

Cómo citar este trabajo: Gutiérrez Hernández, O., Cámara Artigas, R., Senciales González, J. M., & García, L. V. (2018). Modelos predictivos en Biogeografía: aplicación para la modelización de nichos ecológicos en Geografía Física. *Boletín de la Asociación de Geógrafos Españoles*, 78, 88–126. doi: <http://dx.doi.org/10.21138/bage.2395>

Modelos predictivos en Biogeografía: aplicación para la modelización de nichos ecológicos en Geografía Física

Predictive modeling in Biogeography:
applying ecological niche modelling in Physical Geography

Oliver Gutiérrez Hernández 

olivergh@uma.es

*Departamento de Geografía
Universidad de Málaga (España)*

Rafael Cámara Artigas 

rcamara@us.es

*Departamento de Geografía Física y Análisis Geográfico Regional
Universidad de Sevilla (España)*

José María Senciales González 

senciales@uma.es

*Departamento de Geografía
Universidad de Málaga (España)*

Luis V. García 

ventura@cica.es

*Instituto de Recursos Naturales y Agrobiología
Consejo Superior de Investigaciones Científicas (España)*

Resumen

La Biogeografía está en el origen de la Geografía moderna. Desde el principio, el estudio de la distribución de los seres vivos y la interpretación de los nichos ecológicos ha constituido uno de los frentes más activos de la Geografía Física. Sin embargo, en algunos países, los geógrafos abandonaron el enfoque ecológico en los estudios de biogeográficos. En este artículo, se propone un flujo de trabajo general para integrar la modelización de nichos ecológicos en el contexto de la Biogeografía como ciencia geográfica que estudia patrones de biodiversidad. El estudio de caso aborda la predicción del nicho ecológico fundamental del alcornoque en Andalucía.

Palabras clave: modelos de nicho ecológico; ecología; biogeografía; *Quercus suber*.

Abstract

Biogeography was born together with the modern Geography. From the beginning, mapping species distribution and the ecological meaning of the niche became one of the frontiers through which the Physical Geography has advanced. However, in some countries the ecological approach was lost in the biogeographic research carried out by geographers. In this paper, we propose a general workflow integrating Ecological Niche Modeling (ENM) within the framework of Biogeography, conceived as the geographical science that studies biodiversity patterns. In this sense, we applied the proposed workflow in order to predict the potential distribution of cork oak in Andalusia.

Key words: environmental niche modelling; ecology; biogeography; *Quercus suber*.

1 Introducción

1.1 El nicho ecológico y la distribución de las especies

El concepto de nicho es uno de los más fecundos y complejos de la Ecología. A lo largo de la historia, este término ha estado sujeto a diferentes acepciones y confusiones (McInerney & Etienne, 2012). El denominado “nicho Grinneliano” (Grinnell, 1917) enfatiza la componente ambiental, centrándose en los factores que limitan la distribución de las especies; aquí el nicho se concibe como una característica del medio y no de las especies. En sentido opuesto, el “nicho Eltoniano” (Elton, 1927) propone una visión en la que se destaca el papel activo de los seres vivos en comunidad, relegando los factores ambientales a un segundo plