma cluster para saber qué variables estaban más relacionadas entre sí. Teniendo en cuenta, además de lo anterior, el sentido ecológico de cada variable, se fueron eliminando de una en una hasta conseguir un conjunto en el que el VIF global fuera inferior a 5. También se realizaron modelos preliminares a cada paso para asegurar que todas las variables seleccionadas en el conjunto final influían significativamente en el modelo (a través del indicador R^2 que compara los residuos del modelo con un modelo aleatorio).

1.5 Ajuste de los modelos de distribución de especies en el presente

En el presente estudio, todas las modelizaciones se han realizado con algoritmos GLM, que son métodos tradicionales de regresión, en concreto, modelos lineales generalizados (GLM son sus correspondientes siglas en inglés). Es una de las metodologías más frecuentes en estudios sobre distribución de especies, pues responde directamente al planteamiento matemático básico de tratar de explicar una variable por medio de otras. Otros algoritmos también frecuentes en este tipo de estudios son los modelos aditivos generalizados (GAM) y los denominados "*machine learning*" Random Forest y Maxent (este último utilizado en los proyectos BIOFOREST-CC y BIODEF-CC). Para facilitar la interpretación de las curvas de respuesta de las variables, se incluyen también modelizaciones GAM realizadas con los mismos conjuntos de variables que GLM. Por tanto, se utilizó el algoritmo GLM para generar los mapas de idoneidad ambiental presente y futura y el algoritmo GAM para complementar la interpretación de las curvas de respuesta de las variables ambientales.

Generación de pseudo-ausencias

Tanto GLM como los otros algoritmos citados en el párrafo anterior se consideran técnicas discriminantes y necesitan disponer de presencias observadas, o presencias y ausencias observadas que informen sobre la distribución actual de la especie objeto de estudio. Sin embargo, las ausencias son difíciles de obtener e incluyen sesgos en el modelo ya que pudiera ser que la especie no se encuentre en la zona inventariada por incompatibilidad con el clima o por otros motivos, como la desaparición por la acción antrópica, por contingencia (factores

derivados del azar), o que estando presente no haya sido detectada (por ejemplo, por desconocimiento de la especie en el caso de especies invasoras).

Si no se dispone de datos de ausencias verdaderas, el algoritmo de modelización necesitará o generará su propio set de puntos para que el modelo clasifique entre presencias y ausencias (Hanberry et al., 2012). Cuando son generadas directamente por el algoritmo, como en el caso de Maxent, el conjunto de “ausencias” se denomina background (Phillips et al., 2006; Phillips y Dudík, 2008). Cuando debemos crearlas de forma activa, como en el caso de GLM, se denominan pseudo-ausencias. Para crear estas pseudo-ausencias se definimos el número de puntos a generar y si estos deben tener alguna característica determinada o se crearán simplemente al azar. Se recomienda crear pseudo-ausencias con distancia tanto geográfica como ambiental a las presencias conocidas.

En el caso de especies de rango ambiental muy restringido, si se generan pseudo-ausencias al azar por todo el territorio, parecerá que el modelo se ajusta a los datos mejor de lo que realmente lo hace; por tanto, se suele forzar la creación de la mitad de las pseudo-ausencias en un área más acorde a los ambientes propicios para la especie, pero manteniendo el criterio de distancia geográfica y ambiental.

Conjuntos de datos para calibración y evaluación

Una vez preparados los registros de presencias de la especie en el área de estudio, se realizará una partición de los datos en dos grupos (**Figura 4**), datos de calibración y datos de evaluación, lo que implica tener un número suficiente de presencias de la especie (Mateo et al., 2011). El set de datos de calibración se usará para construir o entrenar el modelo y el set de datos de evaluación se usará para evaluar el modelo, es decir, para comprobar la calidad del modelo en datos diferentes a los utilizados para construirlo.

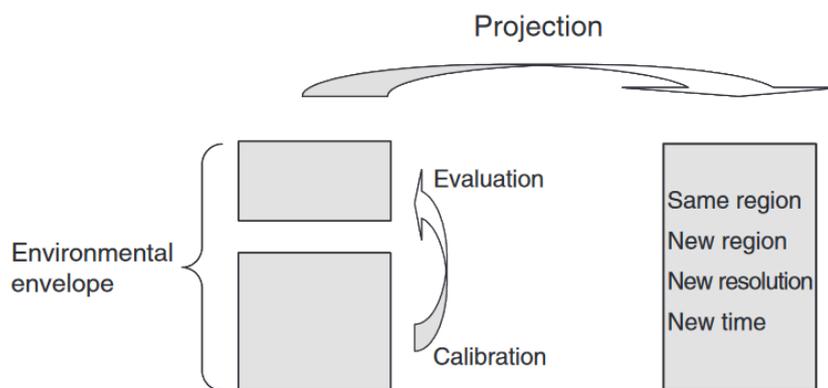


Figura 4. División de los datos para la calibración y evaluación de los modelos de distribución de especies. Fuente: Araújo et al. (2005).

En trabajos anteriores, derivados de los proyectos BIOFOREST-CC y BIODEF-CC se ha utilizado el algoritmo de modelización Maxent. Este algoritmo permite generar modelos solo con datos de presencia ya que genera sus propias "ausencias" denominadas background. (Phillips et al., 2006; Phillips y Dudík, 2008). Los datos de background no son considerados como verdaderas ausencias por el modelo sino puntos al azar dentro del área de estudio que permiten caracterizar el espacio ambiental donde la especie se desarrolla (Phillips y Dudík, 2008). En este estudio se han generado 10.000 puntos de background al azar en la Península Ibérica para capturar la variabilidad ambiental del área de estudio.

Maxent se utiliza frecuentemente en estudios que proyectan las distribuciones actuales de las especies hacia diferentes periodos temporales y espaciales (Elith et al., 2011). Maxent es uno de los algoritmos más usados en MDE, muestra muy buenos resultados en las comparaciones con otros métodos de modelización ecológica (Elith et al., 2006), permite la interpretación biológica de los resultados; y su uso está implementado en la plataforma R a través del paquete "dismo" (Hijmans et al., 2015).

Maxent calcula la probabilidad de distribución de máxima entropía (la distribución más próxima a la uniforme) de forma que el valor esperado de cada variable independiente (valor medio predicho por el modelo) debe coincidir con su media empírica (valor medio observado al muestrear las variables independientes con los datos de presencia de partida). Maxent permite valorar la influencia de las variables ambientales sobre el modelo final y produce otros gráficos en los que se indica la importancia de cada variable ambiental en la distribución de la especie (Elith et al., 2011).

Evaluación de los modelos

Los resultados del MDE deben ser evaluados estadísticamente para ver la capacidad predictiva del modelo; es decir, es necesario evaluar su capacidad de clasificar correctamente las presencias y ausencias observadas de la especie. Para tener una evaluación robusta se requieren datos de validación independientes a los de calibración. Para evaluar los modelos se puede realizar la validación cruzada (*cross-validation*) y la división de los datos en datos de calibración y de test puede ser aleatoria (al azar). En este trabajo se ha realizado una división al azar de los datos de presencia de cada especie en dos grupos, el 70% para calibración y el 30% restante para validación.

Una vez se tiene el set de datos de evaluación del modelo (datos de test), se contrasta la predicción del modelo en esos puntos comparando las presencias/ausencias predichas por el modelo con las presencias/ausencias reales y se evalúa el error que comete el modelo mediante la proporción de casos clasificados incorrectamente. Esto se repite para varias divisiones 70-30% realizadas al azar, se evalúa la bondad de ajuste del modelo en todas ellas y se toma el valor de la mediana de los resultados como valor definitivo de la bondad de ajuste del modelo. Para cuantificar esta capacidad de acierto del modelo y poder comparar con otros modelos alternativos, se pueden utilizar varios parámetros. La mayor parte de estos parámetros dependen del valor umbral (*threshold*) de idoneidad que establezcamos para discriminar entre presencias y ausencias. Entre ellos destacan la **sensibilidad (*sensitivity*)**, que es la capacidad de predecir correctamente las presencias; la **especificidad (*specificity*)**, que es la capacidad de predecir correctamente las ausencias; y el llamado **True Skill Statistic (*TSS*)**, que proporciona una medida combinada de ambas capacidades.

Sin embargo, el **AUC (área bajo la curva ROC)** es el estadístico de validación de MDE más usado en la actualidad. La curva ROC es la representación gráfica de la capacidad discriminativa de un modelo, es decir, la capacidad del modelo de distinguir entre presencias y ausencias (Fielding y Bell, 1997). El eje de ordenadas representa la sensibilidad y el de abscisas representa el

complementario de la especificidad (1-especificidad), o error de comisión (falsos positivos). La curva ROC es independiente del umbral ya que calcula la sensibilidad y el error de comisión para todos los umbrales (**Figura 5**) (Miller, 2010).

El área bajo la curva ROC corresponde a la probabilidad de que, al tomar al azar un par de casos, uno de presencia y otro de ausencia, el modelo dé un valor de idoneidad mayor a la presencia que a la ausencia. El valor de AUC varía entre 0 y 1, siendo 1 el valor que indica que todos los casos se han clasificado correctamente y 0.5 el valor que muestra que el modelo no clasifica los casos de forma distinta a una clasificación al azar. En el caso de usar pseudo-ausencias y no verdaderas ausencias, el AUC se interpreta como la probabilidad de que una presencia elegida al azar tenga mayor idoneidad que una pseudo-ausencia al azar. Este AUC es menos informativo que el generado con verdaderas ausencias ya que entre las pseudo-ausencias puede haber puntos de zonas potencialmente no muestreadas (Merow et al., 2013).

En la curva ROC (**Figura 5**), si la curva roja cae por debajo de la línea negra, indica que el modelo es peor que si se hubiese hecho al azar. En cambio, cuanto más se aproxime la línea roja a la esquina superior izquierda, el modelo tiene mejor capacidad de predicción.

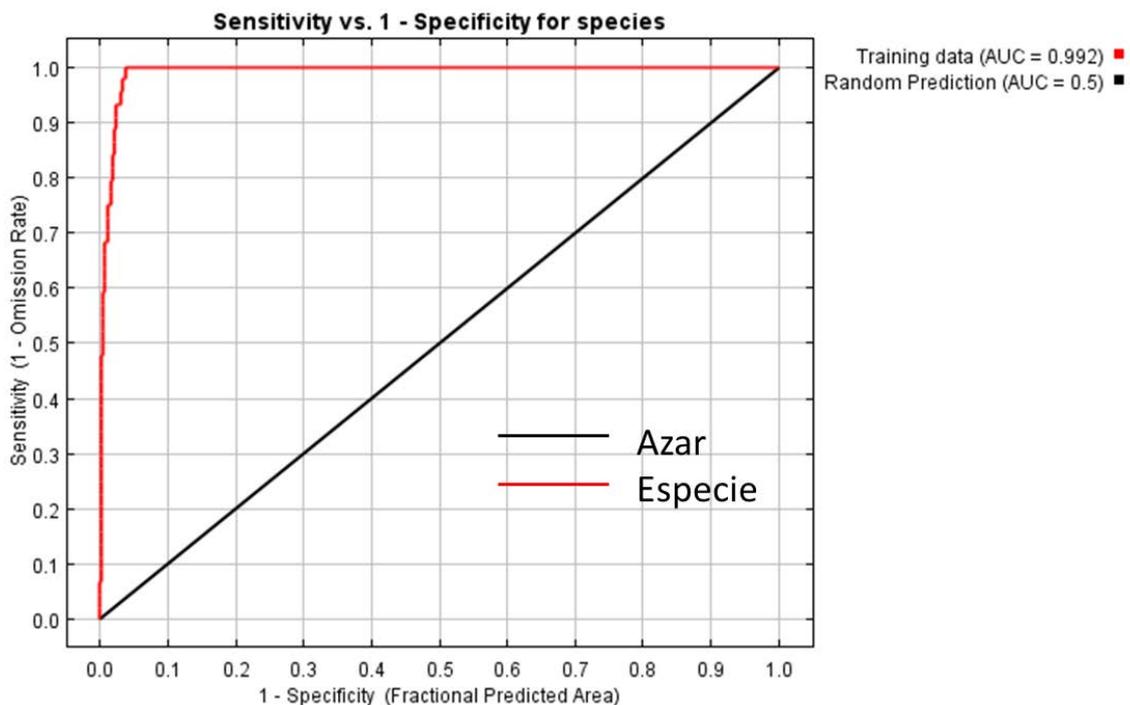


Figura 5. Ejemplo de gráfica del AUC (Área bajo la curva ROC). La curva roja representa el ajuste del modelo a los datos. La línea negra representa la línea esperada si el modelo no fuese mejor que por azar. La sensibilidad (*sensitivity*) es la proporción de presencias observadas correctamente predichas y la especificidad (*specificity*) es la proporción de ausencias observadas correctamente predichas.

El AUC permite comparar cualquier método de modelización y es una medida independiente del umbral de corte (Mateo et al., 2011). Sin embargo, presenta ciertas limitaciones. Por ejemplo, proporciona el mismo valor al error de comisión que al error de omisión cuando no es igual de importante clasificar una presencia como ausencia teniendo certeza de presencia que clasificar

una ausencia como presencia sin tener certeza de ausencia de la especie. Además, no aporta información de la distribución espacial del error del modelo (Mateo et al., 2011).

1.5 Proyección a escenarios futuros

Los modelos generados para el presente se pueden proyectar bajo diferentes condiciones climáticas en el espacio y en el tiempo, como los escenarios climáticos del IPCC5 sobre los que ha trabajado la FIC (actividades 1, 2, 3 y 4 del presente proyecto SPAINCLIM). Para realizar las proyecciones de clima futuras se han utilizado siete escenarios futuros de cambio climático propuestos por el IPCC en su quinto informe (IPCC, 2014). Estos escenarios futuros o trayectorias de concentración representativas (RCP, Representative Concentration Pathways) constituyen diferentes supuestos de emisiones de gases de efecto invernadero (GEI) a la atmósfera para el año 2100 y permiten generar las predicciones de las variables ambientales en el futuro y predecir la distribución futura de las especies mediante el uso de los MDE.

A la hora de proyectar los MDE bajo diferentes escenarios climáticos de los usados para calibrar el modelo hay que interpretar los resultados cautelosamente ya que los modelos se basan en ciertas asunciones, como la conservación del nicho y el equilibrio de la especie con el clima (Mateo et al., 2011).

En vez de elegir un modelo climático concreto, para mayor robustez de las proyecciones futuras, en este estudio los MDE generados en el presente para cada especie fueron proyectados al futuro usando la mediana de los 7 modelos climáticos disponibles. Por cada especie se han obtenido 4 MDE futuros, correspondientes a los dos periodos temporales 2050 (2041-2060) y 2070 (2061-2080) y a las dos trayectorias de concentración representativas (RCP4.5 y RCP8.5).

2 Modelizaciones realizadas

En esta segunda parte del documento se presentan las modelizaciones realizadas para *Arundo donax*, *Baccharis halimifolia*, *Carpobrotus acinaciformis* y *Opuntia ficus-indica*. La primera de estas plantas ya ha sido estudiada en proyectos anteriores (BIODEF-CC y BIOFOREST-CC). Lo que se aporta en este caso es el resultado de las mejoras efectuadas sobre los modelos iniciales, dando lugar a un MDE más fiable. Esta profundización en el estudio de la idoneidad ambiental de *Arundo donax* constituye el fundamento del proyecto final de grado de Joe D. Zambrano Suárez (todavía inédito).

El proceso de modelización suele ser fuertemente iterativo, pues, con frecuencia, la observación de los primeros resultados sugiere mejoras a implementar en los datos de entrada del modelo, por lo que tiene que volver a rodarse una y otra vez hasta lograr la conformidad de los analistas. Por tanto, para cada especie, se recogen detalles metodológicos importantes además de los resultados principales, ya que dichas características metodológicas pueden considerarse, conceptualmente, como parte de los resultados. Se incluyen también algunos comentarios breves sobre las idoneidades climáticas predichas, pero la discusión conjunta de resultados se presenta en un apartado propio.

2.1 *Arundo donax*

2.1.1 Datos de entrada en el MDE

Debido al gran número de presencias de *Arundo donax* en España y a la diversidad de ambientes en que se sitúan, el modelo se ha realizado solo con datos de España peninsular e Islas Baleares. Aunque se ve una densidad mucho mayor en el Este de la península Ibérica, también hay registros en el interior peninsular, en Galicia, en la costa cantábrica oriental, alto Ebro y Andalucía. La ventaja de incluir únicamente España peninsular e Islas Baleares es que nos permite trabajar con datos de índices bioclimáticos más precisos, como ya se ha explicado en apartados anteriores.

Obtención de presencias de la especie

Tras la descarga de datos de GBIF, se eliminaron los puntos duplicados y sin coordenadas, los que carecen de información de la incertidumbre y los que están demasiado cerca de otras presencias (distancia mínima de 1,5 km). De todo ese conjunto se eligieron solamente las presencias localizadas en España peninsular e Islas Baleares, obteniéndose un total de 6552 registros de presencia listos para introducir en el modelo (**Figura 6**). Este conjunto tan grande de presencias se dividió en seis subconjuntos para poder probar los algoritmos con sets de datos independientes, como se explica en el ajuste del modelo.

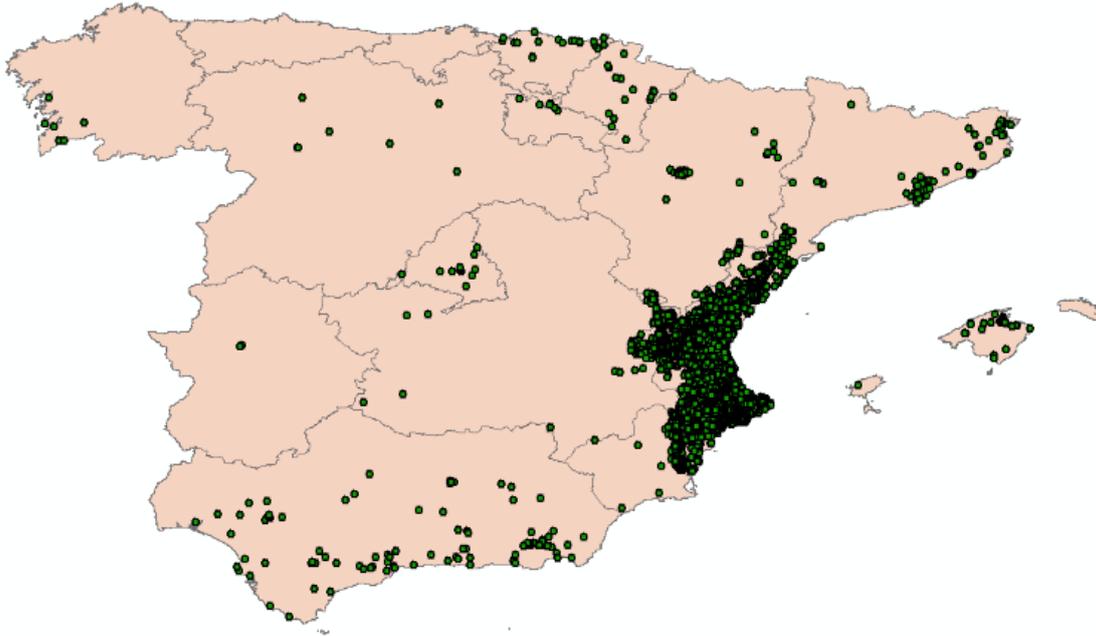


Figura 6. Registros de *Arundo donax* utilizados en la construcción del modelo. Fuente: elaboración propia a partir de datos de GBIF.org.

Generación de pseudo-ausencias

Para garantizar la distancia tanto geográfica como ecológica a las presencias de la especie, se definió un área susceptible de contener pseudo-ausencias que cumpliera dichos requisitos. De esta forma, el área incluye todos los territorios de España peninsular y Baleares que estén situados a más de 30 km de cualquier presencia y bajo una envuelta ambiental que no esté contenida en el 95% del intervalo de confianza del conjunto de presencias.

Para respetar la proporción adecuada respecto a los registros de presencia (cinco conjuntos de 1000 registros y un conjunto de 1552 registros), se crearon diferentes sets de pseudo-ausencias con el tripe de registros que las presencias a las que se asociaban. Antes de introducir los datos en el algoritmo, es necesario dividirlos en datos de calibración (70%) y datos de evaluación (30%). Finalmente, se crearon 24 conjuntos de datos para calibrar y evaluar los modelos GLM y otros 24 para los modelos GAM.

Selección de variables ambientales

A partir del conjunto de 19 índices bioclimáticos y la pendiente, se fueron descartando variables correlacionadas entre sí, con alta colinealidad y con menor sentido ecológico, siempre de una en una. Las variables seleccionadas, por tanto, son aquellas que mayor sentido ecológico presentaron y que cumplieran los requisitos de baja correlación entre sí (Pearson $r \leq 0.75$), baja colinealidad ($VIF < 5$) y significancia en el modelo GLM. Se trata de los cinco parámetros recogidos en la **Tabla 3**. En todo momento se utilizaron los datos generados por la FIC para

España peninsular y Baleares en el manejo de las variables correspondientes a los índices bioclimáticos.

Tabla 3. Variables explicativas para *Arundo donax*.

Código	Nombre
Bio2	Rango diario de temperaturas (media mensual (max temp - min temp))
Bio11	Temperatura media del trimestre más frío
Bio15	Estacionalidad en la precipitación (Coeficiente de variación)
Bio18	Precipitación del trimestre más cálido
Pendiente	Pendiente del terreno

2.1.2 Ajuste del MDE

Los modelos se generaron con el algoritmo GLM, las cinco variables seleccionadas (**Tabla 3**) y los seis sets de presencias y pseudo-ausencias independientes entre sí, realizando en cada uno dos repeticiones mediante subsampling y dos mediante bootstrapping. Por tanto, el número total de modelos con los que se calculó la bondad de ajuste fueron 24. Para estimar esta bondad de ajuste se tomaron los valores de las medianas de los estadísticos de evaluación (**Tabla 4**). El estadístico AUC se acerca mucho a la esquina superior izquierda, lo que nos indica que el modelo tiene una alta capacidad predictiva (**Figura 7**).

Tabla 4. Estadísticos de evaluación del modelo de distribución de *Arundo donax*.

	Minimo	1er cuartil	Mediana	Media	3er cuartil	Máximo
Threshold	0.1653	0.1988	0.2725	0.2698	0.3213	0.4000
TSS	0.8000	0.8250	0.8370	0.8372	0.8502	0.8750
Sensitivity	0.8760	0.8960	0.9200	0.9152	0.9360	0.9520
Specificity	0.8480	0.9140	0.9280	0.9220	0.9367	0.9570

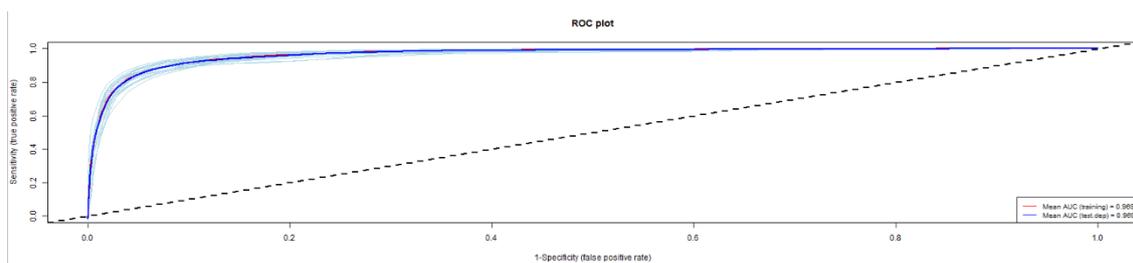


Figura 7. AUC (ROC plot) del modelo de distribución de *Arundo donax*. Fuente: elaboración propia.

Idoneidad ambiental actual

En la **Figura 8** se muestran los valores de idoneidad ambiental obtenidos para *Arundo donax* en España peninsular. Como se puede apreciar en dicha figura, existen amplias zonas de elevada idoneidad con una marcada tendencia hacia el Este de la península Ibérica. Las zonas de alta potencialidad más extensas y continuas se dan en Levante (Almería-Castellón) y el interior del

valle del Ebro. En segundo lugar estarían las zonas favorables detectadas en Andalucía (interior y costa mediterránea), el valle del Tajo, Cataluña y la costa cantábrica. Fuera de estas zonas, cabe destacar que las áreas claramente desfavorables (verde intenso) no son rotundas y continuas salvo en Galicia y Castilla León y algunos enclaves menores que podemos asociar al Sistema Ibérico en Aragón, al Pirineo con cierta penetración hacia el interior catalán y al Sur de Andalucía. En general, se aprecian muchas áreas de idoneidad moderada en contacto con las de alta y baja potencialidad, lo que podría indicar cierta disposición a cambios futuros.

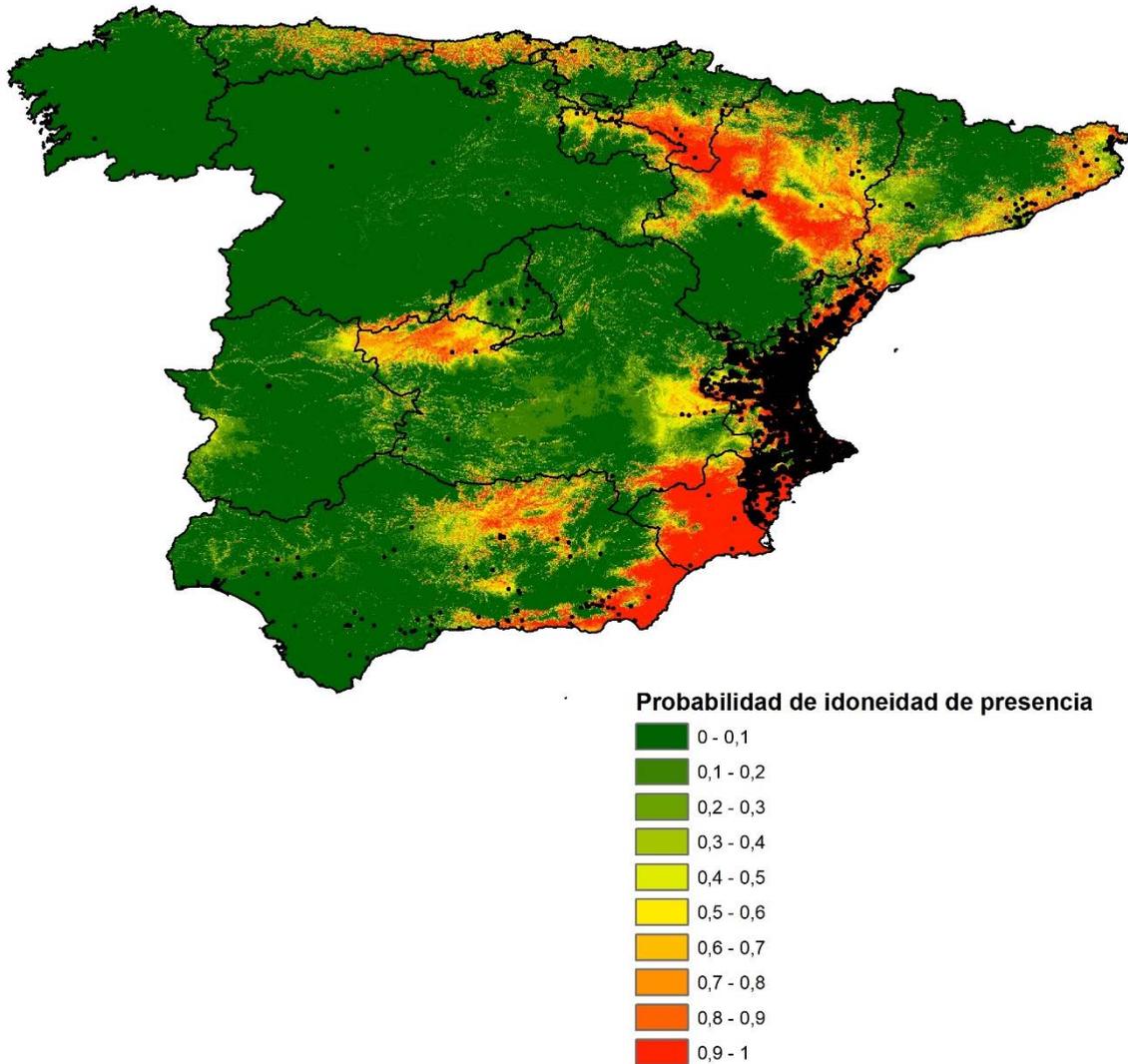


Figura 8. Idoneidad ambiental para *Arundo donax* en España peninsular. Los puntos negros representan las presencias actuales. Las líneas negras representan los límites autonómicos. Fuente: elaboración propia.

En la **Figura 9** se muestra una ampliación del mapa anterior para centrar la atención en las cuencas del Júcar y del Segura. Como se puede ver en dicha figura, *Arundo donax* ha ocupado ya gran parte de su área potencial, especialmente en la Comunidad Valenciana. En Murcia y Almería, sin embargo, existen zonas de elevada idoneidad con escasas presencias actuales. Al margen de posibles sesgos de muestreo, es probable que esta diferencia esté relacionada con

la actividad humana. En concreto, esta especie parece estar muy ligada a terrenos agrícolas abandonados, canales de riego de campos de cultivo y otros espacios antrópicos.

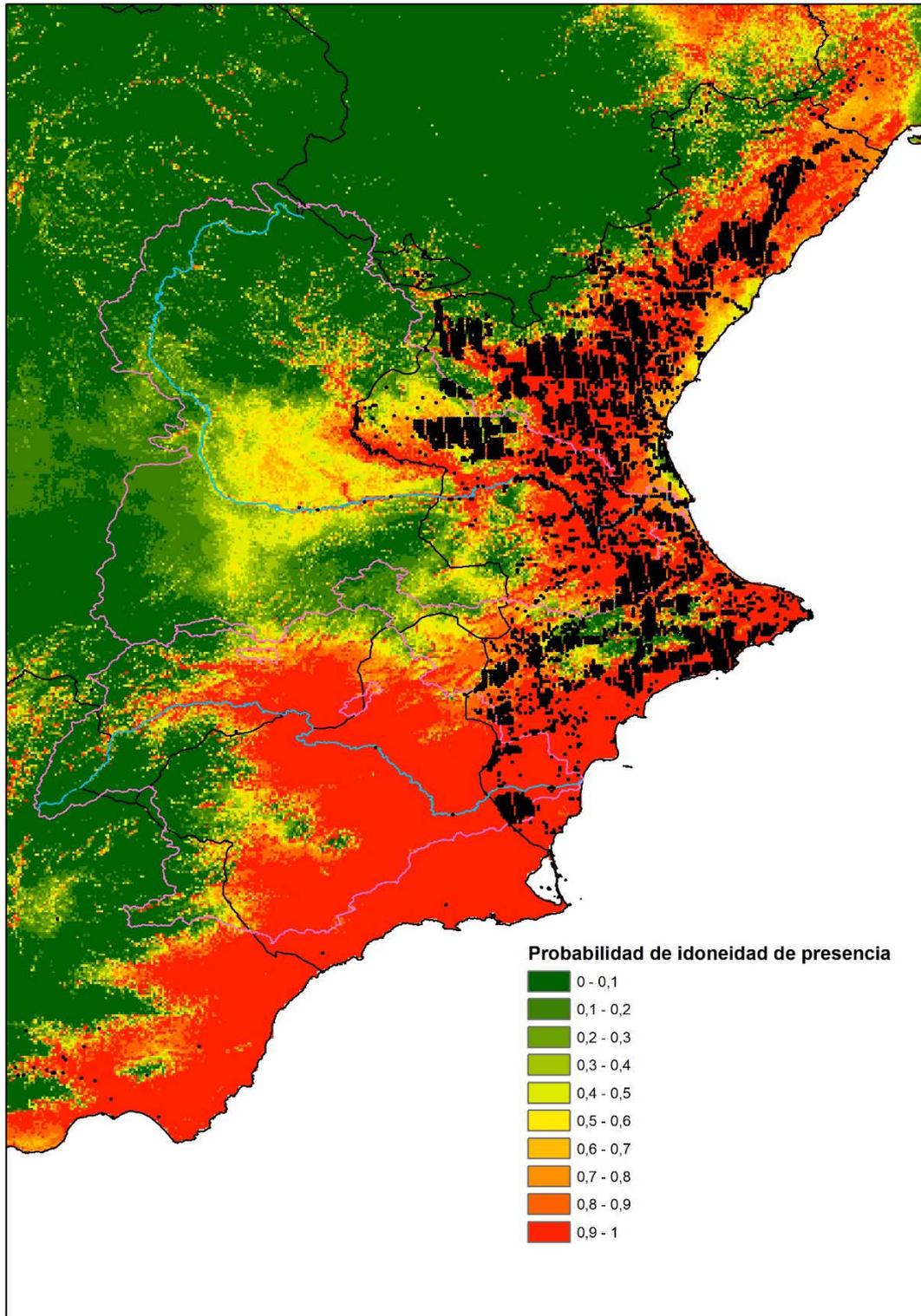


Figura 9. Idoneidad ambiental para *Arundo donax* en las cuencas del Júcar y del Segura. Los puntos negros representan las presencias actuales. Las líneas rosas representan los límites de las cuencas y las azules los cauces principales. Las líneas negras representan los límites autonómicos. Fuente: elaboración propia.

Curvas de respuesta a las variables

Las cinco variables predictoras tienen significancia en el modelo, pero no todas tienen el mismo peso a la hora de explicar la distribución de los registros (**Figura 22**). Ordenadas de mayor a menor importancia en el modelo son: Bio11 (temperatura media del trimestre más frío), Bio15 (estacionalidad en la precipitación), Bio18 (precipitación del trimestre más cálido), Bio2 (rango diario de temperaturas) y pendiente (pendiente topográfica a escala 1km²).

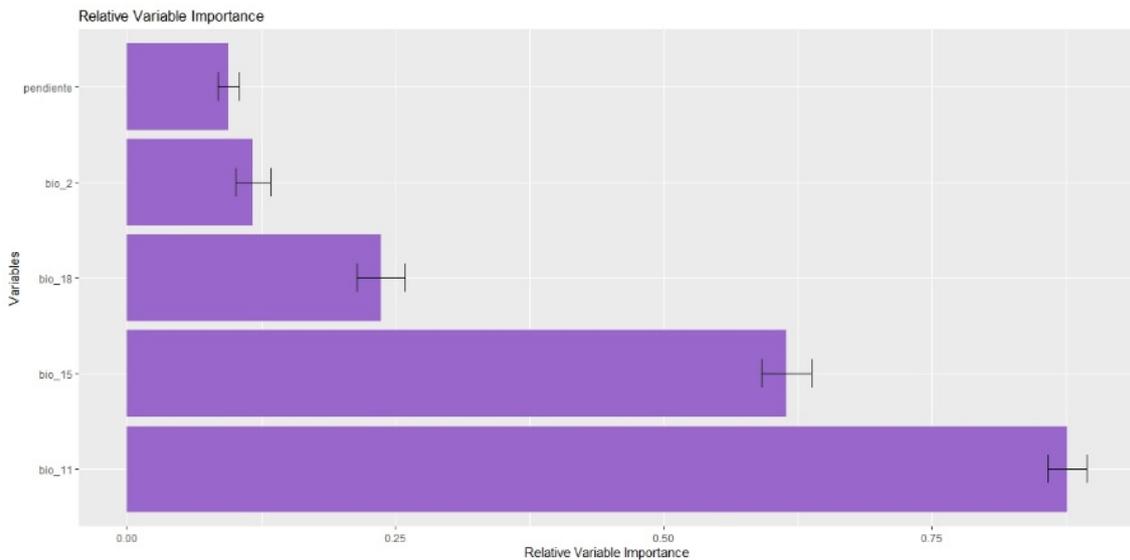


Figura 10. Importancia relativa de las variables en el modelo GLM para *Arundo donax*. Fuente: elaboración propia.

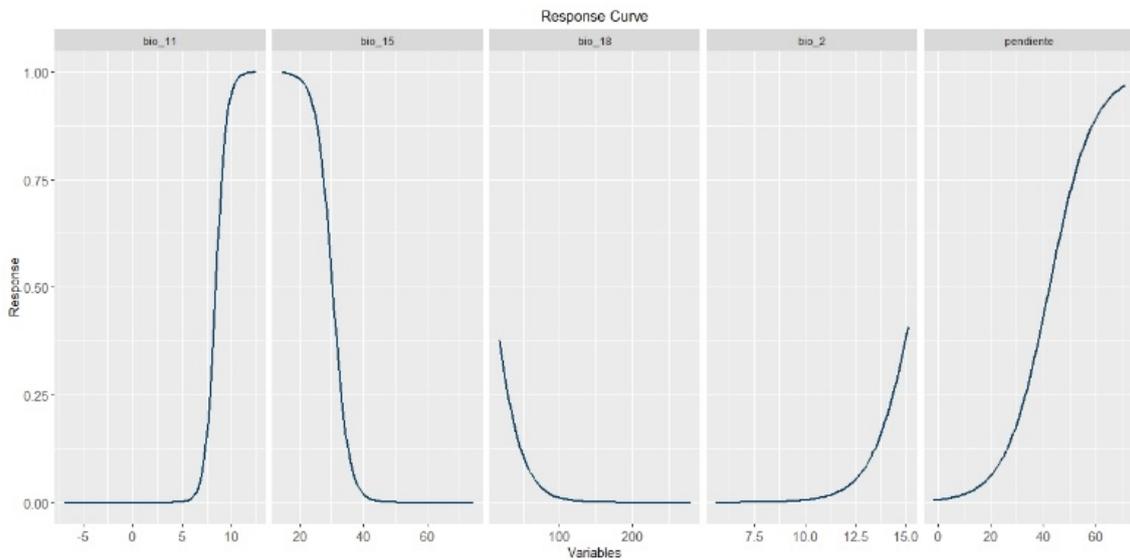


Figura 11. Curvas de respuesta a las variables ambientales en el modelo GLM para *Arundo donax*. Fuente: elaboración propia.

Como era de esperar, la idoneidad aumenta con la temperatura media del trimestre más frío (Bio11), mientras que la respuesta al resto de variables es sorprendente en algunos casos y

sencillamente no prevista con antelación en otros. En concreto, se observa una respuesta positiva al aumento del rango diario de temperaturas (Bio2) y a la pendiente, junto con una respuesta negativa al aumento de la precipitación del trimestre más cálido (Bio18) y la estacionalidad de la precipitación (Bio15) (Figura 11). Puesto que los modelos GLM obligan a que todas las curvas de respuesta presenten una forma sigmoïdal, para un análisis más detallado, es conveniente complementar estas tendencias iniciales del modelo GLM con las curvas de respuesta de un modelo GAM efectuado sobre el mismo conjunto de variables.

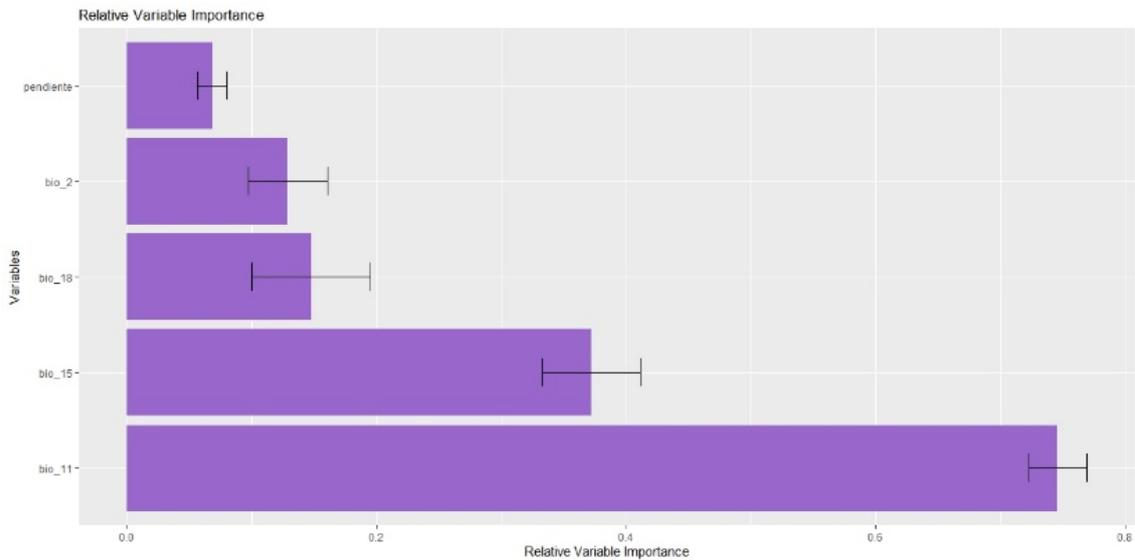


Figura 12. Importancia relativa de las variables en el modelo GAM para *Arundo donax*. Fuente: elaboración propia.

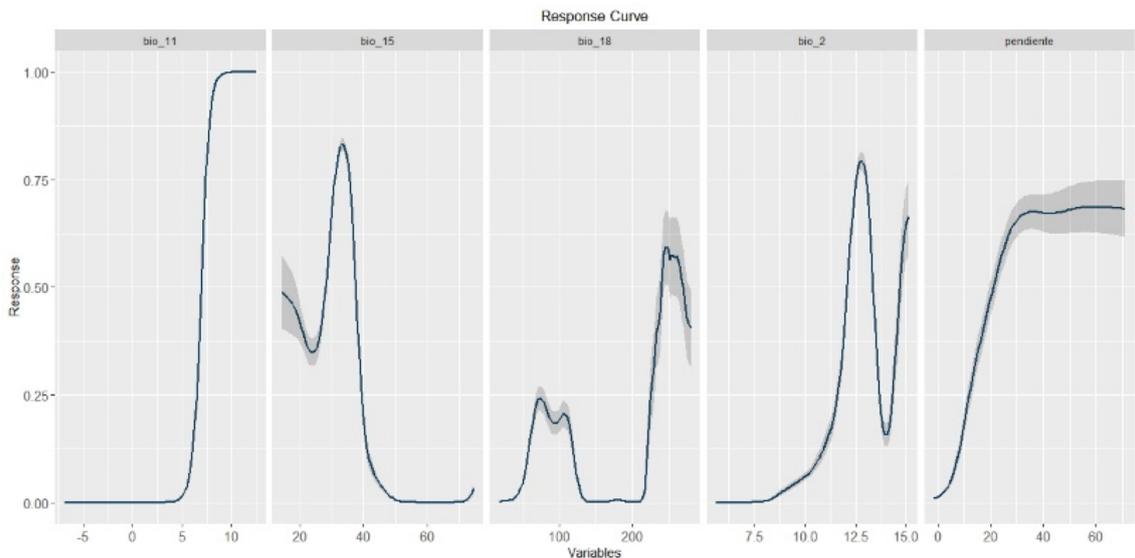


Figura 13. Curvas de respuesta a las variables ambientales en el modelo GAM para *Arundo donax*. Las zonas sombreadas en gris indican la existencia de incertidumbre en esa franja de la curva y marcan el intervalo de confianza. Fuente: elaboración propia.

La importancia relativa de las variables en el modelo GAM es muy similar a la que presentan en el modelo GLM (**Figura 12**). Las curvas de respuesta son acordes con las tendencias marcadas por el modelo GLM pero nos informan de algunos límites mejor ajustados (**Figura 13**).

La única variable que mantiene una curva sigmoide es la temperatura media del trimestre más frío (Bio11), que también es la que mayor peso tiene en el modelo, por lo que parece estar actuando como limitante para la especie, así que un aumento de las temperaturas invernales en el futuro podría favorecer claramente una expansión de *Arundo donax*. La estacionalidad de la precipitación (Bio15), sin embargo, parece mostrar un intervalo óptimo, por encima del cual disminuye y se mantiene en valores bajos la idoneidad, y por debajo del cual se aprecia una zona de gran incertidumbre que podría imitar también el comportamiento de la anterior. Es decir, que un cierto nivel de estacionalidad en la precipitación favorece a esta especie. Seguramente, esto es consecuencia de su preferencia por ambientes mediterráneos, tal como sucede con la precipitación del trimestre más cálido (Bio18). En ambos casos, por tanto, la respuesta observada parece estar relacionada con la distribución preferentemente mediterránea de *Arundo donax*, más que con el beneficio directo de la sequía estival o la irregularidad de las precipitaciones. El rango diurno de temperaturas (Bio2) también se asocia a una curva de respuesta compleja, con valores altos más bien beneficiosos. En nuestra opinión, es el reflejo de la tolerancia a la continentalidad que demuestra esta especie con algunas presencias en el interior peninsular. La pendiente parece favorecer a *Arundo donax* hasta un valor aproximado del 40%, que, a la escala de trabajo utilizada, es muy elevado, pues solo se da en la zona pirenaica y en puntos dispersos del resto de sistemas montañosos ibéricos. La interpretación de esta respuesta, por tanto, no es sencilla. Es probable que en las cuadrículas de 1km² con valores altos de pendiente coexistan zonas de mayor pendiente con enclaves más llanos propicios para esta planta, como, por ejemplo, los pies de los barrancos que delimitan ramblas.

2.1.3 Modelización de la distribución potencial futura:

Los cuatro escenarios futuros analizados muestran una idoneidad fuertemente creciente para *Arundo donax*. Las zonas claramente favorables de gran amplitud pasan de estar situadas en la mitad Este de la península Ibérica a cubrir casi todo el territorio, con la excepción de los macizos montañosos y el Noroeste (**Figura 14**).

Este aumento tan grande de la zona ambientalmente favorable para *Arundo donax* se produce, incluso, en los escenarios más favorables (RCP 4.5), seguramente por la cercanía actual a las temperaturas invernales limitantes para la especie. En el escenario de peores previsiones para el intervalo más lejano (RCP 8.5 para 2061-2080) se incrementa todavía más el territorio favorable. Ante esta situación, es difícil establecer medidas de gestión adecuadas, más allá de alertar de la posible amenaza creciente en el futuro. Para mejorar el enfoque hacia las áreas más sensibles, será necesario cruzar la información aquí obtenida sobre idoneidad climática con otras fuentes de datos que incluyan interacciones bióticas y actividades humanas. La superposición de estas capas informativas podría proporcionar mapas de riesgos más ajustados a la realidad, pues permitiría descartar aquellas áreas en las que *Arundo donax* no prospera bien por competencia con especies nativas (ej: bosques de ribera bien conservados), o donde las

actividades agrícolas podrían facilitar su control. Sin embargo, este tipo de estudios de mayor complejidad es difícil de abordar a escalas tan amplias. Con la intención de facilitar la aplicación de los resultados aquí obtenidos a posibles trabajos en esa línea, se presenta también una ampliación del detalle de los mapas en las dos cuencas hidrográficas más afectadas (Júcar y Segura) tanto para el escenario RCP 4.5 (**Figura 15**) como para el RCP 8.5 (**Figura 16**).

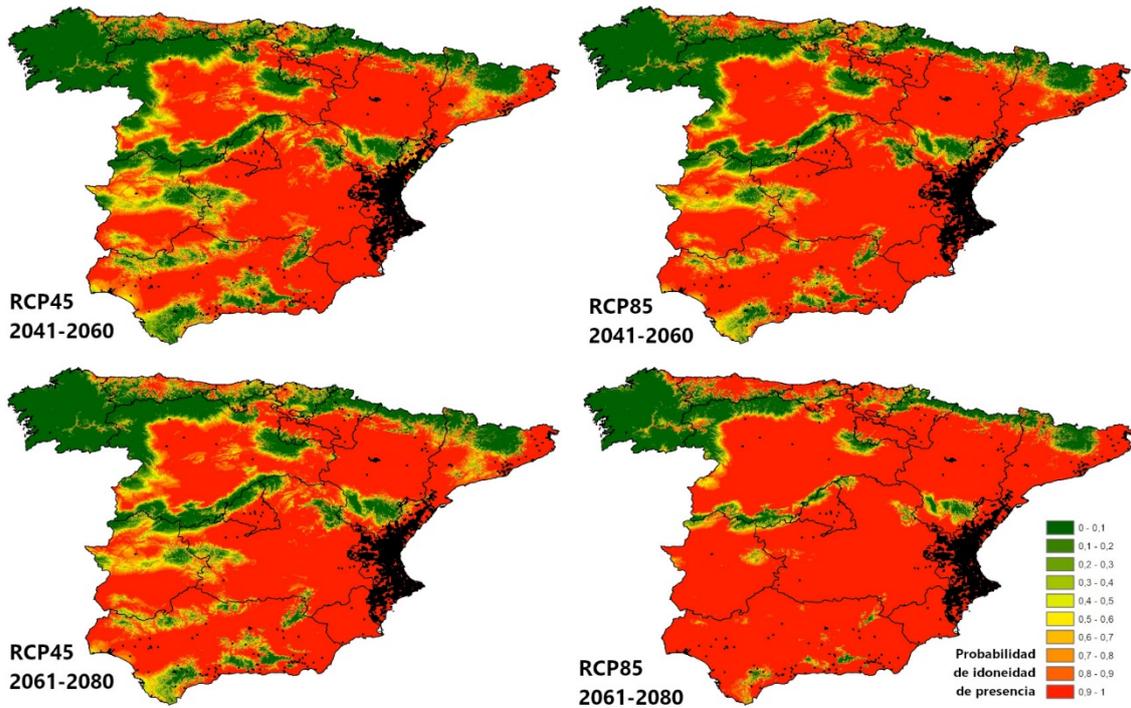


Figura 14. Idoneidad ambiental para *Arundo donax* al proyectar el modelo GLM bajo las condiciones climáticas futuras (escenarios RCP 4.5 y RCP 8.5 para los periodos 2041-2060 y 2061-2080). Fuente: elaboración propia.

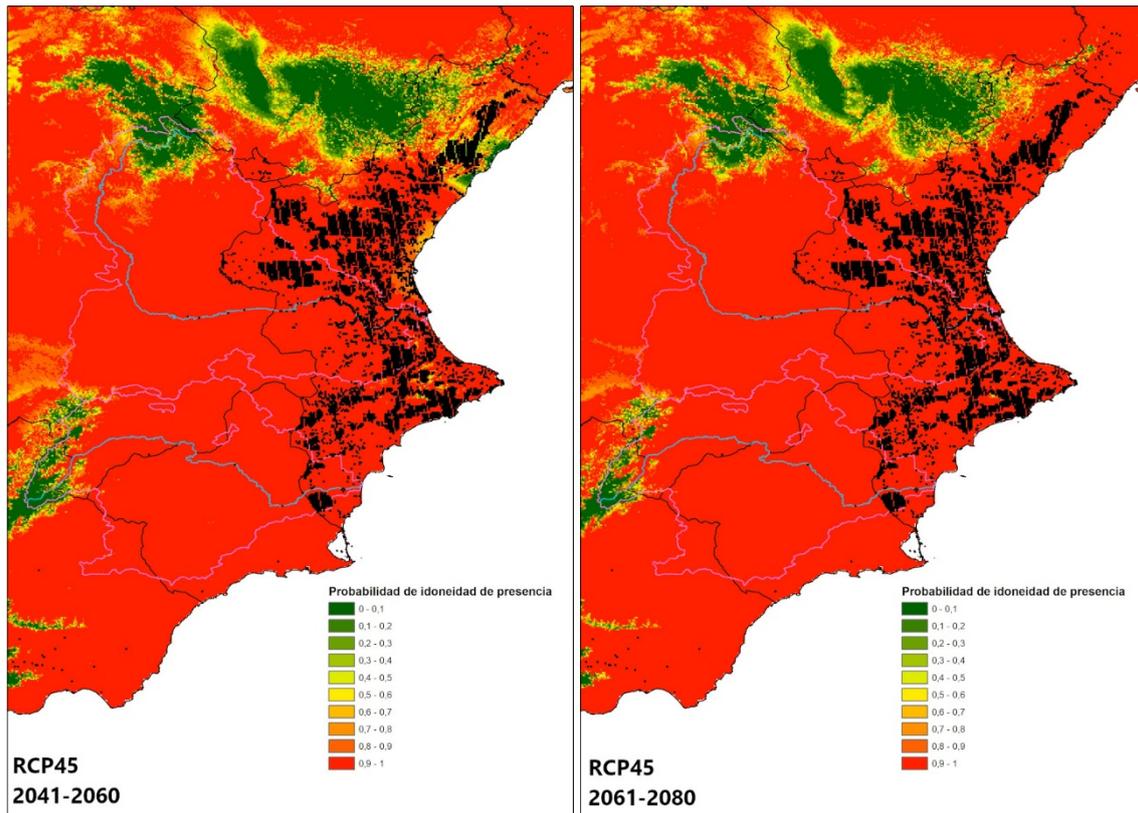


Figura 15. Ampliación sobre las cuencas del Júcar y del Segura del mapa de idoneidad ambiental para *Arundo donax* al proyectar el modelo GLM bajo las condiciones climáticas futuras del escenario RCP 4.5 para los periodos 2041-2060 y 2061-2080. Fuente: elaboración propia.

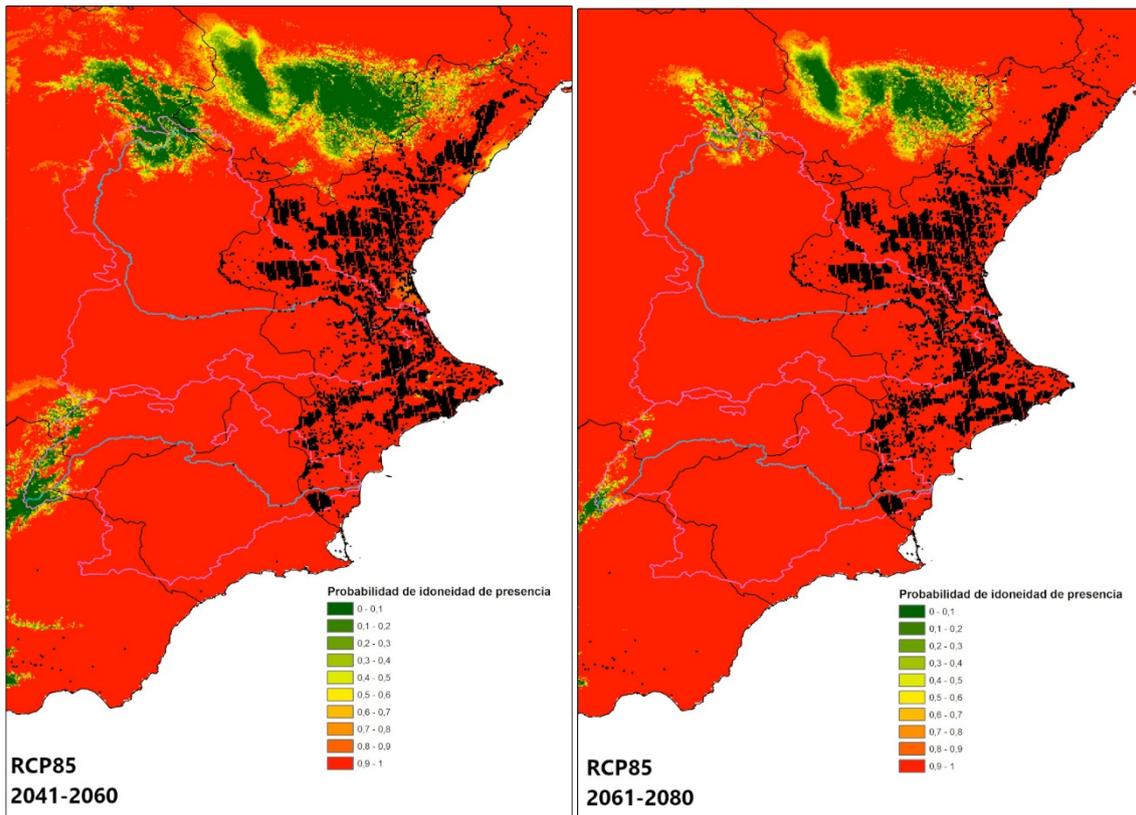


Figura 16. Ampliación sobre las cuencas del Júcar y del Segura del mapa de idoneidad ambiental para *Arundo donax* al proyectar el modelo GLM bajo las condiciones climáticas futuras del escenario RCP 8.5 para los periodos 2041-2060 y 2061-2080. Fuente: elaboración propia.

2.2 *Baccharis halimifolia*

2.2.1 Datos de entrada en el MDE

Debido al escaso número de presencias de *Baccharis halimifolia* en la península Ibérica, se han incluido datos de España, Francia y Países Bajos en la construcción del modelo y, por tanto, se han utilizado datos climáticos de WORLDCLIM. El objetivo ha sido recoger un rango ecológico suficientemente amplio. Se trata de una planta adaptada a ambientes húmedos costeros, con tolerancia a la salinidad, por lo que suele aparecer en marismas y estuarios. Actualmente ha penetrado poco en la costa cantábrica ibérica, probablemente procedente de las poblaciones más numerosas de la costa sur-atlántica francesa. En la **Figura 17** se presenta un mapa con la distribución actual en el Oeste de Europa (datos obtenidos de GBIF.org).

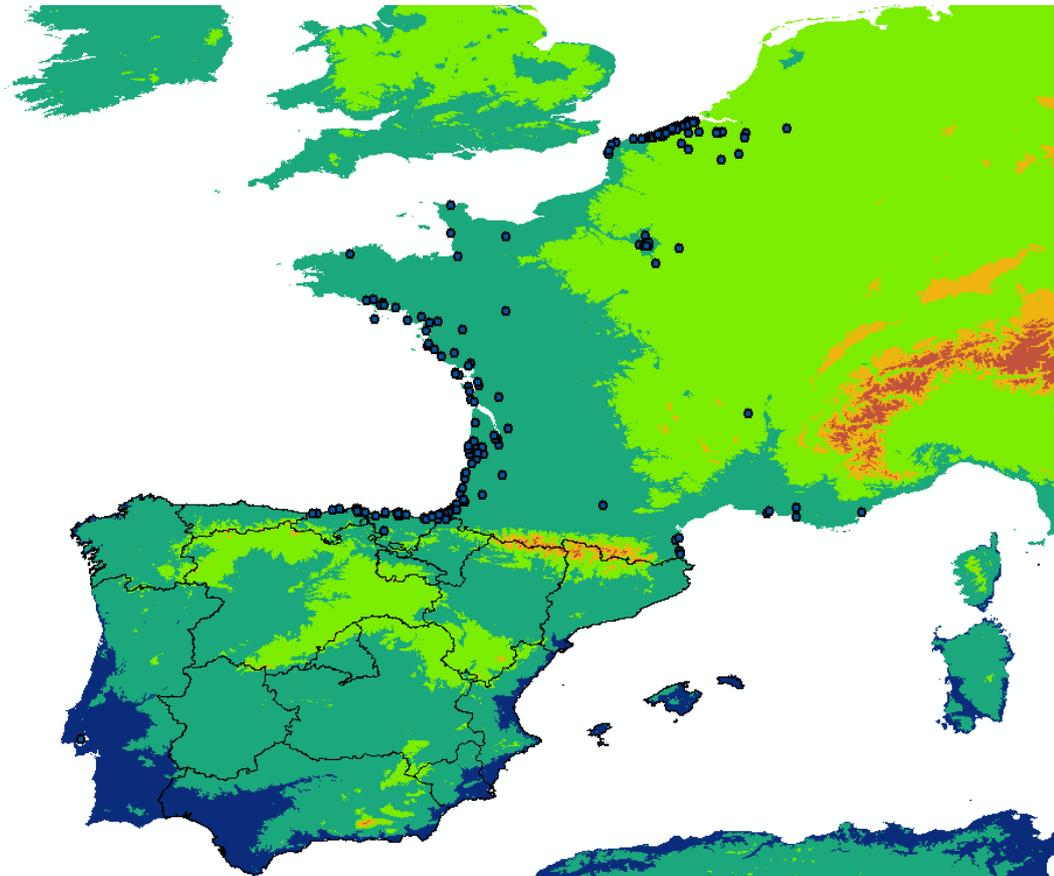


Figura 17. Distribución actual de *Baccharis halimifolia* en el Oeste de Europa (datos obtenidos de GBIF.org) sobre un mapa con siluetas de comunidades autónomas españolas.

Obtención de presencias de la especie

Tras la descarga de datos de GBIF para toda el área de estudio (**Figura 1**), se eliminaron los puntos duplicados y sin coordenadas, los que carecen de información de la incertidumbre y los que están demasiado cerca de otras presencias (distancia mínima de 1,5 km). A continuación se eligieron solamente las presencias localizadas en España, Francia y Países Bajos, obteniéndose un conjunto de 190 registros de presencia listos para introducir en el modelo (**Figura 18**).

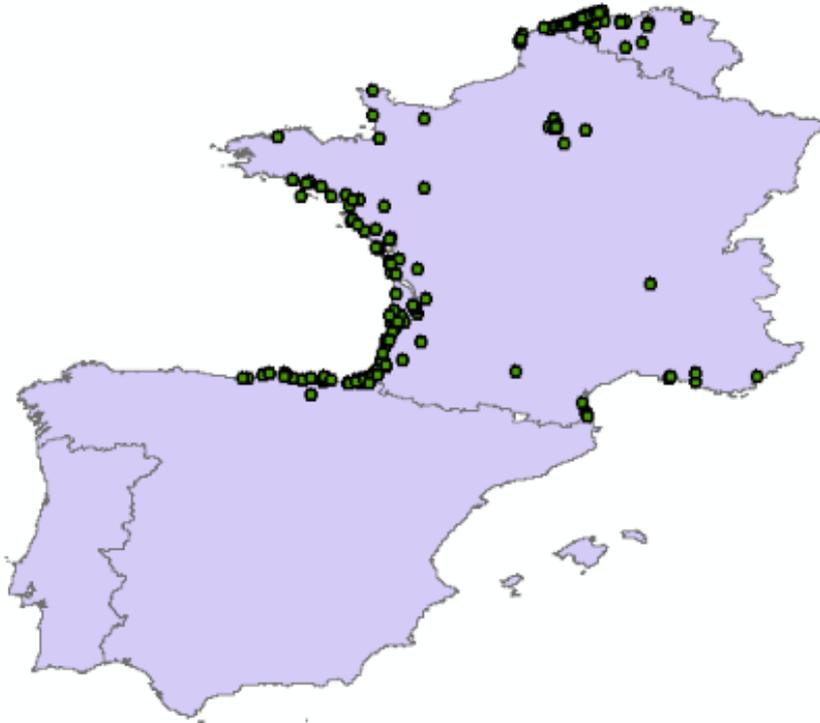


Figura 18. Registros de *Baccharis halimifolia* utilizados en la construcción del modelo. Fuente: elaboración propia a partir de los datos de GBIF.org.

Generación de pseudo-ausencias

Para garantizar la distancia tanto geográfica como ecológica a las presencias de la especie, se definió un área susceptible de contener pseudo-ausencias que cumpliera dichos requisitos. De esta forma, el área incluye todos los territorios de los países seleccionados que estén situados a más de 30 km de cualquier presencia y bajo una envuelta ambiental que no esté contenida en el 95% del intervalo de confianza del conjunto de presencias. Como *Baccharis halimifolia* tiene unos requerimientos ecológicos que hacen altamente improbable su aparición en áreas de elevada altitud en la zona de estudio, también se impuso la generación de la mitad del conjunto de pseudo-ausencias dentro de la zona situada a menos de 400 m de altitud. Esta última restricción evita la sobreestimación de la bondad de ajuste del modelo.

Para respetar la proporción adecuada respecto a los registros de presencia (190), se crearon 10 conjuntos de 570 pseudo-ausencias cada uno. Antes de introducir los datos en el algoritmo, es necesario dividirlos en datos de calibración (70%) y datos de evaluación (30%). Finalmente, se crearon 100 conjuntos de datos para calibrar y evaluar los modelos GLM y otros 100 para los modelos GAM.

Selección de variables ambientales

A partir del conjunto de 19 índices bioclimáticos y la pendiente, se fueron descartando variables correlacionadas entre sí, con alta colinealidad y con menor sentido ecológico, siempre de una en una. Las variables seleccionadas, por tanto, son aquellas que mayor sentido ecológico presentaron y que cumplían los requisitos de baja correlación entre sí (Pearson $r \leq 0.75$), baja colinealidad ($VIF < 5$) y significancia en el modelo GLM. Se trata de los seis parámetros recogidos en la **Tabla 5**. Este conjunto parece adecuado para detectar ambientes oceánicos templados con poca sequía estival, que se ajustan bien a los requerimientos conocidos de *Baccharis halimifolia*.

Tabla 5. Variables explicativas para *Baccharis halimifolia*.

Código	Nombre
Bio1	Temperatura media anual
Bio2	Rango diario de temperaturas (media mensual (max temp - min temp))
Bio4	Estacionalidad en la temperatura (desviación estándar*100)
Bio15	Estacionalidad en la precipitación (Coeficiente de variación)
Bio18	Precipitación del trimestre más cálido
Pendiente	Pendiente del terreno

2.2.2 Ajuste del MDE

El modelo generado con el algoritmo GLM y las seis variables seleccionadas (**Tabla 5**) se rodó sobre 100 combinaciones diferentes de divisiones de datos para calibración y evaluación. Para estimar la bondad de ajuste del modelo se tomaron los valores de las medianas de los estadísticos de evaluación (**Tabla 6**). El estadístico AUC se acerca mucho a la esquina superior izquierda, lo que nos indica que el modelo tiene una alta capacidad predictiva (**Figura 19**).

Tabla 6. Estadísticos de evaluación del modelo de distribución de *Baccharis halimifolia*.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo
Threshold	0.09615	0.25367	0.30859	0.32959	0.42134	0.59772
TSS	0.7890	0.8540	0.8710	0.8752	0.8950	0.9470
Sensitivity	0.8950	0.9470	0.9650	0.9616	0.9820	1.0000
Specifity	0.8130	0.8950	0.9180	0.9136	0.9315	0.9880

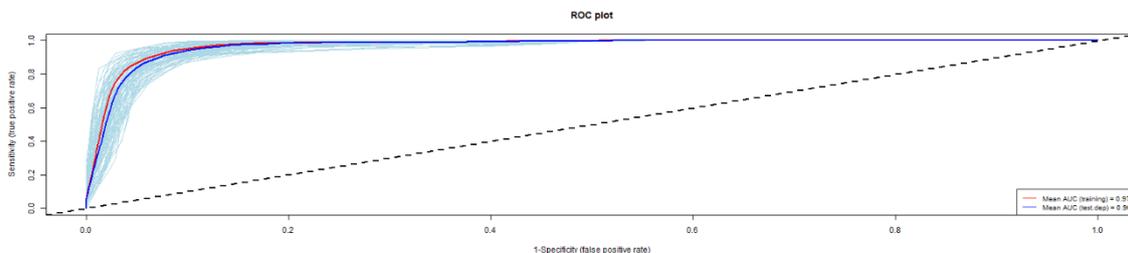


Figura 19. AUC (ROC plot) del modelo de distribución de *Baccharis halimifolia*. Fuente: elaboración propia.

Idoneidad ambiental actual

En la **Figura 20** se muestran los valores de idoneidad ambiental obtenidos para *Baccharis halimifolia* para España peninsular. Este mapa se obtiene de la proyección del modelo bajo las variables ambientales presentes calculadas de forma más precisa solo para España peninsular. En la **Figura 21**, sin embargo, se muestran los valores de idoneidad ambiental obtenidos para el Oeste de Europa. Aunque nuestro objeto de estudio está claramente centrado en la España peninsular, en este caso resulta especialmente interesante observar el comportamiento del modelo en Francia, que es el país vecino que actualmente presenta poblaciones más agresivas de esta planta.

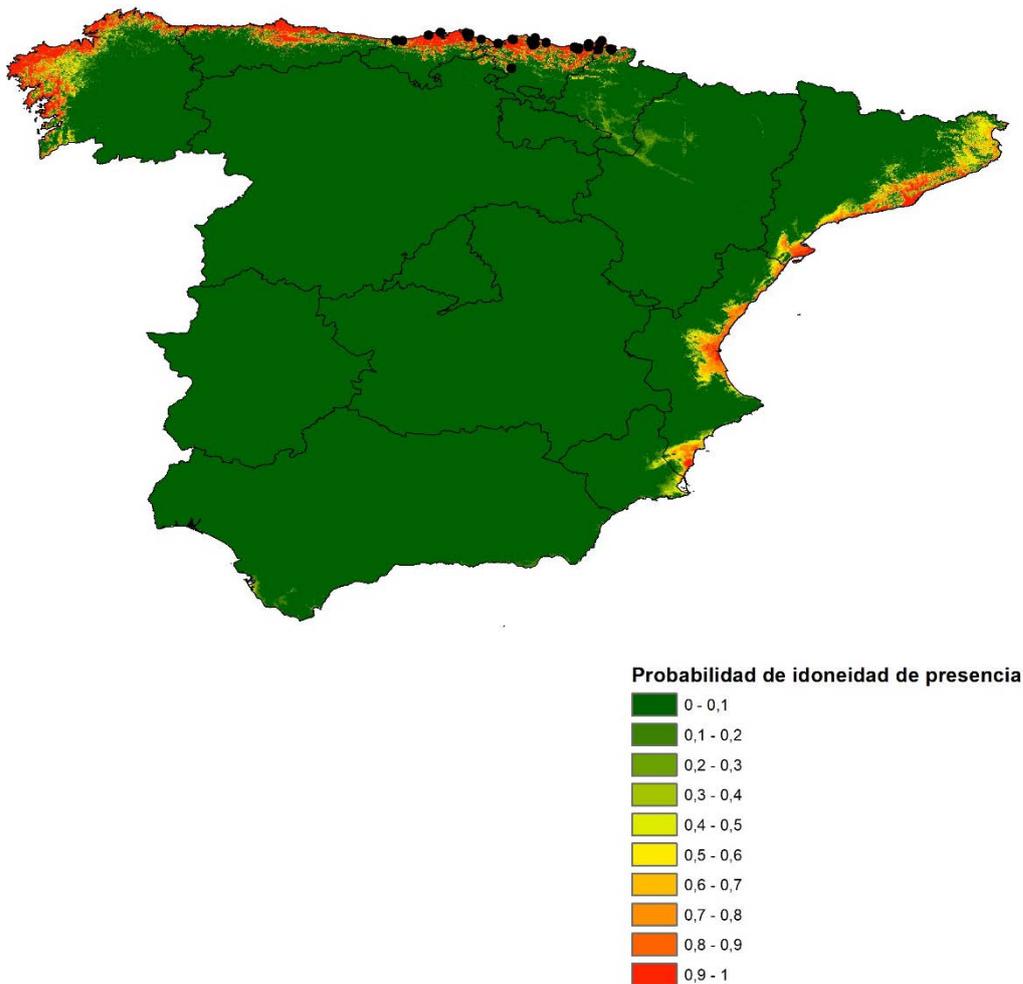


Figura 20. Idoneidad ambiental para *Baccharis halimifolia* en España peninsular. Los puntos negros representan las presencias actuales. Las líneas negras corresponden a los límites autonómicos. Fuente: elaboración propia.

Como se puede ver en la **Figura 20** *Baccharis halimifolia* no ha ocupado toda su área potencial, pues existen zonas de alta idoneidad en las que no se han registrado ocurrencias de la especie. Es el caso de Asturias, Galicia y algunas zonas de la costa mediterránea, donde se localizan áreas favorables para esta planta. Muy probablemente esto se debe a la falta de presión de propágulos o, dicho de otra manera, a que la especie no ha llegado. Sería importante vigilar los posibles

corredores que conecten las áreas actualmente colonizadas para contener la expansión antes de que puedan colonizar estas áreas actualmente libres.

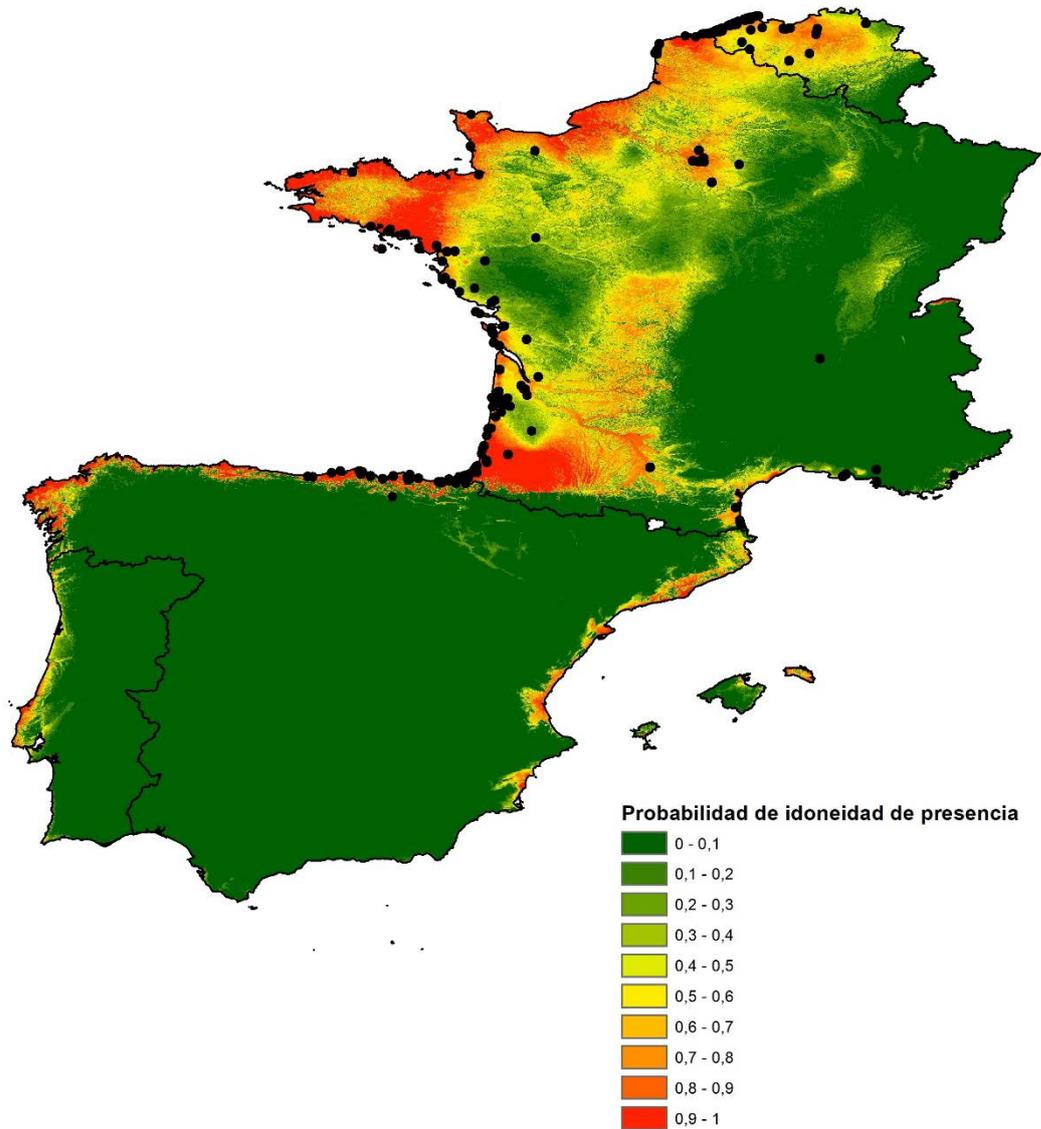


Figura 21. Idoneidad ambiental para *Baccharis halimifolia* en el Oeste de Europa. Los puntos negros representan las presencias actuales. Las líneas negras delimitan fronteras nacionales. Fuente: elaboración propia.

Según se puede apreciar en la **Figura 21**, en Francia existen grandes áreas con alta disponibilidad climática para *Baccharis halimifolia*, especialmente en la costa atlántica. Sería conveniente revisar la posibilidad de que esta especie pueda sobrevivir lejos de la costa, pues en toda la mitad oeste de Francia, incluido el sur, existen zonas climáticamente favorables para ella. La escasez de presencias en todo este territorio puede deberse a la ausencia/escasez de propágulos (como parece suceder en la costa gallega) pero resulta más probable otra hipótesis. Es muy posible que la expansión de *Baccharis halimifolia* esté ligada a algún factor ecológico más complejo. Por ejemplo, su elevada tolerancia a la salinidad tanto del suelo como del hálito marino, le otorga una ventaja competitiva en entornos costeros que no tiene en otros lugares, por lo que quizá solo logre vencer la competencia con las especies nativas en zonas costeras. Sin embargo, este

fenómeno podría cambiar si en el futuro desciende la idoneidad de las especies nativas con las que compite.

Curvas de respuesta a las variables

Las seis variables predictoras tienen significancia en el modelo, pero no todas tienen el mismo peso a la hora de explicar la distribución de los registros (**Figura 22**). Ordenadas de mayor a menor importancia en el modelo son: Bio1 (temperatura media anual), Bio15 (estacionalidad en la precipitación), Bio4 (estacionalidad en la temperatura), Bio2 (rango diario de temperaturas), pendiente (pendiente topográfica a escala 1km²) y Bio18 (precipitación del trimestre más cálido).

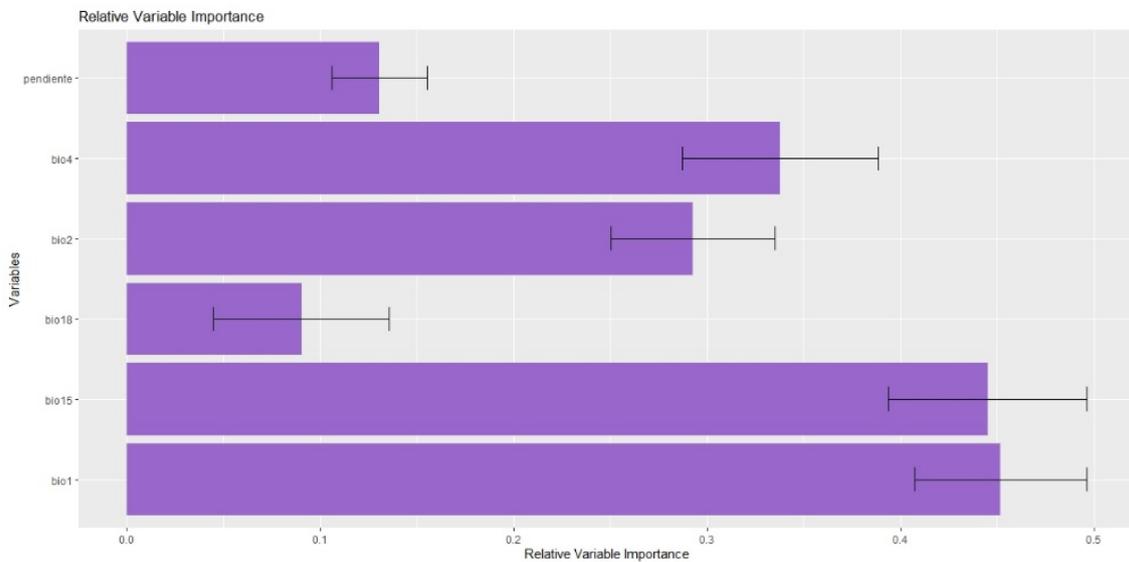


Figura 22. Importancia relativa de las variables en el modelo GLM para *Baccharis halimifolia*. Fuente: elaboración propia.

Como era de esperar, la idoneidad aumenta con la temperatura media anual (Bio1) y con la precipitación del trimestre más cálido (Bio18) y disminuye con el resto de variables (**Figura 23**). Puesto que los modelos GLM obligan a que todas las curvas de respuesta presenten una forma sigmoideal, para un análisis más detallado, es conveniente complementar estas tendencias iniciales del modelo GLM con las curvas de respuesta de un modelo GAM efectuado sobre el mismo conjunto de variables.

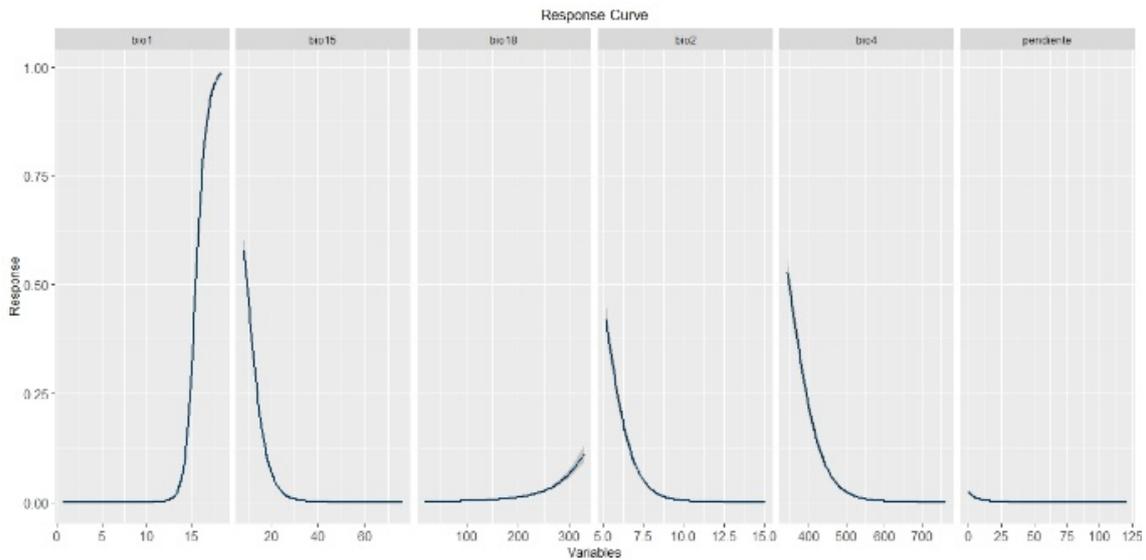


Figura 23. Curvas de respuesta a las variables ambientales en el modelo GLM para *Baccharis halimifolia*. Fuente: elaboración propia.

La importancia relativa de las variables en el modelo GAM es muy similar a la que presentan en el modelo GLM, con la única excepción de las dos variables de menor importancia, que invierten su orden: la pendiente es menos importante en el modelo GAM que la precipitación del trimestre más cálido (**Figura 24**). Las curvas de respuesta son acordes con las tendencias marcadas por el modelo GLM pero nos informan de algunos límites mejor ajustados (**Figura 25**).

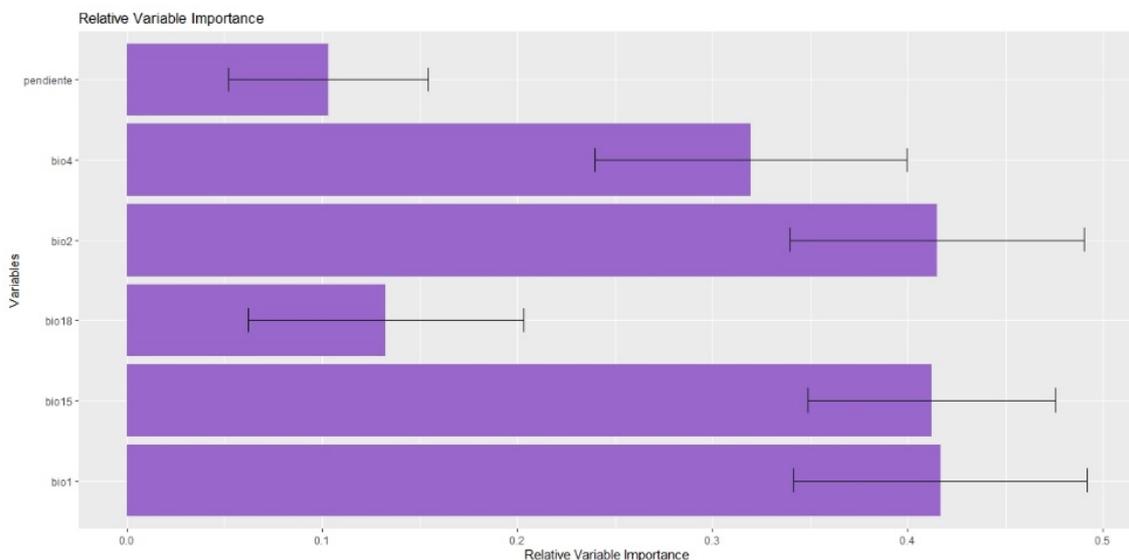


Figura 24. Importancia relativa de las variables del modelo GAM para *Baccharis halimifolia*. Fuente: elaboración propia.

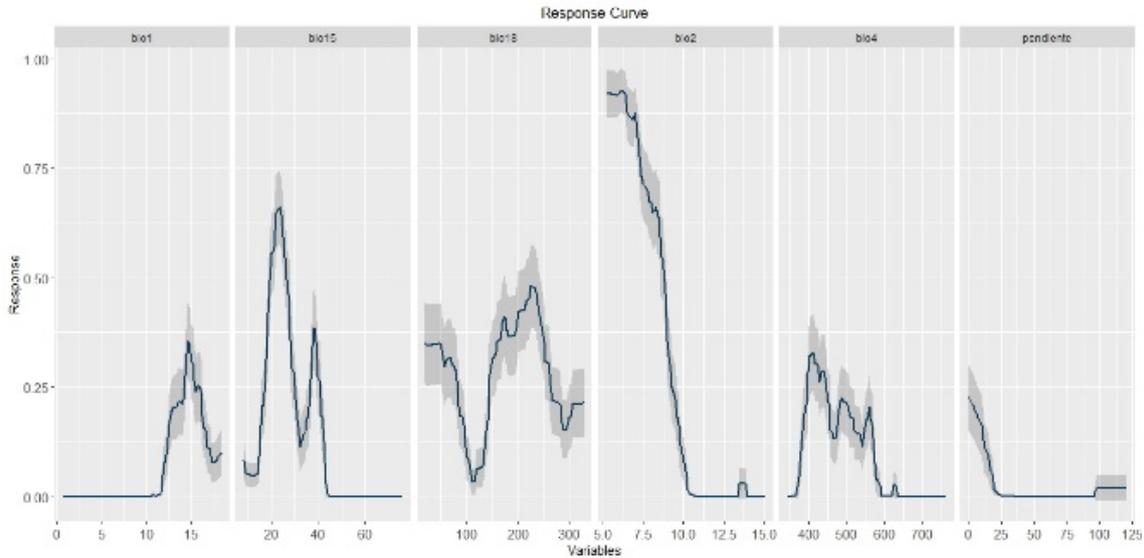


Figura 25. Curvas de respuesta a las variables ambientales en el modelo GAM para *Baccharis halimifolia*. Las zonas sombreadas en gris indican la existencia de incertidumbre en esa franja de la curva y marcan el intervalo de confianza. Fuente: elaboración propia.

2.2.3 Modelización de la distribución potencial futura:

Los cuatro escenarios futuros analizados muestran una idoneidad creciente para *Baccharis halimifolia*, especialmente en la costa mediterránea, como se puede apreciar en la **Figura 26**. El interior peninsular se mantiene claramente desfavorable para esta especie si exceptuamos el valle del Ebro.

En los dos escenarios de peores previsiones (RCP 8.5 para 2041-2060 y para 2061-2080) sí aparecen áreas de alta idoneidad en el valle del Ebro que podrían llegar a conectar, incluso la costa cantábrica con la mediterránea. Sin embargo, es necesario realizar la misma reflexión que en el caso de los territorios interiores del Suroeste de Francia en la actualidad: es posible que las interacciones bióticas jueguen un papel fundamental en la expansión de esta especie y los fenómenos de competencia logren frenar el establecimiento de *Baccharis halimifolia* lejos de los ambientes costeros. Por tanto, esta tendencia creciente de la idoneidad climática hacia el interior del valle del Ebro podría tener poca importancia. De todos modos, si la ventaja competitiva viene marcada por la salinidad, también es conveniente recordar que existen suelos salinos en el valle del Ebro, por lo que antes de descartar definitivamente la preocupación lejos de la costa, es necesario profundizar más en el conocimiento de esta especie.

El análisis de las zonas costeras requiere una atención cuidadosa. El modelo prevé un aumento muy marcado de la idoneidad en la costa mediterránea, penetrando hacia el interior en una franja anchura creciente según empeora el escenario y/o avanza el intervalo temporal, desde Cataluña hasta Almería. Sin embargo, en la costa galaico-cantábrica, se aprecia una marcada diferencia entre el oeste (Galicia), donde la idoneidad disminuye en los escenarios futuros, y la

costa cantábrica (extremo Este de Galicia, Asturias, Cantabria y Euskadi), donde se mantiene con valores muy altos que incluso crecer hacia el interior.

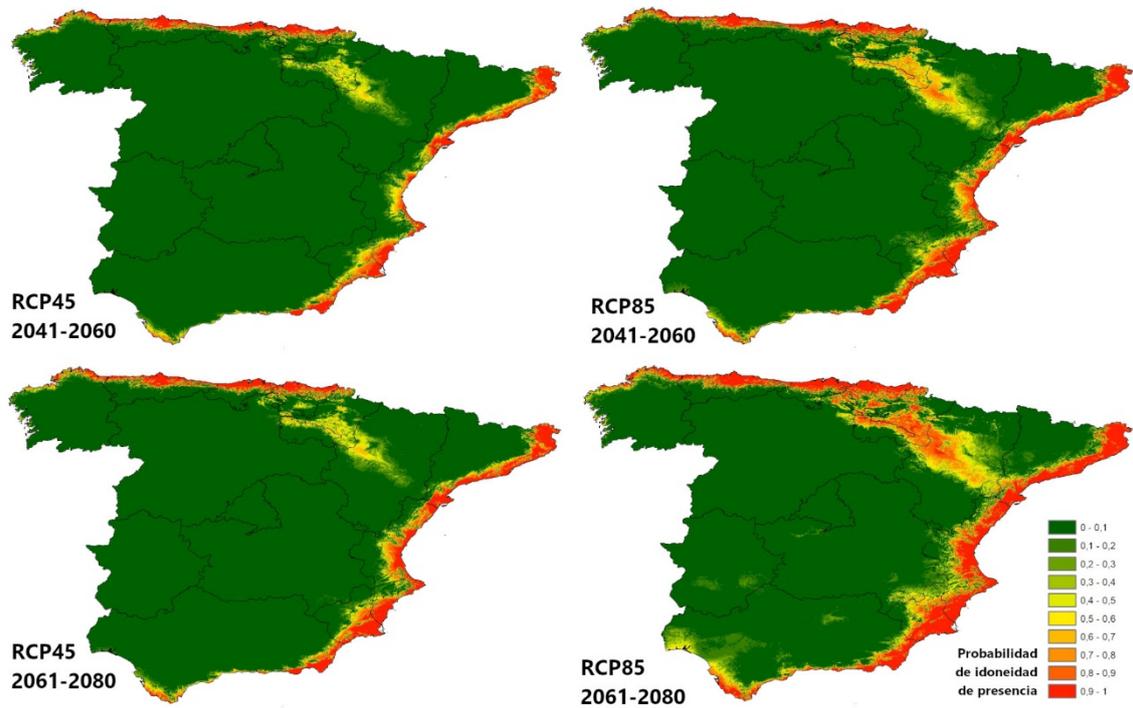


Figura 26. Idoneidad ambiental para *Baccharis halimifolia* al proyectar el modelo GLM bajo las condiciones climáticas futuras (escenarios RCP 4.5 y RCP 8.5 para los periodos 2041-2060 y 2061-2080). Fuente: elaboración propia.

2.3 *Carpobrotus acinaciformis*

2.3.1 Datos de entrada en el MDE

Debido al escaso número de presencias de *Carpobrotus acinaciformis* en la península Ibérica, se han incluido datos de varios países del Sur de Europa en la construcción del modelo, junto con los valores de índices bioclimáticos extraídos de la base datos WOLDCLIM. Se trata de una planta termófila de ambientes costeros, tanto dunares como rocosos, procedente de Sudáfrica. Actualmente está citada en todo el arco europeo mediterráneo, así como la costa atlántica andaluza, Canarias y alguna otra presencia atlántica puntual. En la **Figura 26** se presenta un mapa con la distribución actual en el Sur de Europa (datos obtenidos de GBIF.org).

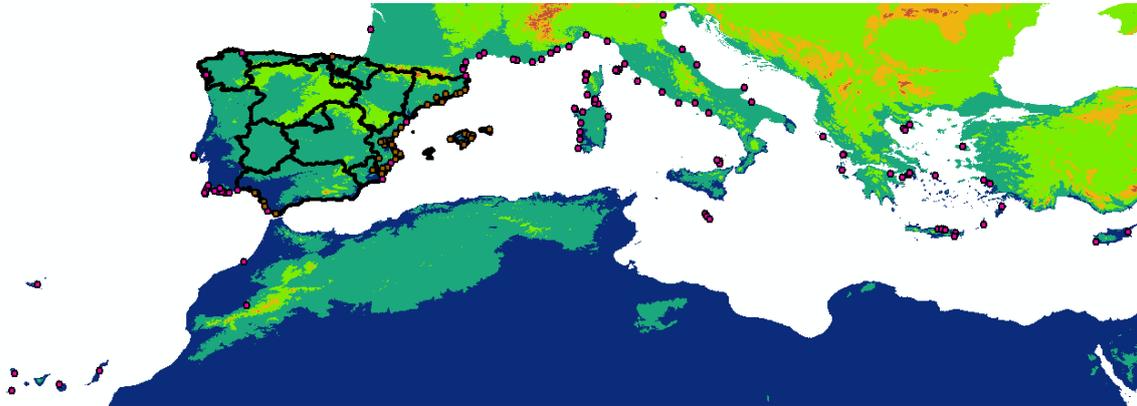


Figura 27. Distribución actual de *Carpobrotus acinaciformis* en el Sur de Europa (datos obtenidos de GBIF.org). Las líneas negras delimitan las comunidades autónomas en la España peninsular.

Obtención de presencias de la especie

Tras la descarga de datos de GBIF para toda el área de estudio (**Figura 1**), se eliminaron los puntos duplicados y sin coordenadas, los que carecen de información de la incertidumbre y los que están demasiado cerca de otras presencias (distancia mínima de 1,5 km). A continuación se eligieron solamente las presencias localizadas en España (incluyendo Canarias), Portugal (incluyendo Madeira), Francia, Italia, Malta, Grecia y Chipre, obteniéndose un conjunto de 168 registros de presencia listos para introducir en el modelo (**Figura 28**).



Figura 28. Registros de *Carpobrotus acinaciformis* utilizados en la construcción del modelo. Fuente: elaboración propia a partir de los datos de GBIF.org.

Generación de pseudo-ausencias

Para garantizar la distancia tanto geográfica como ecológica a las presencias de la especie, se definió un área susceptible de contener pseudo-ausencias que cumpliera dichos requisitos. De esta forma, el área incluye todos los territorios de los países seleccionados que estén situados a más de 30 km de cualquier presencia y bajo una envuelta ambiental que no esté contenida en el 95% del intervalo de confianza del conjunto de presencias. Como *Carpobrotus acinaciformis* tiene unos requerimientos ecológicos que hacen altamente improbable su aparición en áreas alejadas de la costa, también se impuso la generación de la mitad del conjunto de pseudo-ausencias dentro de la zona situada a 20 km como máximo de distancia a la costa. Esta última restricción evita la sobreestimación de la bondad de ajuste del modelo.

Para respetar la proporción adecuada respecto a los registros de presencia (168), se crearon 10 conjuntos de 504 pseudo-ausencias cada uno. Antes de introducir los datos en el algoritmo, es necesario dividirlos en datos de calibración (70%) y datos de evaluación (30%). Finalmente, se crearon 100 conjuntos de datos para calibrar y evaluar los modelos GLM y otros 100 para los modelos GAM.

Selección de variables ambientales

A partir del conjunto de 19 índices bioclimáticos y la pendiente, se fueron descartando variables correlacionadas entre sí, con alta colinealidad y con menor sentido ecológico, siempre de una en una. Las variables seleccionadas, por tanto, son aquellas que mayor sentido ecológico presentaron y que cumplieran los requisitos de baja correlación entre sí (Pearson $r \leq 0.75$), baja colinealidad ($VIF < 5$) y significancia en el modelo GLM. Se trata de los cinco parámetros recogidos en la **Tabla 5**. Este conjunto parece adecuado para detectar ambientes costeros termófilos de la región mediterránea en los que suele aparecer *Carpobrotus acinaciformis*.

Tabla 7. Variables explicativas para *Carpobrotus acinaciformis*.

Código	Nombre
Bio2	Rango diario de temperaturas (media mensual (max temp - min temp))
Bio11	Temperatura media del trimestre más frío
Bio15	Estacionalidad en la precipitación (Coeficiente de variación)
Bio18	Precipitación del trimestre más cálido
Pendiente	Pendiente del terreno

2.3.2 Ajuste del MDE

El modelo generado con el algoritmo GLM y las cinco variables seleccionadas (**Tabla 7**) se rodó sobre 100 combinaciones diferentes de divisiones de datos para calibración y evaluación. Para estimar la bondad de ajuste del modelo se tomaron los valores de las medianas de los estadísticos de evaluación (**Tabla 8**). El estadístico AUC se acerca mucho a la esquina superior izquierda, lo que nos indica que el modelo tiene una alta capacidad predictiva (**Figura 29**).

Tabla 8. Estadísticos de evaluación del modelo de distribución de *Carpobrotus acinaciformis*.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo
Threshold	0.1485	0.2077	0.2606	0.2705	0.3185	0.5328
TSS	0.6980	0.7830	0.8080	0.8035	0.8260	0.8880
Sensitivity	0.7760	0.8980	0.9180	0.9184	0.9390	1.0000
Specificity	0.8150	0.8640	0.8895	0.8852	0.9030	0.9540

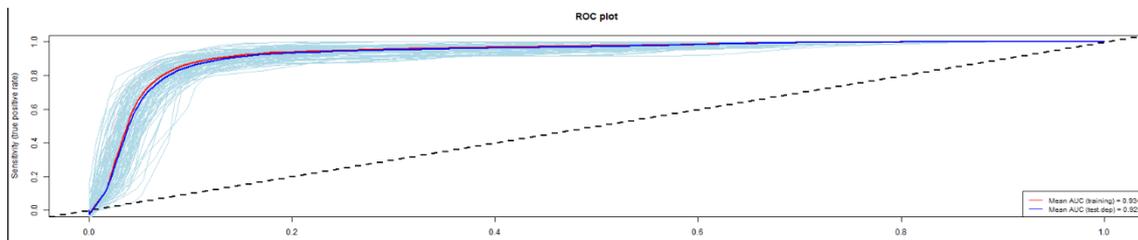


Figura 29. AUC (ROC plot) del modelo de distribución de *Carpobrotus acinaciformis*. Fuente: elaboración propia.

Idoneidad ambiental actual

En la **Figura 30** se muestran los valores de idoneidad ambiental obtenidos para *Carpobrotus acinaciformis* en España peninsular. Aunque nuestro objeto de estudio está claramente centrado en este territorio, también se muestra el mapa de idoneidad para toda la zona incluida en la construcción del modelo en la **Figura 31**.

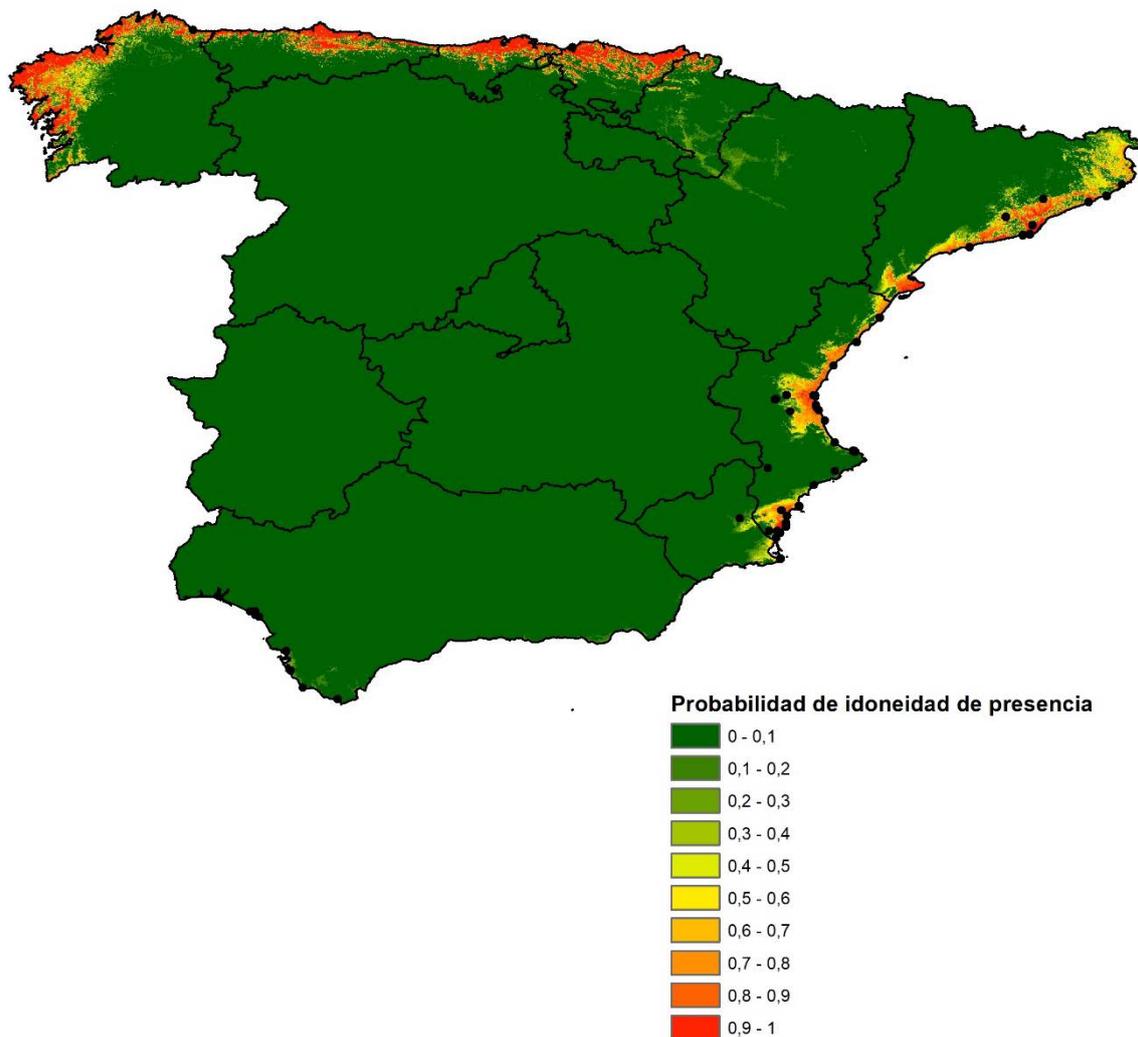


Figura 30. Idoneidad ambiental para *Carpobrotus acinaciformis* en España peninsular. Los puntos negros representan las presencias actuales. Las líneas negras corresponden a los límites autonómicos. Fuente: elaboración propia.

La proyección sobre las condiciones climáticas más precisas de España peninsular (**Figura 30**) muestra un área potencial totalmente circunscrita a la costa. Sin embargo, lejos de ceñirse a la costa mediterránea, donde ahora es más abundante, presenta altos valores de idoneidad en todo el litoral cantábrico y gallego, donde actualmente no se registran apenas presencias. Destacan también los bajos valores de idoneidad que presenta la costa andaluza atlántica, donde sí pueden encontrarse algunos ejemplares en la actualidad. Al ser una especie originaria de una región mediterránea, es posible que en la costa cantábrica encuentre algún factor limitante que no ha tenido en cuenta el modelo, pero esta discrepancia entre alta idoneidad y baja densidad actual también podría deberse, sencillamente, a la falta de presión de propágulos en esa zona, por lo que no puede descartarse la vigilancia con fines preventivos. En este sentido, es importante señalar que otra especie del mismo género (*Carpobrotus edulis*) sí logra colonizar con gran éxito ambientes costeros atlánticos de distinto tipo (tanto arenosos como rocosos), pero se trata de una especie que también en su área de origen despliega un rango ecológico más amplio que *C. acinaciformis* (Sanz-Elorza et al., 2004).

En contraste, en la **Figura 31** se aprecia que existen zonas de idoneidad media o incluso alta en la costa andaluza atlántica, así como en la costa portuguesa hasta la desembocadura del Duero aproximadamente, en la costa adriática italiana, en las costas de Sicilia y Cerdeña y, más ligeramente, en el Golfo de León. Es muy posible que la escala de representación oculte otros enclaves más locales de la franja estrictamente costera, pero lo que más llama la atención son precisamente esas áreas medias que penetran hacia el interior en lugares de baja continentalidad como los valles del Tajo y del Guadalquivir o las islas mediterráneas.

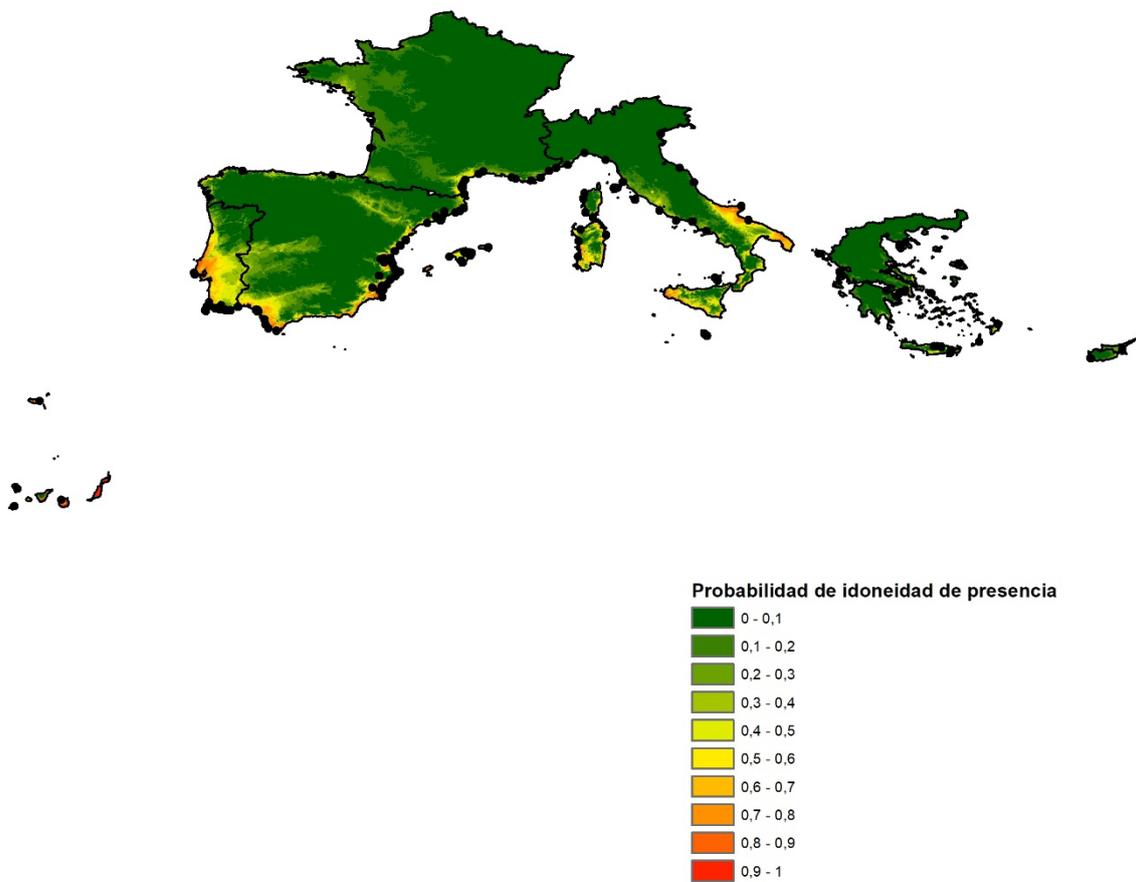


Figura 31. Idoneidad ambiental para *Carpobrotus acinaciformis* en el Sur de Europa (España, Portugal, Francia, Italia, Grecia, Malta y Chipre). Los puntos negros representan las presencias actuales. Fuente: elaboración propia.

Curvas de respuesta a las variables

Las cinco variables predictoras tienen significancia en el modelo, pero no todas tienen el mismo peso a la hora de explicar la distribución de los registros (**Figura 32**). Ordenadas de mayor a menor importancia en el modelo son: Bio11 (temperatura media del trimestre más frío), Bio18 (precipitación del trimestre más cálido), Bio15 (estacionalidad en la precipitación), Bio2 (rango diario de temperaturas) y pendiente (pendiente topográfica a escala 1km²).

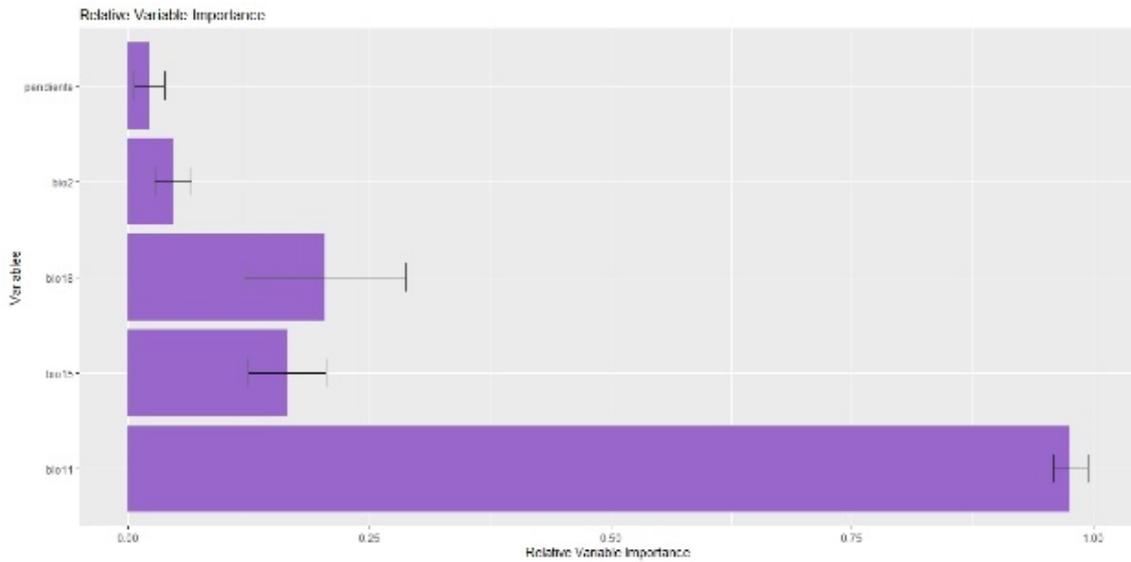


Figura 32. Importancia relativa de las variables en el modelo GLM para *Carpobrotus acinaciformis*. Fuente: elaboración propia.

La idoneidad aumenta con la temperatura media del trimestre más frío (Bio11) y con la precipitación del trimestre más cálido (Bio18) y disminuye con el resto de variables (**Figura 33**). Puesto que los modelos GLM obligan a que todas las curvas de respuesta presenten una forma sigmoidal, para un análisis más detallado, es conveniente complementar estas tendencias iniciales del modelo GLM con las curvas de respuesta de un modelo GAM efectuado sobre el mismo conjunto de variables.

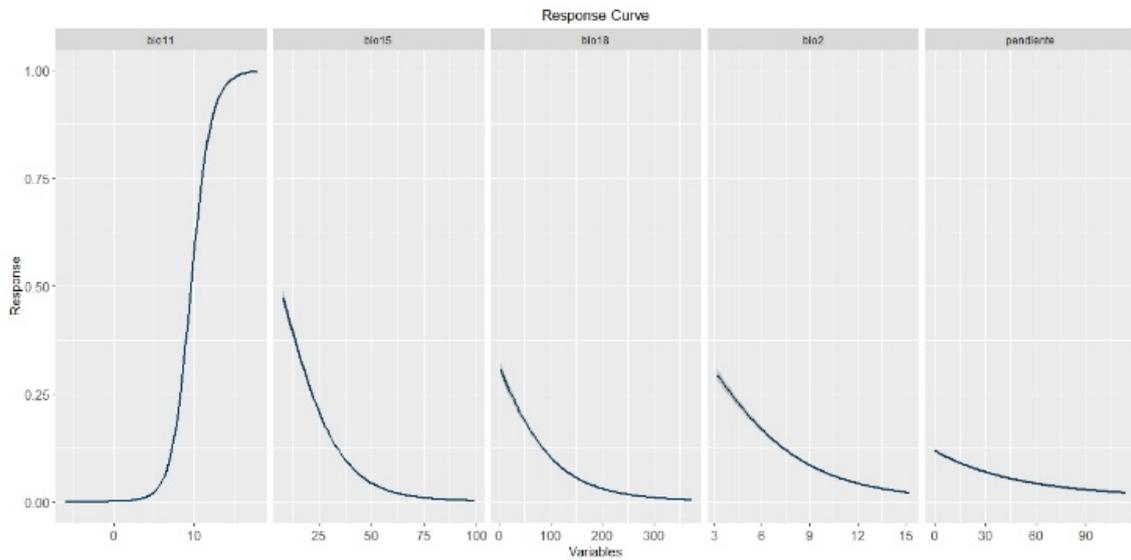


Figura 33. Curvas de respuesta a las variables ambientales en el modelo GLM para *Carpobrotus acinaciformis*. Fuente: elaboración propia.

La importancia relativa de las variables en el modelo GAM es muy similar a la que presentan en el modelo GLM, con la temperatura media del trimestre más frío (Bio11) mucho más destacada

que el resto (**Figura 34**). Las curvas de respuesta son acordes con las tendencias marcadas por el modelo GLM pero mucho más complejas (**Figura 35**).

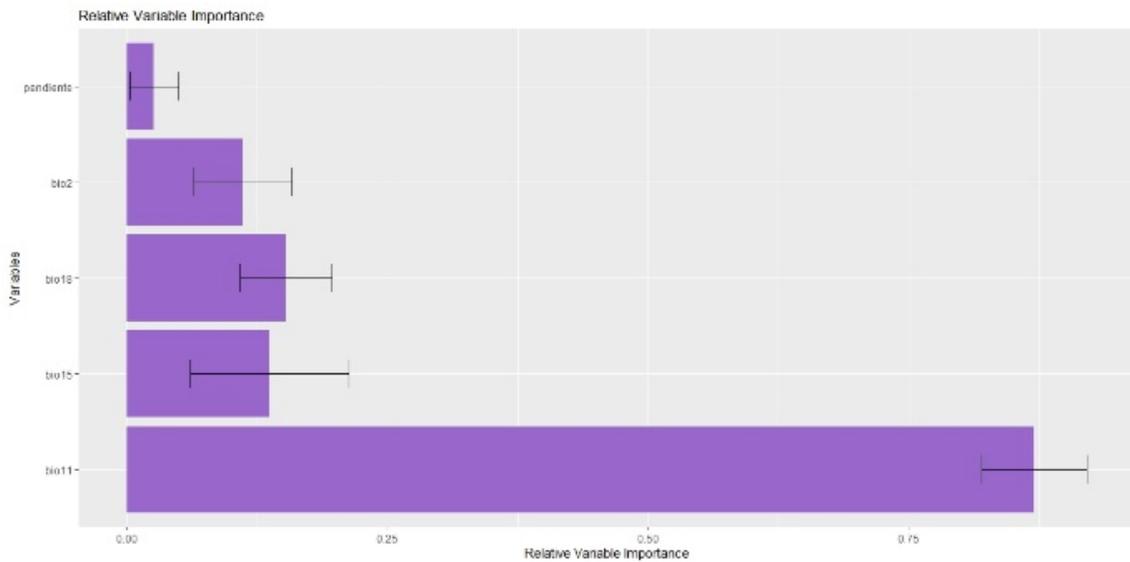


Figura 34. Importancia de las variables del modelo GAM para *Carpobrotus acinaciformis*. Fuente: elaboración propia.

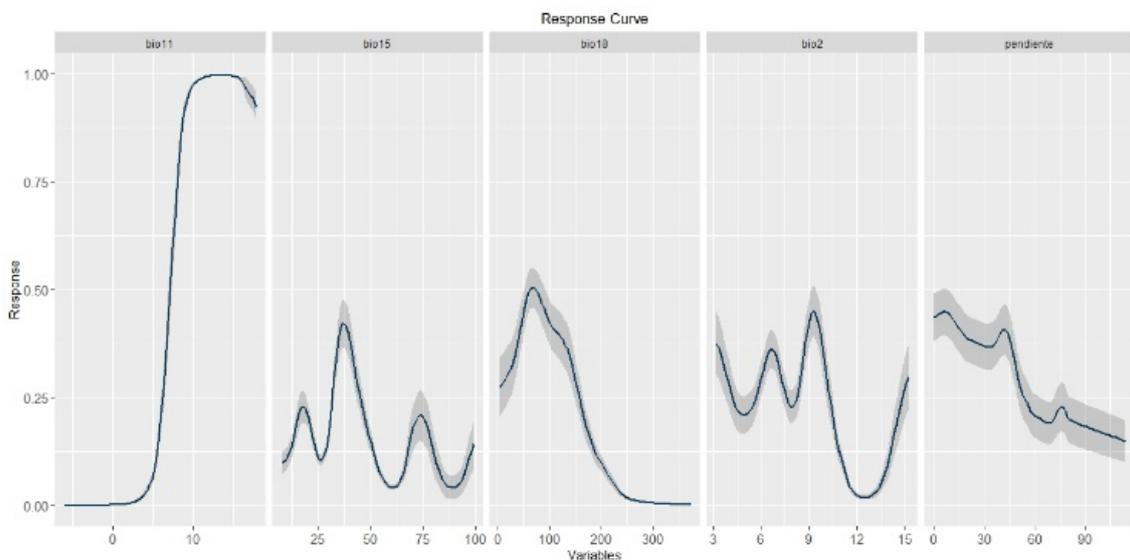


Figura 35. Curvas de respuesta a las variables ambientales en el modelo GAM para *Carpobrotus acinaciformis*. Las zonas sombreadas en gris indican la existencia de incertidumbre en esa franja de la curva y marcan el intervalo de confianza. Fuente: elaboración propia.

La idoneidad crece claramente con la temperatura media del trimestre más frío (Bio11), aunque en este caso parece estabilizarse a partir de los 12°C aproximadamente y luego comienza una zona de incertidumbre que podría ser compatible con una bajada de la idoneidad por encima de los 15°C. La precipitación del trimestre más cálido (Bio18) se muestra desfavorable por encima de los 250-300 mm, lo que podría indicar cierta afinidad por ambientes mediterráneos con sequía estival. La estacionalidad de la precipitación (Bio15) se asocia a una curva de respuesta con altibajos y zonas de incertidumbre difícil de interpretar. El rango diurno de temperaturas

(Bio2) también presenta una curva con algunos altibajos, pero se puede observar una tendencia decreciente bastante marcada, en coherencia con los ambientes costeros en los que el efecto del mar atenúa la variación térmica. La pendiente conserva un intervalo de incertidumbre a lo largo de toda la curva de respuesta, a pesar de su clara tendencia decreciente y además se trata de la variable con menor peso en el modelo (tanto GLM como GAM), bastante por detrás del resto. No es extraño si recordamos que esta planta es capaz de crecer tanto en ecosistemas dunares, con frecuencia localizados en zonas llanas, como en acantilados rocosos de fuerte pendiente.

2.3.3 Modelización de la distribución potencial futura

Los cuatro escenarios futuros contemplados en este estudio muestran un aumento de las zonas con alta idoneidad climática para *Carpobrotus acinaciformis* (Figura 36). Probablemente, este marcado aumento se debe al incremento esperado en las temperaturas invernales, pues Bio11 (temperatura media del trimestre más frío) es la variable más limitante para esta especie según el modelo propuesto.

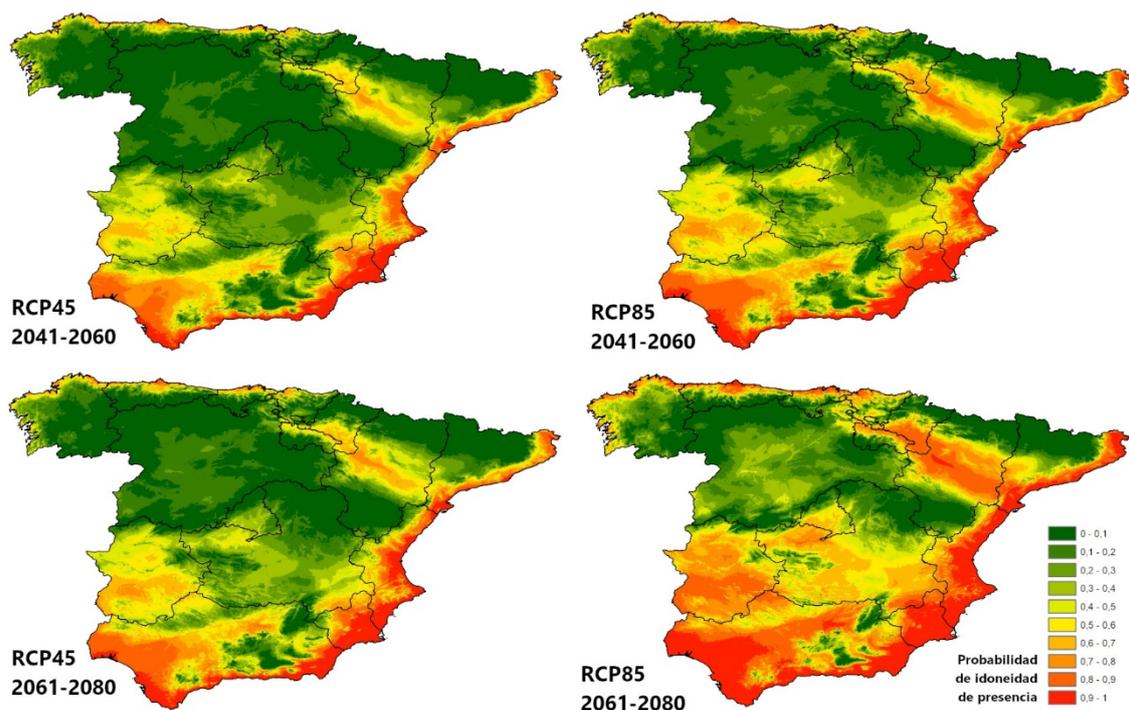


Figura 36. Idoneidad ambiental para *Carpobrotus acinaciformis* al proyectar el modelo GLM bajo las condiciones climáticas futuras (escenarios RCP 4.5 y RCP 8.5 para los periodos 2041-2060 y 2061-2080). Fuente: elaboración propia.

En los escenarios más desfavorables, la penetración hacia el interior puede ser muy amplia tanto en la costa mediterránea como en el valle del Guadalquivir. Sin embargo, sería necesario un estudio más profundo que permitiera descartar o no las zonas alejadas de la costa como zonas de alto riesgo, gracias a la incorporación de factores que no se han podido tener en cuenta en este modelo, como las interacciones bióticas o los requerimientos edáficos precisos.

La costa cantábrica no está exenta de áreas idóneas para el crecimiento de esta planta, aunque sea en una franja costera muy estrecha. Es precisamente en esta estrecha franja costera donde se dan los ecosistemas sensibles que puede llegar a transformar esta especie en caso de establecerse en ellos. Por tanto, es conveniente mantener todos los programas de vigilancia en estas zonas netamente costeras y especialmente vulnerables con el fin de evitar la posible colonización.

2.4 *Opuntia ficus-indica*

La modelización de la distribución potencial de *Opuntia ficus-indica* está todavía en fase de estudio y profundización, ya que ha resultado ser un caso especialmente complejo. Se presentan aquí los últimos datos disponibles, que son ya muy prometedores, aunque todavía no se haya concluido el trabajo.

2.4.1 Datos de entrada en el MDE

Esta especie muestra una distribución difícil de interpretar debido a su uso humano desde hace largo tiempo. A la región Europea-Mediterránea llegó en el siglo XVI como introducción voluntaria procedente de América, donde ya estaba ampliamente distribuida como cultivo agrícola precolombino. Actualmente, se sigue empleando en algunos países para usos medicinales, alimentarios y de gestión del territorio (fijación de taludes, protección y aislamiento).

Tal como se ve en la figura **Figura 37**, *Opuntia ficus-indica* presenta una distribución amplia en el área de estudio y abarca un rango ecológico también amplio (datos obtenidos de GBIF.org). Por tanto, además de las presencias ibéricas, se han tenido en cuenta también presencias de otros países europeos con el fin de recoger toda su variabilidad ecológica. Las presencias de países del Norte de África se han descartado por su bajo número en comparación con las descripciones encontradas en la bibliografía, que apuntan a una extensa distribución en dichos lugares. Por tanto, aumentaría mucho el riesgo de generar pseudo-ausencias en lugares con verdaderas presencias no recogidas en la base de datos de GBIF.

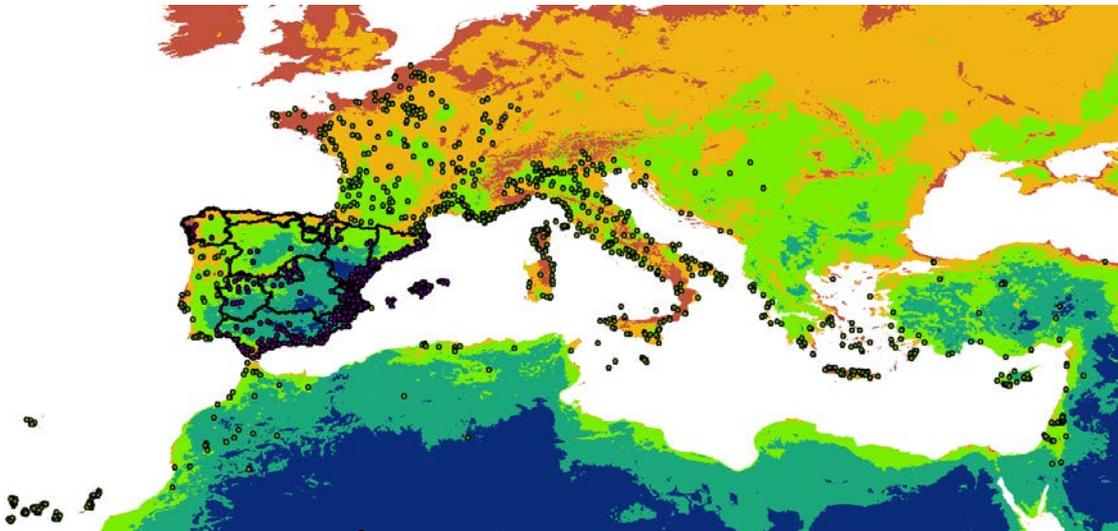


Figura 37. Distribución de *Opuntia ficus-indica* en el área de estudio (datos obtenidos de GBIF.org). Las líneas negras delimitan las comunidades autónomas en la España peninsular.

La decisión de incluir varios países europeos determina que la base de datos climática sea WORLDCLIM. Las primeras pruebas de modelización y la observación de la distribución de los datos hizo sospechar del origen de algunos de ellos, especialmente en Francia, donde parece poco creíble que esta especie tenga capacidad para invadir tantos enclaves de marcado carácter

eurosiberiano. Con el fin de mejorar los datos de entrada en el modelo, se ha añadido otra fuente de información que permita filtrar los registros de presencias en función de la densidad de población humana. Se trata del mapa de densidad de población de GEOSTAT 2011 grid dataset (EUROSTAT).

Obtención de presencias de la especie

Tras la descarga de datos de GBIF para toda el área de estudio (**Figura 1**), se eliminaron los puntos duplicados y sin coordenadas, los que carecen de información de la incertidumbre y los que están demasiado cerca de otras presencias (distancia mínima de 1,5 km). A continuación se eligieron solamente las presencias localizadas en la zona continental y/o mediterránea de Portugal, España, Francia e Italia, obteniéndose un conjunto de 1791 registros de presencia listos para introducir en los modelos.

Tras la observación del resultado de la modelización preliminar, se decidió investigar la influencia de las presencias dudosas en la construcción del modelo, bajo la hipótesis de que aquellas ligadas a enclaves con mayor densidad humana podrían estar sesgando el modelo. Para ello, se establecieron diferentes conjuntos de datos de entrada y se construyeron modelos con todos ellos. Posteriormente, se compararon dichos modelos para detectar diferencias y definir posibles mejoras aplicables no solo en este caso sino también en otras especies invasoras.

El filtro utilizado para definir los diferentes conjuntos de datos de presencia fue el mapa de densidad de población de GEOSTAT 2011 grid dataset (EUROSTAT) (**Figura 38**). Se trata de una cartografía con resolución de 1 km² que puede incorporarse fácilmente a las herramientas de análisis geográfico utilizadas en este trabajo (ArcMap 10.1). Para establecer las zonas de diferente densidad de población humana, el presente trabajo se apoya directamente en las categorías ya establecidas en dicha cartografía: lugares con densidad de población menor de 15 habitantes/km², entre 15 y 50 habitantes/km², entre 50 y 100 habitantes/km², entre 100 y 500 habitantes/km² y mayor de 500 habitantes/km². Los conjuntos de datos (datasets) a introducir en el algoritmo serán, por tanto, los definidos en la **Tabla 9**. Como puede verse, incluso en el caso más restrictivo el número de presencias supera los 700 registros.

Tabla 9. Conjuntos de datos de entrada según la densidad de población humana asociada al lugar donde se sitúa cada presencia de *Opuntia ficus-indica* (utilizando las categorías establecidas en el mapa de densidad de población de GEOSTAT 2011 grid dataset, EUROSTAT).

Nombre del dataset	Descripción	Nº de presencias
SDM_GLM_15	Presencias en lugares con menos de 15 hab/km ²	741
SDM_GLM_50	Presencias en lugares con menos de 50 hab/km ²	886
SDM_GLM_100	Presencias en lugares con menos de 100 hab/km ²	1000
SDM_GLM_500	Presencias en lugares con menos de 500 hab/km ²	1263
SDM_GLM_mas500	Presencias (todas)	1791

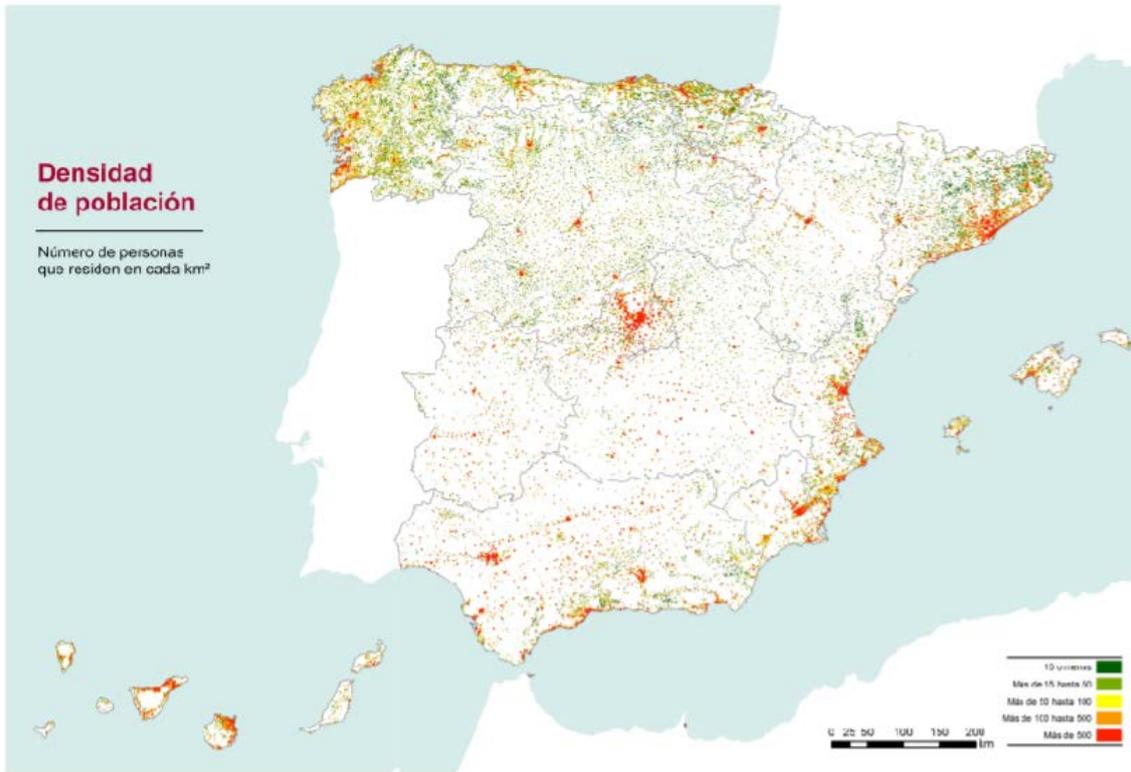


Figura 38. Mapa de densidad de población de GEOSTAT 2011 grid dataset (EUROSTAT) para España. Fuente: EUROSTAT <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography>.

Generación de pseudo-ausencias

Con el objetivo de garantizar la distancia tanto geográfica como ecológica a las presencias de la especie, se definió un área susceptible de contener pseudo-ausencias que cumpliera dichos requisitos. De esta forma, el área incluye todos los territorios de los países seleccionados que estén situados a más de 30 km de cualquier presencia y bajo una envuelta ambiental que no esté contenida en el 95% del intervalo de confianza del conjunto de presencias. Dentro de esta área común se crearon los conjuntos de pseudo-ausencias asociados a cada set de presencias, con el objetivo de que en los casos más restrictivos no cayera ninguna pseudo-ausencia en puntos donde pudiese estar alguna de las presencias descartadas.

Para respetar la proporción adecuada respecto a los registros de presencia (741, 886, 1000, 1263 o 1791, según cada dataset), se crearon 10 conjuntos del triple de pseudo-ausencias en cada caso. Antes de introducir los datos en el algoritmo, es necesario dividirlos en datos de calibración (70%) y datos de evaluación (30%). Finalmente, se crearon 100 conjuntos de datos para calibrar y evaluar cada modelo con el algoritmo GLM.

Selección de variables ambientales

A partir del conjunto de 19 índices bioclimáticos y la pendiente, se fueron descartando variables correlacionadas entre sí, con alta colinealidad y con menor sentido ecológico, siempre de una en una. Las variables seleccionadas, por tanto, son aquellas que mayor sentido ecológico presentaron y que cumplían los requisitos de baja correlación entre sí (Pearson $r \leq 0.75$), baja colinealidad (VIF < 5) y significancia en el modelo GLM. Se trata de los cinco parámetros recogidos en la **Tabla 10**.

Tabla 10. Variables explicativas para *Opuntia ficus-indica*.

Código	Nombre
Bio2	Rango diario de temperaturas (media mensual (max temp - min temp))
Bio8	Temperatura media del trimestre más húmedo
Bio11	Temperatura media del trimestre más frío
Bio12	Precipitación anual
pendiente	Pendiente del terreno

Según la bibliografía consultada, esta especie tolera mal las bajas temperaturas, especialmente cuando coinciden con un nivel alto de humedad en sus tejidos, mientras que muestra algo más de resistencia al frío cuando coincide con un periodo prolongado de sequía que vacía de contenido acuoso los tejidos de reserva. También es muy sensible al encharcamiento, presentando cierta amplitud en el rango de precipitaciones anuales soportadas en función de las características del suelo. Es posible que la pendiente favorezca el drenaje y favorezca de este modo la supervivencia de esta planta. Su resistencia a la continentalidad es bien conocida, soportando temperaturas relativamente bajas junto a un límite superior muy elevado que apenas llega a alcanzarse en el área de estudio. Por tanto, las cinco variables seleccionadas parecen describir muy bien algunos de los factores ya descritos en publicaciones previas.

2.4.2 Ajuste de los MDE

El modelo generado con el algoritmo GLM y las cinco variables seleccionadas (**Tabla 10**) se rodó sobre 10 combinaciones diferentes de divisiones de datos para calibración y evaluación en cada dataset (**Tabla 9**). Para estimar la bondad de ajuste del modelo se tomaron los valores de las medianas de los estadísticos de evaluación. Al tener diferente número de presencias, el AUC no es directamente comparable, pero también se aportan los gráficos correspondientes (**Figura 39**, **Figura 40**, **Figura 41**, **Figura 42** y **Figura 43**).

El modelo construido con todas las presencias no presenta unos estadísticos de evaluación demasiado buenos, como ya se había detectado en el análisis preliminar (**Tabla 11**). Los otros cuatro modelos presentan valores aceptables, resultando el modelo construido con el dataset SDM_GLM_15 (el más restrictivo) es el que muestra mejores resultados en los estadísticos de evaluación (**Tabla 12**).

Tabla 11. Estadísticos de evaluación de los modelos de distribución de *Opuntia ficus-indica* construidos con todas las presencias (dataset SDM_GLM_mas500), con las presencias en zonas de menos de 500 habitantes/km² (dataset SDM_GLM_500), con las presencias en zonas de menos de 100 habitantes/km² (dataset SDM_GLM_100), con las presencias en zonas de menos de 50 habitantes/km² (dataset SDM_GLM_50) y con las presencias en zonas de menos de 15 habitantes/km² (dataset SDM_GLM_15).

		Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo
SDM_GLM_mas500	Threshold	0.1428	0.1849	0.2087	0.2087	0.2231	0.3368
	TSS	0.7520	0.7780	0.7880	0.7877	0.7980	0.8180
	Sensitivity	0.8420	0.8890	0.9030	0.9018	0.9140	0.9460
	Specificity	0.8520	0.8765	0.8880	0.8862	0.8972	0.9270
SDM_GLM_500	Threshold	0.2003	0.2171	0.2297	0.2342	0.2440	0.3332
	TSS	0.8070	0.8320	0.8430	0.8417	0.8520	0.8850
	Sensitivity	0.8670	0.8932	0.9040	0.9030	0.9120	0.9440
	Specificity	0.9200	0.9330	0.9385	0.9389	0.9453	0.9620
SDM_GLM_100	Threshold	0.1471	0.2500	0.2907	0.2958	0.3457	0.4746
	TSS	0.8290	0.8638	0.8690	0.8715	0.8825	0.9040
	Sensitivity	0.8690	0.8990	0.9130	0.9107	0.9230	0.9560
	Specificity	0.9230	0.9560	0.9610	0.9609	0.9680	0.9830
SDM_GLM_50	Threshold	0.1963	0.2616	0.2978	0.2992	0.3356	0.4713
	TSS	0.8580	0.8840	0.8940	0.8936	0.9030	0.9280
	Sensitivity	0.9020	0.9200	0.9280	0.9298	0.9390	0.9620
	Specificity	0.9450	0.9590	0.9640	0.9638	0.9692	0.9860
SDM_GLM_15	Threshold	0.1412	0.2586	0.2676	0.3071	0.3480	0.4709
	TSS	0.8730	0.9020	0.9120	0.9123	0.9220	0.9430
	Sensitivity	0.9050	0.9320	0.9460	0.9453	0.9560	0.9770
	Specificity	0.9360	0.9610	0.9670	0.9672	0.9730	0.9900

Tabla 12. Resumen comparativo de los evaluadores de bondad de ajuste más importantes.

Dataset	TSS (mediana)	Sensibilidad (mediana)	AUC (mediana)
SDM_GLM_15	0,91	0,95	0,98
SDM_GLM_50	0,89	0,93	0,98
SDM_GLM_100	0,87	0,91	0,97
SDM_GLM_500	0,84	0,9	0,97
SDM_GLM_mas500	0,79	0,9	0,96

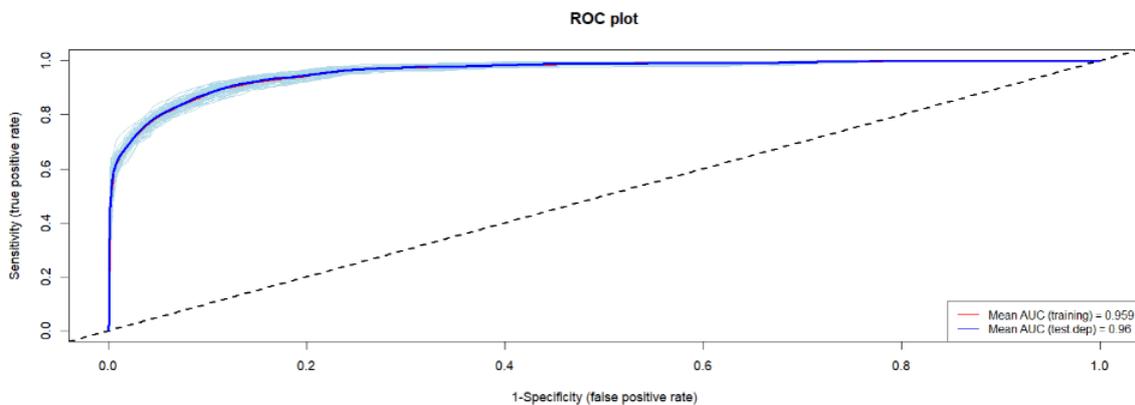


Figura 39. AUC (ROC plot) del modelo de distribución de *Opuntia ficus-indica* construido con todas las presencias (dataset SDM_GLM_mas500). Fuente: elaboración propia.

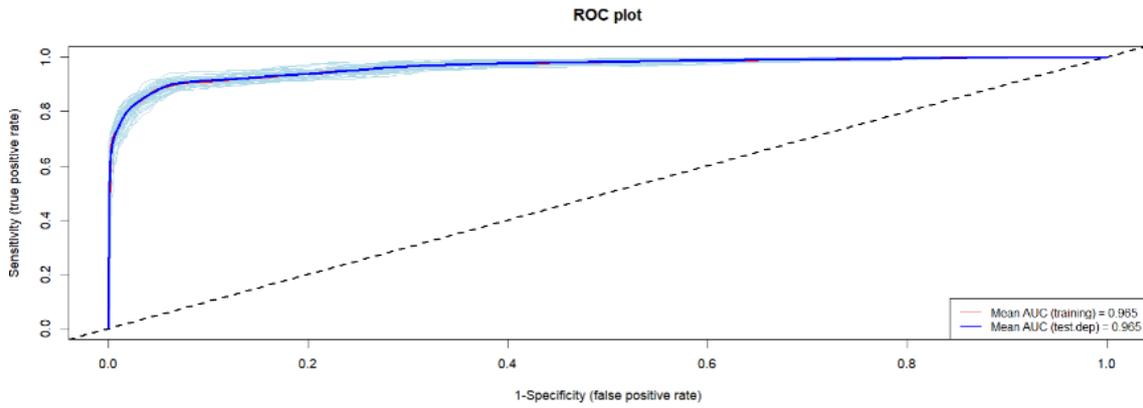


Figura 40. AUC (ROC plot) del modelo de distribución de *Opuntia ficus-indica* construido con las presencias en zonas de densidad de población menor de 500 habitantes/km² (dataset SDM_GLM_500). Fuente: elaboración propia.

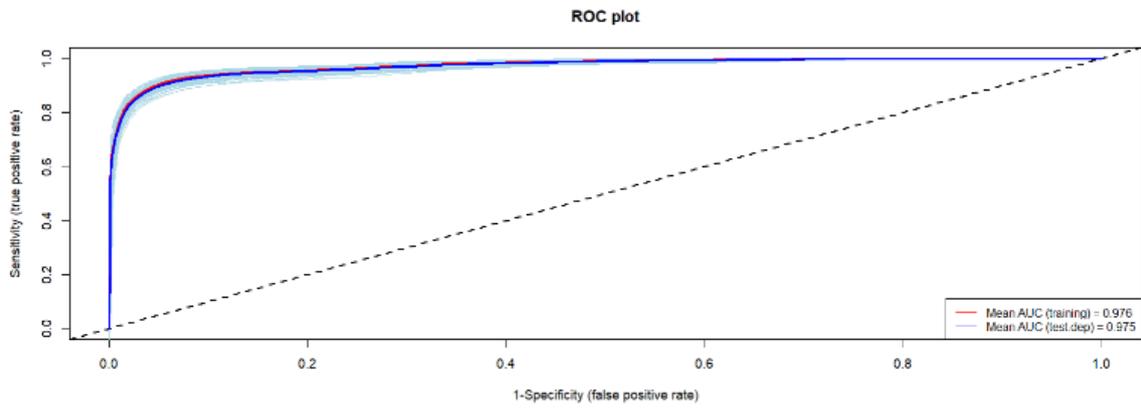


Figura 41. AUC (ROC plot) del modelo de distribución de *Opuntia ficus-indica* construido con las presencias en zonas de densidad de población menor de 100 habitantes/km² (dataset SDM_GLM_100). Fuente: elaboración propia.

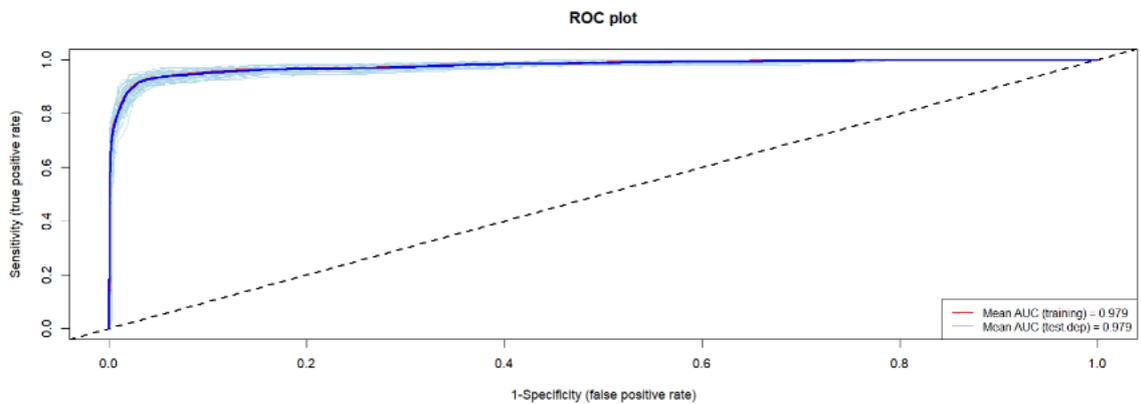


Figura 42. AUC (ROC plot) del modelo de distribución de *Opuntia ficus-indica* construido con las presencias en zonas de densidad de población menor de 50 habitantes/km² (dataset SDM_GLM_50). Fuente: elaboración propia.

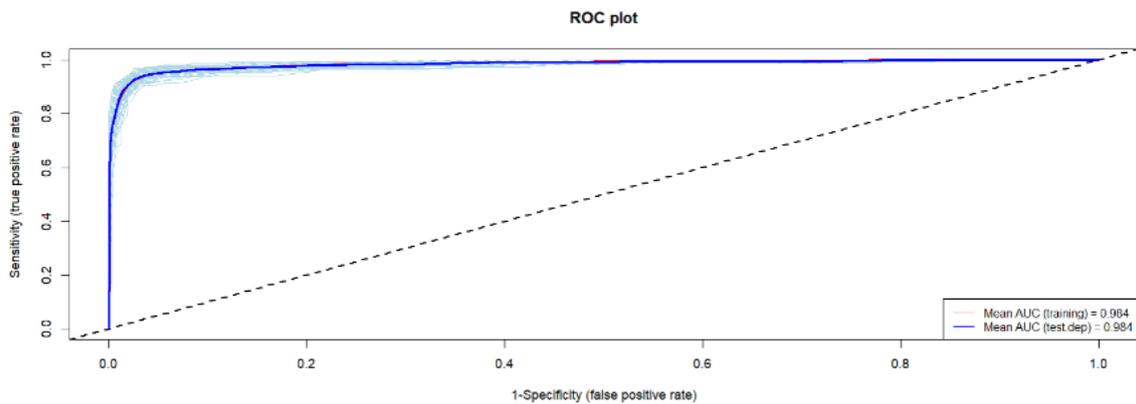


Figura 43. AUC (ROC plot) del modelo de distribución de *Opuntia ficus-indica* construido con las presencias en zonas de densidad de población menor de 15 habitantes/km² (dataset SDM_GLM_15). Fuente: elaboración propia.

Idoneidad ambiental actual

Cada uno de los modelos construidos proporciona un mapa de idoneidad ambiental sobre la zona de estudio (**Figura 44**, **Figura 45**, **Figura 46**, **Figura 47** y **Figura 48**). Comparando los dos extremos, el modelo hecho con todas las presencias y el más restrictivo que solo contiene presencias de zonas con menos de 15 habitantes/km², se observan diferencias importantes, especialmente en la región eurosiberiana. La eliminación de presencias dudosas parece mejorar el modelo y, al mismo tiempo, reducir el área de mayor idoneidad. En todos los casos, las zonas más favorables de la península Ibérica se presentan en la costa mediterránea y la mitad sur de la costa atlántica, con fuertes incursiones hacia el interior por los valles del Tajo, del Guadiana, del Guadalquivir y del Ebro. A escala europea, se aprecian valores elevados por toda la costa mediterránea con especial preocupación en las islas (Baleares, Sicilia, Cerdeña y Córcega).

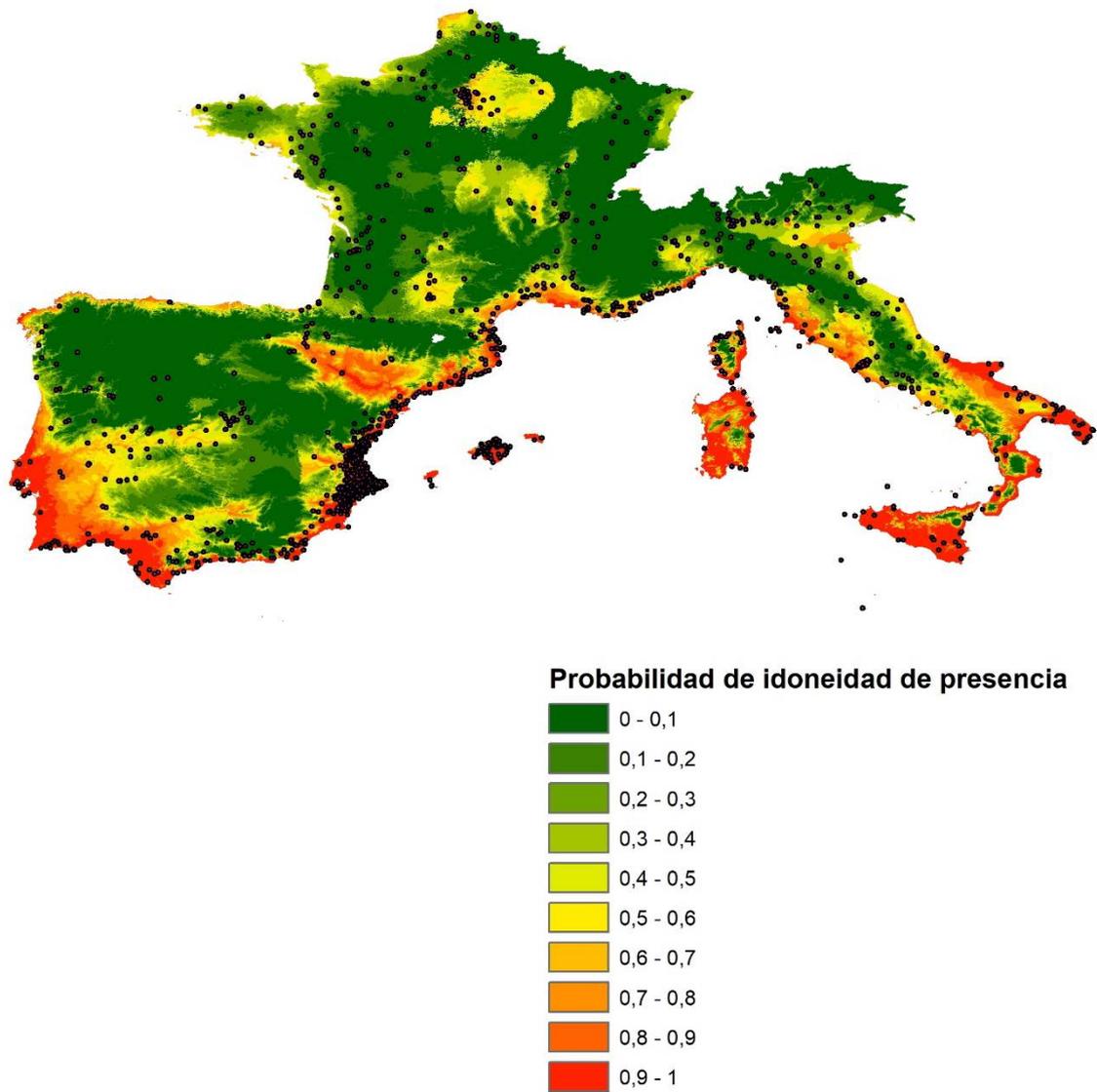


Figura 44. Idoneidad ambiental para *Opuntia ficus-indica* en el Suroeste de Europa (España, Portugal, Francia e Italia) según el modelo construido con todas las presencias de la especie (dataset SDM_GLM_mas500). Los puntos negros representan las presencias actuales incluidas en el modelo. Fuente: elaboración propia.

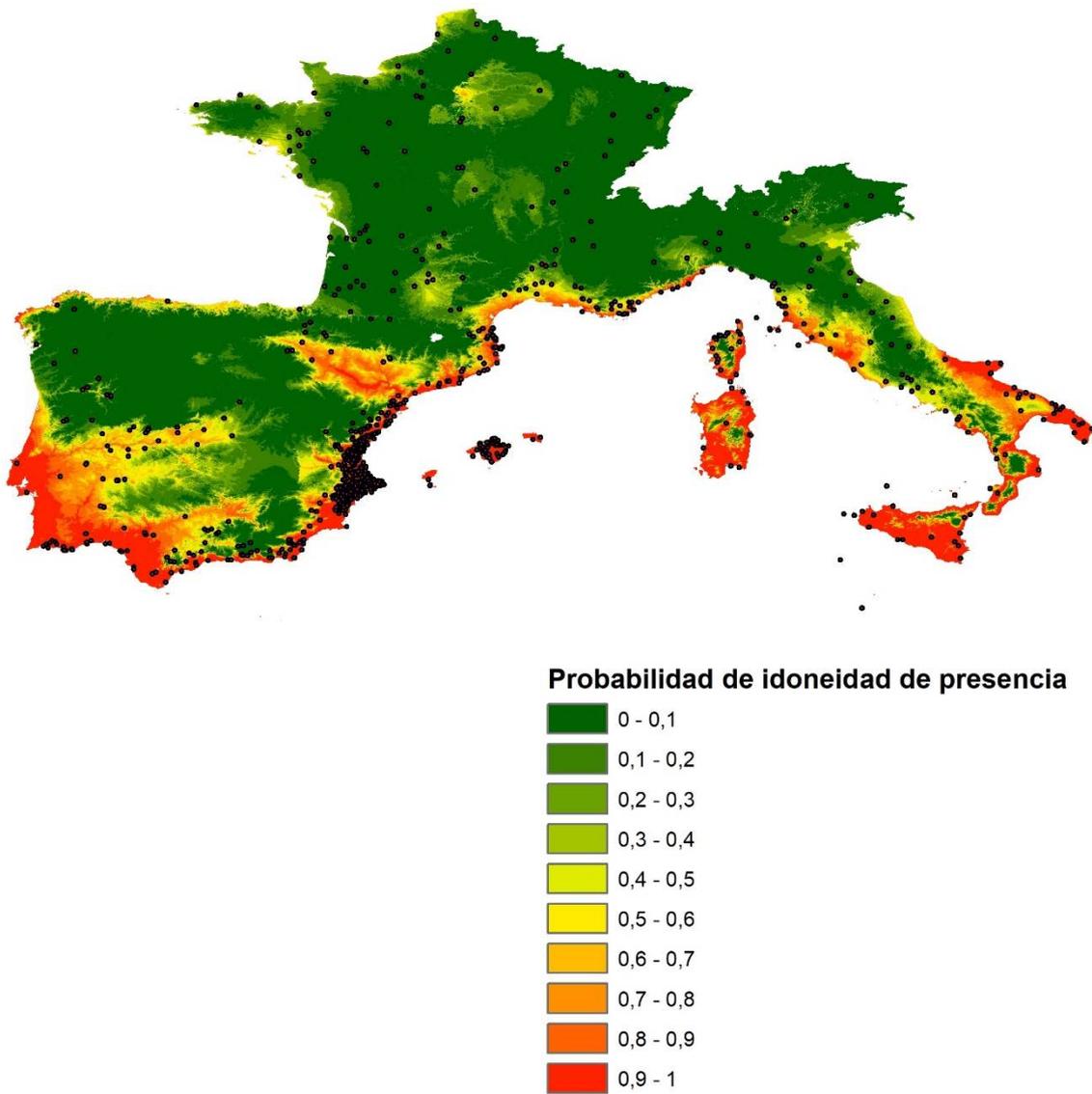


Figura 45. Idoneidad ambiental para *Opuntia ficus-indica* en el Suroeste de Europa (España, Portugal, Francia e Italia) según el modelo construido con las presencias de zonas con menos de 500 habitantes/km² (dataset SDM_GLM_500). Los puntos negros representan las presencias actuales incluidas en el modelo. Fuente: elaboración propia.

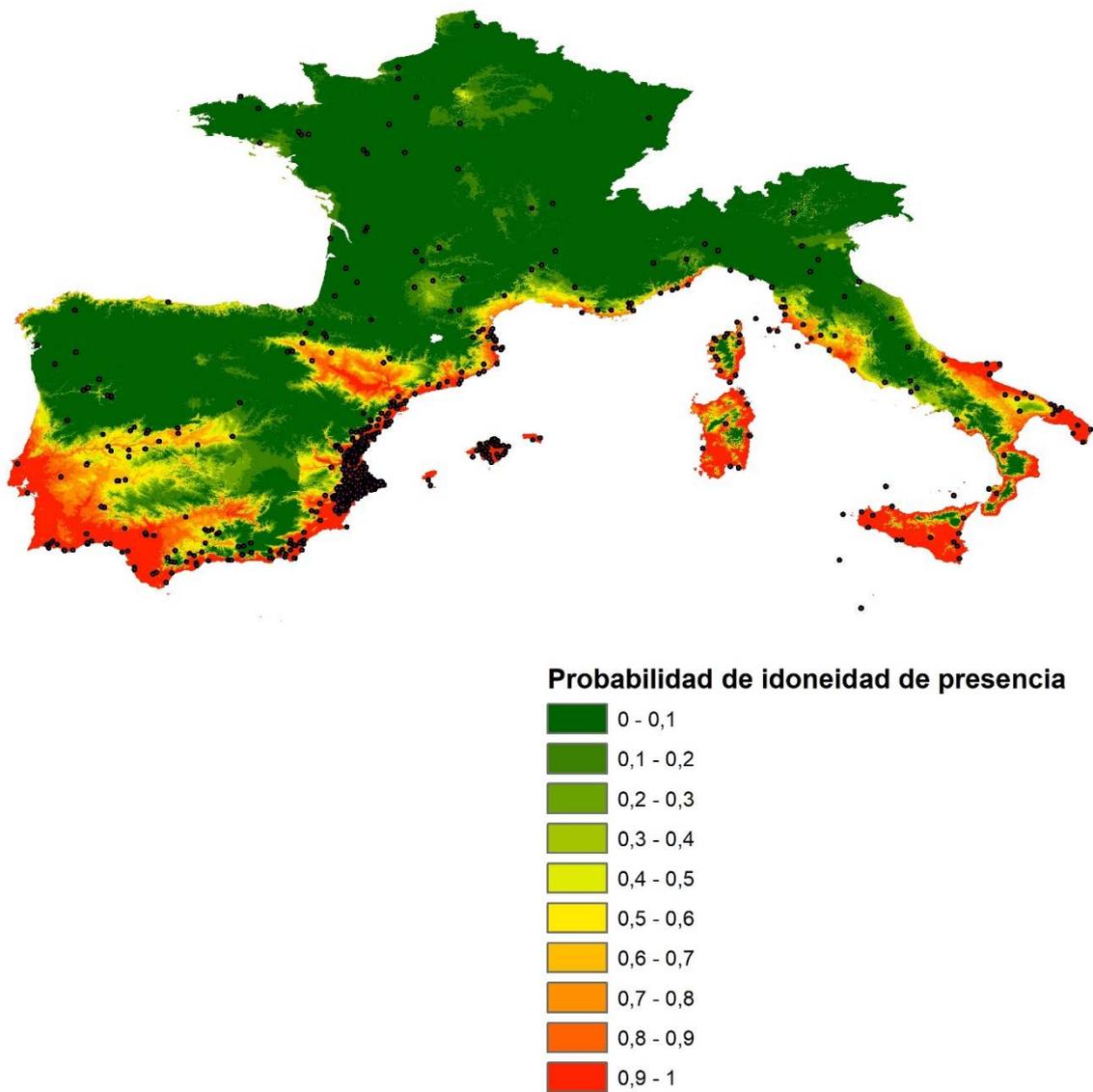


Figura 46. Idoneidad ambiental para *Opuntia ficus-indica* en el Suroeste de Europa (España, Portugal, Francia e Italia) según el modelo construido con las presencias de zonas con menos de 100 habitantes/km² (dataset SDM_GLM_100). Los puntos negros representan las presencias actuales incluidas en el modelo. Fuente: elaboración propia.

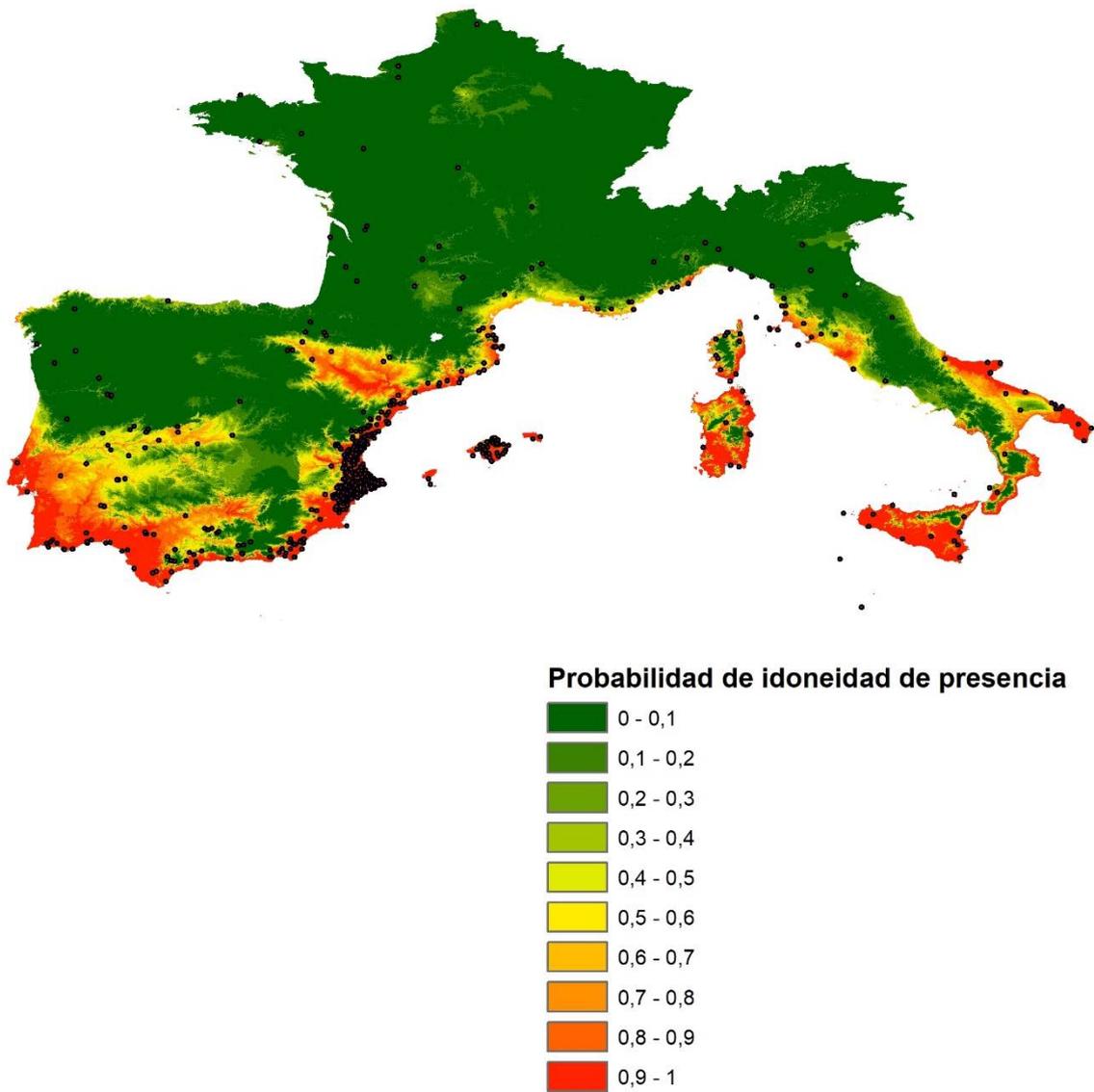


Figura 47. Idoneidad ambiental para *Opuntia ficus-indica* en el Suroeste de Europa (España, Portugal, Francia e Italia) según el modelo construido con las presencias de zonas con menos de 50 habitantes/km² (dataset SDM_GLM_50). Los puntos negros representan las presencias actuales incluidas en el modelo. Fuente: elaboración propia.

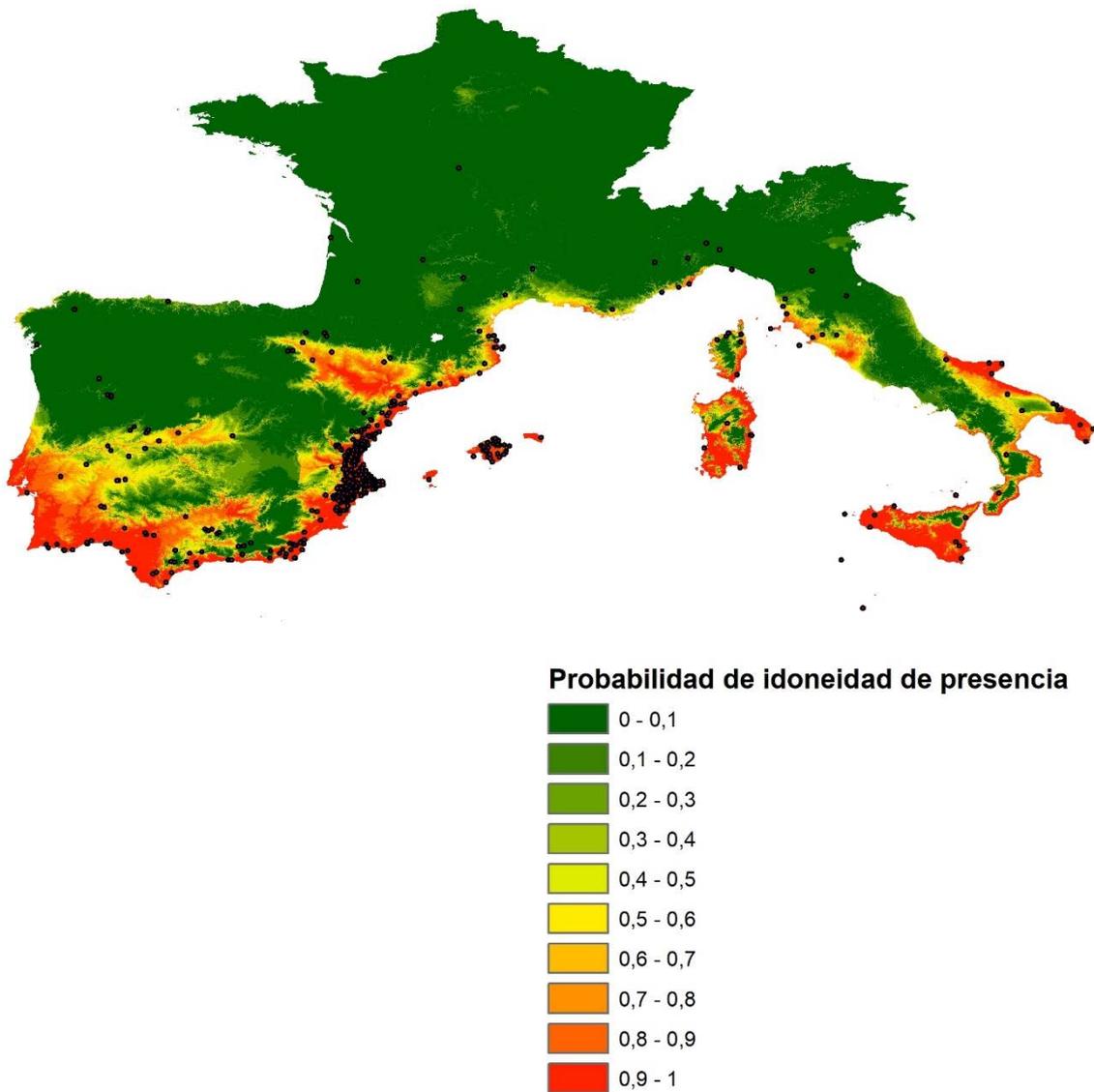


Figura 48. Idoneidad ambiental para *Opuntia ficus-indica* en el Suroeste de Europa (España, Portugal, Francia e Italia) según el modelo construido con las presencias de zonas con menos de 15 habitantes/km² (dataset SDM_GLM_15). Los puntos negros representan las presencias actuales incluidas en el modelo. Fuente: elaboración propia.

Curvas de respuesta a las variables

Las variables predictoras también presentan diferencias en cuanto a su importancia relativa en los diferentes modelos, aunque la temperatura media del trimestre más frío (Bio11) siempre es la de mayor peso (**Figura 49, Figura 51, Figura 53, Figura 55 y Figura 57**). La precipitación anual (Bio12) es la segunda en importancia en los cuatro modelos con restricciones y la cuarta en el modelo que incluye todas las presencias, pues en este caso tienen más peso la temperatura media del trimestre más húmedo (Bio8) y la amplitud térmica diaria (Bio2). La pendiente tiene una influencia prácticamente nula en el modelo con todas las presencias, pero va mostrando más peso en los sucesivos modelos hasta convertirse en la cuarta variable en el modelo más restrictivo.

La tendencia de las curvas de respuesta se muestra similar en los cuatro modelos que descartan datos por elevada densidad de población humana (**Figura 52, Figura 54, Figura 56 y Figura 58**). Solo el modelo construido con todas las presencias muestra diferencia en la curva de la pendiente, que, por otro lado, apenas tiene influencia en dicho modelo (**Figura 49, Figura 50**). La idoneidad aumenta con la temperatura media del trimestre más frío (Bio11), con la temperatura media del trimestre más húmedo (Bio8) y con la pendiente, y disminuye con la precipitación anual (Bio12) y la amplitud térmica diaria (Bio2). Esta última variable tiene baja importancia en el modelo más restrictivo y una tendencia mucho más atenuada que en los otros modelos. Este resultado parece más coherente con los estudios consultados sobre la caracterización ecológica de *Opuntia ficus-indica*, que se considera tolerante a la continentalidad. Dicho de otro modo, el análisis de la importancia relativa de las variables y las curvas de respuesta también parece mostrar que el modelo más restrictivo es el de mayor calidad, ya que presenta mayor coincidencia con los factores ecológicos conocidos.

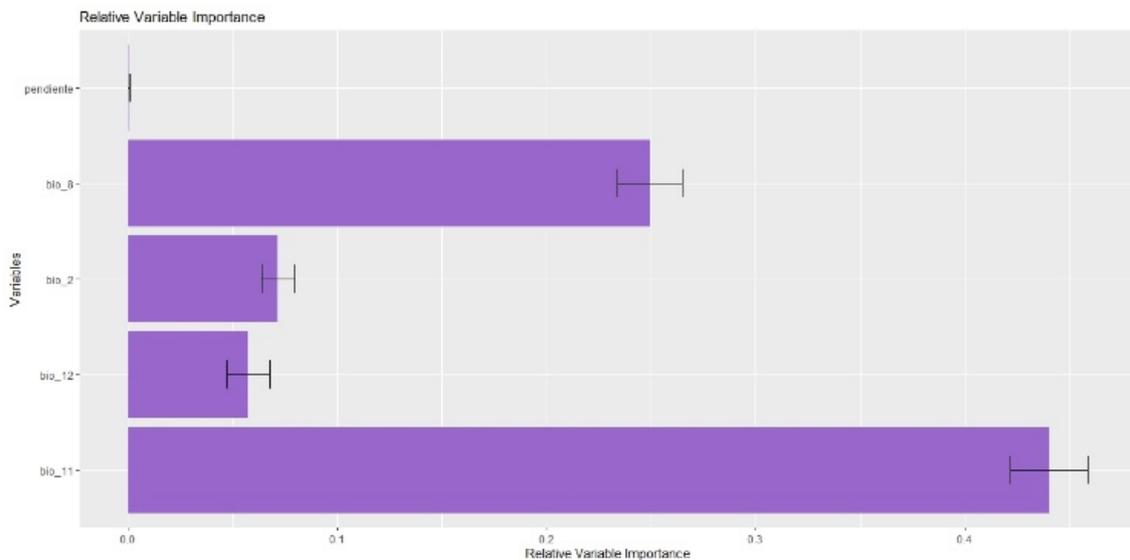


Figura 49. Importancia relativa de las variables en el modelo GLM para *Opuntia ficus-indica* con todas las presencias (dataset SDM_GLM_mas500). Fuente: elaboración propia.

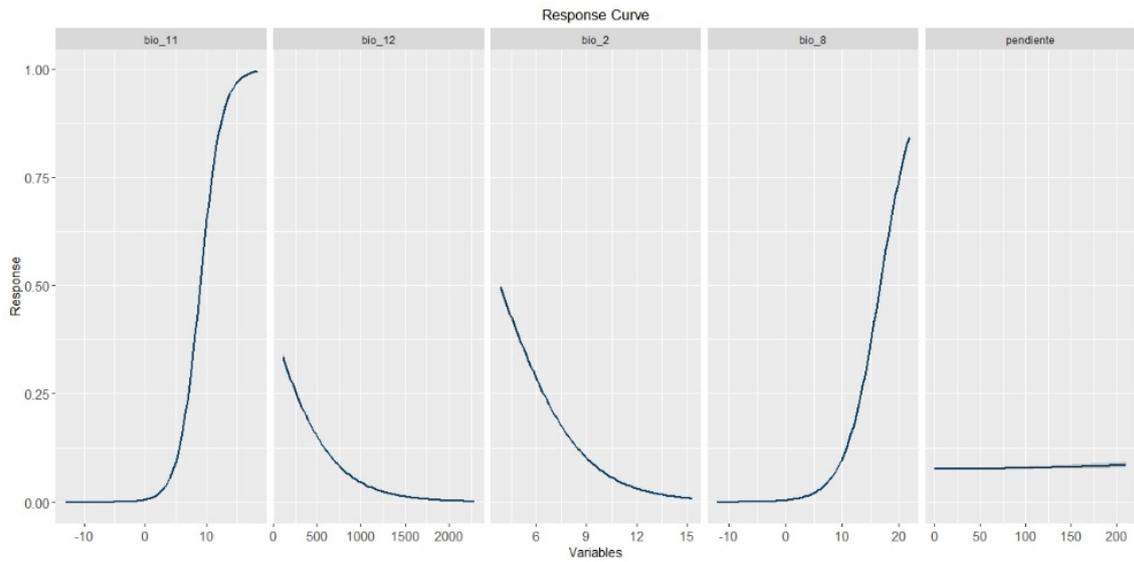


Figura 50. Curvas de respuesta a las variables ambientales en el modelo GLM para *Opuntia ficus-indica* con todas las presencias (dataset SDM_GLM_mas500). Fuente: elaboración propia.

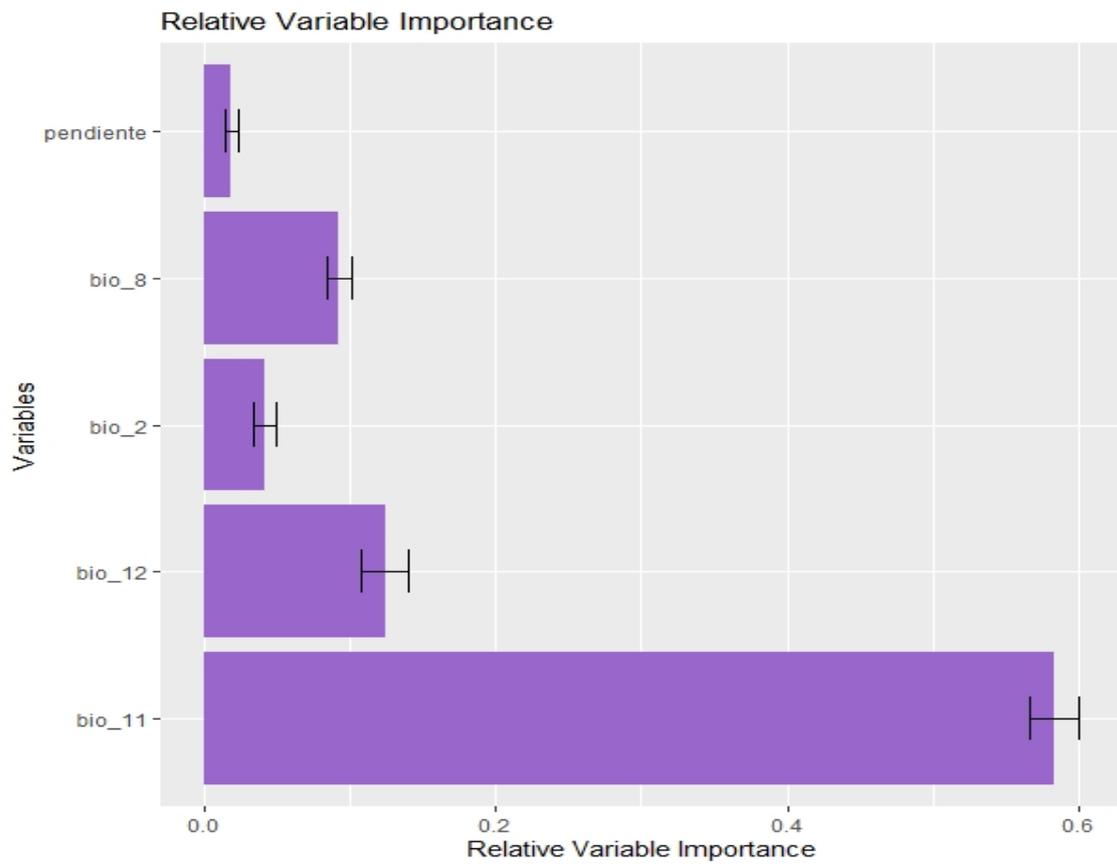


Figura 51. Importancia relativa de las variables predictoras en el modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas con menos de 500 habitantes/km² (dataset SDM_GLM_500). Fuente: elaboración propia.

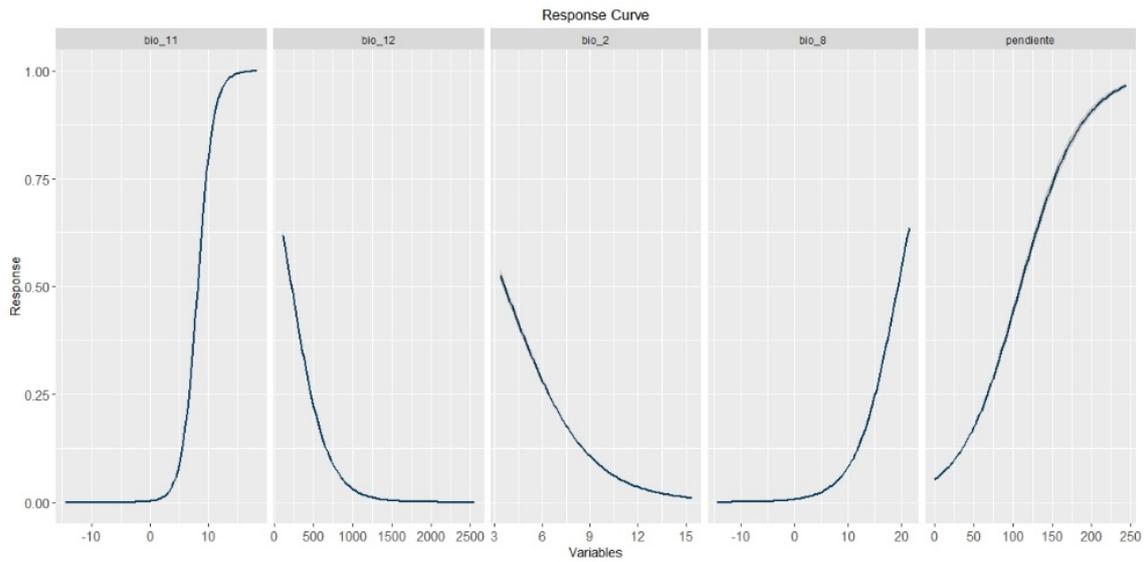


Figura 52. Curvas de respuesta a las variables del modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas de menos de 500 habitantes/km² (dataset SDM_GLM_500). Fuente: elaboración propia.

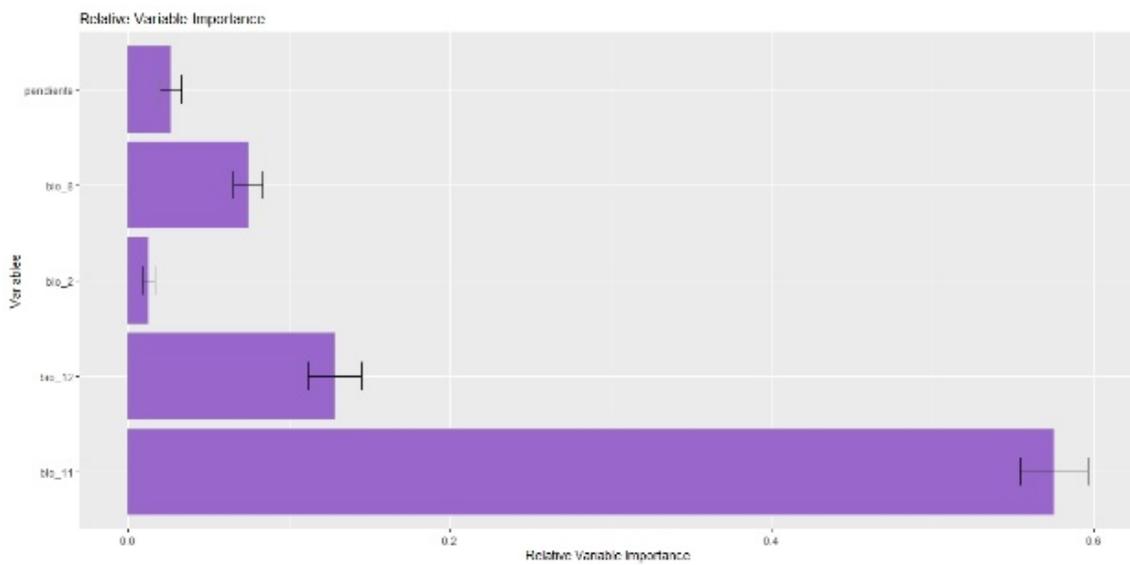


Figura 53. Importancia relativa de las variables predictoras en el modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas con menos de 100 habitantes/km² (dataset SDM_GLM_100). Fuente: elaboración propia.

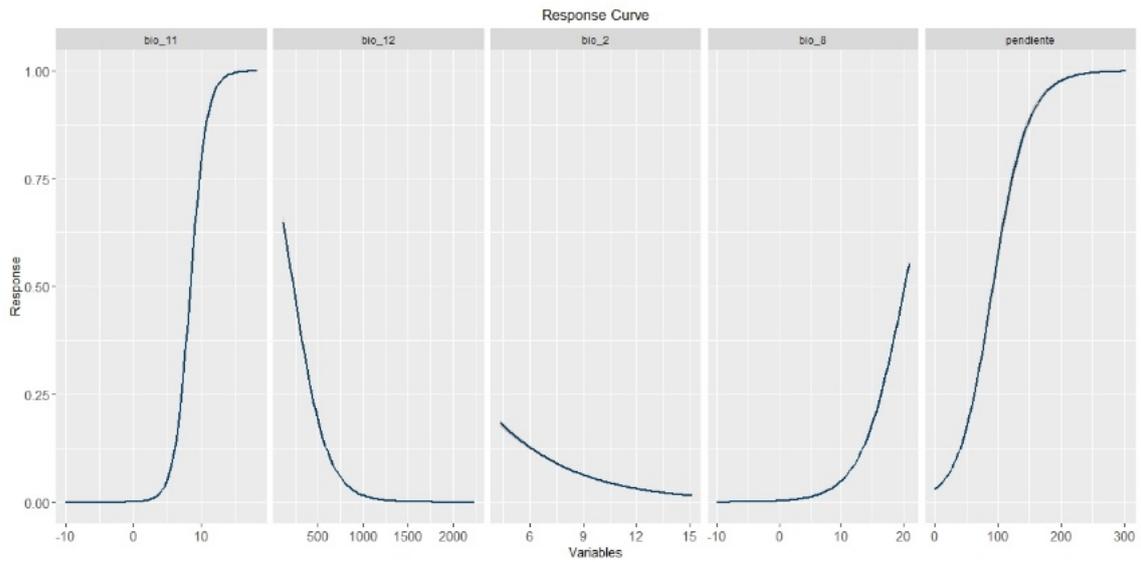


Figura 54. Curvas de respuesta a las variables del modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas de menos de 100 habitantes/km² (dataset SDM_GLM_100). Fuente: elaboración propia.

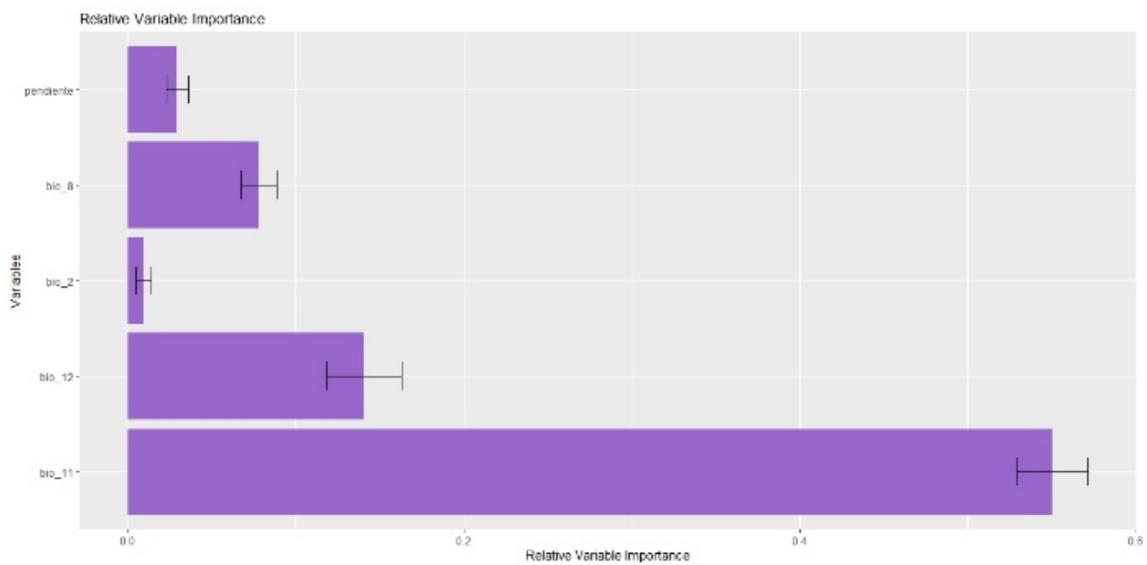


Figura 55. Importancia relativa de las variables predictoras en el modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas con menos de 50 habitantes/km² (dataset SDM_GLM_50). Fuente: elaboración propia.

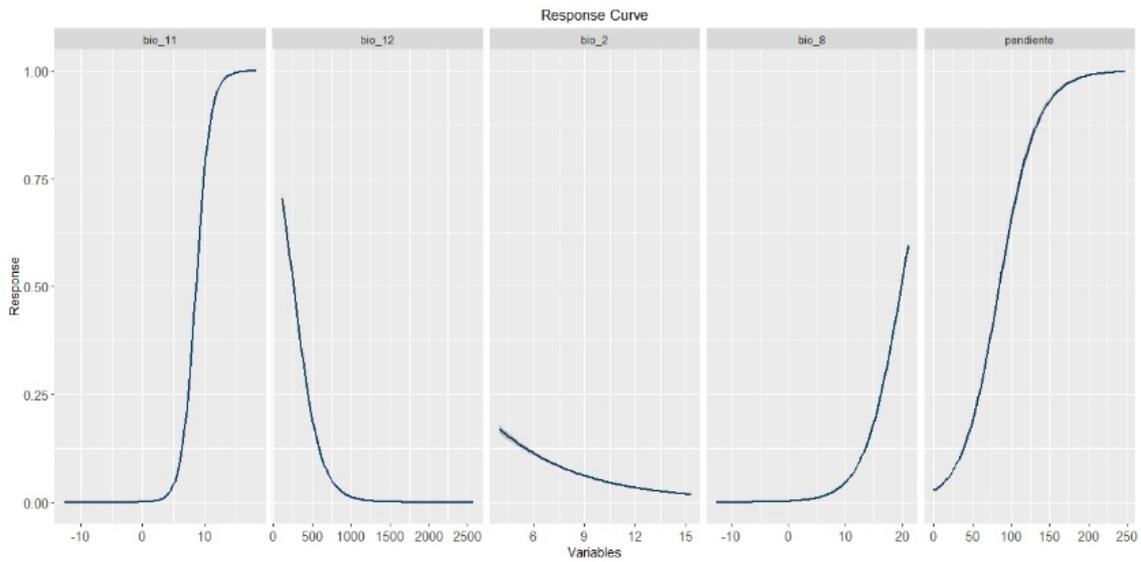


Figura 56. Curvas de respuesta a las variables del modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas de menos de 50 habitantes/km² (dataset SDM_GLM_50). Fuente: elaboración propia.

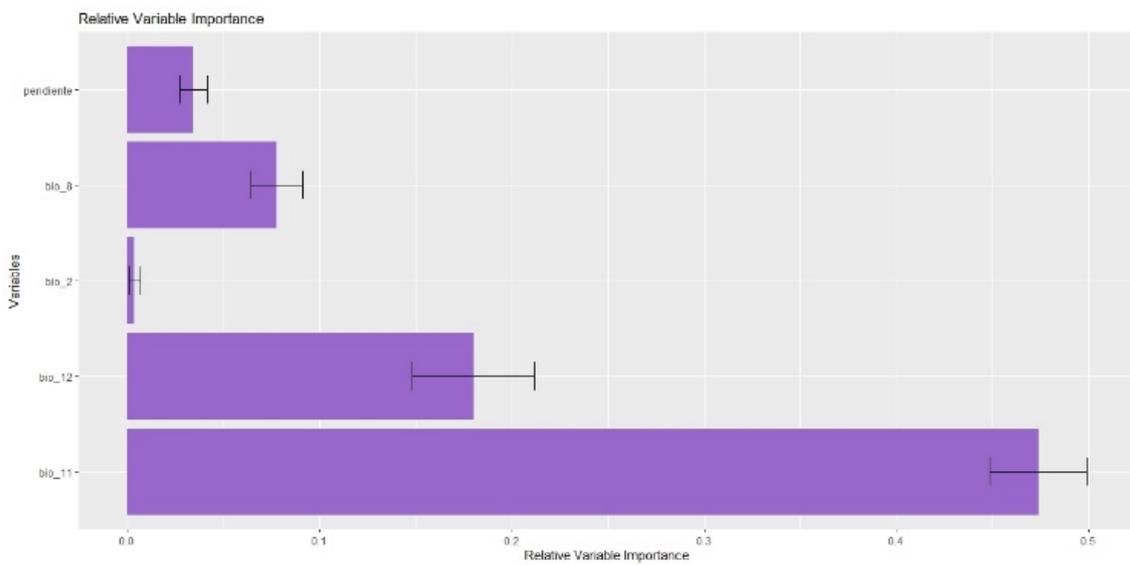


Figura 57. Importancia relativa de las variables predictoras en el modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas con menos de 15 habitantes/km² (dataset SDM_GLM_15). Fuente: elaboración propia.

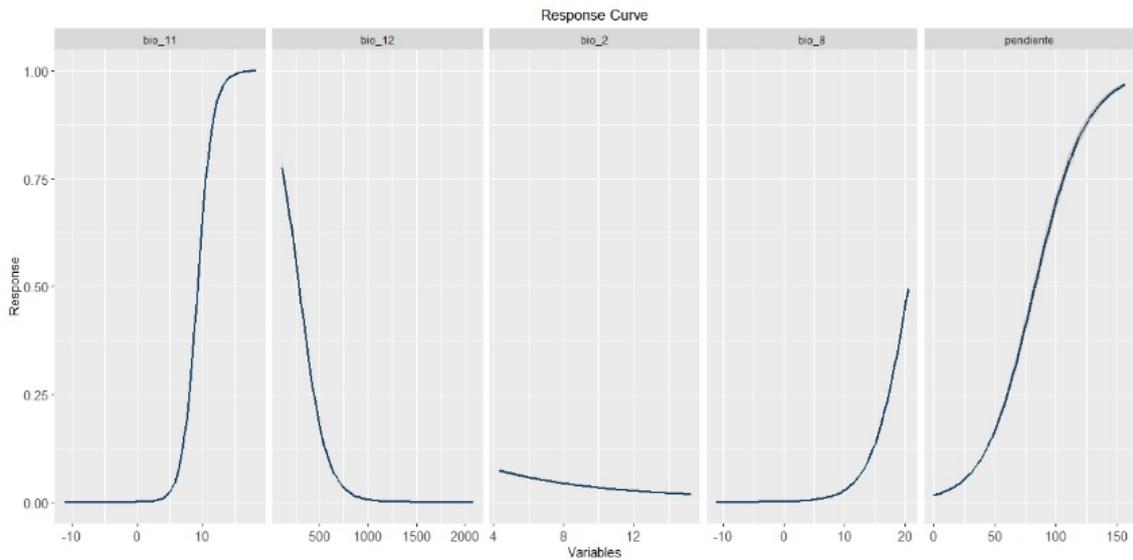


Figura 58. Curvas de respuesta a las variables del modelo GLM para *Opuntia ficus-indica* realizado con las presencias de zonas de menos de 15 habitantes/km² (dataset SDM_GLM_15). Fuente: elaboración propia.

2.4.3 Modelización de la distribución potencial futura

Actualmente la investigación se encuentra todavía en fase de evaluación y comparación de diferentes modelos con el objetivo de seleccionar el más fiable. Por tanto, no se han realizado proyecciones bajo condiciones climáticas de escenarios futuros porque el nivel de incertidumbre derivado de trabajar a la vez con varias propuestas de modelo superaría nuestras capacidades de interpretación y discusión de resultados. Consideramos preferible esperar a la finalización de la fase actual.

Referencias

- Araújo M, Pearson R, Thuiller W, Erhard M. 2005. Validation of species–climate impact models under climate change. *Glob. Change Biol.* 11: 1504–1513.
- Arribas P, Abellán P, Velasco J, Biton D T, Lobo J M, Millán A, Sánchez-Fernández D. 2012. La vulnerabilidad de las especies frente al cambio climático, un reto urgente para la conservación de la biodiversidad. *Ecosistemas* 21: 79-84.
- Beck J, Böller M, Erhardt A, Schwanghart W. 2014. Spatial bias in the GBIF database and its effect on modeling species’ geographic distributions. *Ecol. Inform.* 19: 10–15.
- Elith J, Graham C H, Anderson R P, Dudík M, Ferrier S, Guisan A, Hijmans R J, Huettmann F, Leathwick J R, Lehmann A, et al. 2006. Novel methods improve prediction of species’ distributions from occurrence data. *Ecography* 29: 129-151.
- Elith J, Leathwick J R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Elith J, Phillips S J, Hastie T, Dudík M, Chee Y E, Yates C J. 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17: 43-57.
- EUROSTAT. GEOSTAT 2011 grid dataset
<https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography> (consultado en noviembre de 2020).
- Felicísimo Á M (Coord.) 2011. Impactos, vulnerabilidad y adaptación al cambio climático de la biodiversidad española. 2. Flora y vegetación. Oficina Española de Cambio Climático, Ministerio de Medio Ambiente y Medio Rural y Marino. Madrid. 552 pp.
- Ficetola G F, Thuiller W, Miaud C. 2007. Prediction and validation of the potential global distribution of a problematic alien invasive species - the American bullfrog. *Divers. Distrib.* 13: 476–485.

- Fick S E, Hijmans R J. 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315. <https://doi.org/10.1002/joc.5086>
- Fielding A H, Bell J F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24: 38-49.
- Gallien L, Douzet R, Pratte S, Zimmermann N E, Thuiller W. 2012. Invasive species distribution models—how violating the equilibrium assumption can create new insights. *Glob. Ecol. Biogeogr.* 21: 1126-1136.
- Global Biodiversity Information Facility. URL: <http://www.gbif.org> [Noviembre 2020]
- Hanberry B B, He H S, Palik B J. 2012. Pseudoabsence Generation Strategies for Species Distribution Models. *PLoS ONE* 7(8).
- Harris R M B, Porfirio L L, Hugh S, Lee G, Bindoff N L, Mackey B. 2013. To Be Or Not to Be ? Variable selection can change the projected fate of a threatened species under future climate. *Ecol. Manage. Restor.* 14: 230-234.
- Hijmans R J, Cameron S E, Parra J L, Jones P G, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25: 1965–1978.
- Hijmans R J, Elith J. 2013. Species distribution modeling with R. R CRAN Project. URL: <http://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf>
- IPCC, 2007: Cambio climático 2007: Informe de síntesis. Contribución de los Grupos de trabajo I, II y III al Cuarto Informe de evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático, in: Pachauri R K, Reisinger A (Eds.) IPCC, Ginebra.
- IPCC, 2014: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change in: Stocker T F, Qin D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex V., Midgley P.M. (Eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

- Kumar S, Stohlgren T J. 2009. Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. *J. Ecol. Nat. Environ.* 1: 94-98.
- Mateo R G, Felicísimo A M, Muñoz J. 2010. Effects of the number of presences on reliability and stability of MARS species distribution models: the importance of regional niche variation and ecological heterogeneity. *Journal of Vegetation Science* 21: 908–922. DOI: 10.1111/j.1654-1103.2010.01198.x
- Mateo R G, Felicísimo A M, Muñoz J. 2011. Species distributions models: A synthetic revision. *Rev. Chil. Hist. Nat.* 84: 217-240.
- Mateo R G, Vanderpoorten A, Muñoz J, Laenen B, Désamoré A. 2013. Modeling species distributions from heterogeneous data for the biogeographic regionalization of the European bryophyte flora. *PloS ONE.* 8.
- Merow C, Smith M J, Silander J A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36: 1058–1069.
- Miller J. 2010. Species distribution modeling. *Geogr. Compass.* 4: 490-509.
- Moreno-Amat E. 2017. Reducción de la incertidumbre de los modelos de distribución de especies con datos ecológicos y paleoecológicos. Tesis Doctoral. Universidad Politécnica de Madrid.
- Phillips S J, Anderson R P, Schapire R E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190: 231–259.
- Phillips S J, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161-175.
- Prentice I C, Cramer W, Harrison S P, Leemans R, Monserud R A, Solomon A M. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* 19: 117–134.

- Santini L, Benítez-López A, Cengic M, Maiorano L, Huijbregts M A J. 2020. Assessing the reliability of species distribution projections in climate change research. BioRxiv doi: <https://doi.org/10.1101/2020.06.10.143917>.
- Sanz Elorza M, Dana Sánchez E D, Sobrino Vesperinas E (eds.). 2004. Atlas de las Plantas Alóctonas Invasoras en España. Dirección General para la Biodiversidad. Madrid. 384 pp.
- Schwartz M W. 2012. Using niche models with climate projections to inform conservation management decisions. *Biol. Conserv.* 155: 149-156.
- Settele J, Kudrna O, Harpke A, Kühn I, Van Swaay C, Verovnik R, Warren M, Wiemers M, Hanspach J, Hickler T, et al. 2008. Climatic risk Atlas of European Butterflies. BIORISK, Biodiversity and Ecosystem Risk Assessment. 1, 1-710. Pensoft, Sofia, Moscow.
- Stockwell D R, Peterson A T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological modelling* 148(1): 1-13.
- UICN. 2012. Categorías y Criterios de La Lista Roja de la UICN: Versión 3.1. Segunda edición. Gland, Suiza y Cambridge. Reino Unido: UICN. 34 pp.
- van Proosdij A S, Sosef M S, Wieringa J J, Raes N. 2016. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39(6): 542-552.
- Warren D L, Matzke N J, Iglesias T L. 2020. Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography* 47(1), 167-180.
- Wisn M S, Hijmans R J, Li J, Peterson A T, Graham C H, Guisan A, NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. *Diversity and distributions* 14(5): 763-773.

Referencias software utilizado

ArcMap 10.1: ESRI. 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

dismo package: Hijmans R J, Phillips S, Leathwick J, Elith J. 2015. dismo: Species Distribution Modeling. R <http://CRAN.R-project.org/package=dismo>

HH package: Heiberger R M. 2015. HH: Statistical Analysis and Data Display: Heiberger and Holland. R <http://CRAN.R-project.org/package=HH>

Maxent: Phillips S J, Anderson R P, Schapire R E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259, <http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026>.
<http://www.cs.princeton.edu/~schapire/maxent/>

raster package: Hijmans R J. 2015. raster: Geographic Data Analysis and Modeling. R. <http://CRAN.R-project.org/package=raster>

R Core Team: 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

rgdal package: Bivand R, Keitt T, Rowlingson B. 2015. rgdal: Bindings for the Geospatial Data Abstraction Library. <http://CRAN.R-project.org/package=rgdal>

rJava package: Urbanek S. 2016. rJava: Low-Level R to Java Interface. R <http://CRAN.R-project.org/package=rJava>

R studio: R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.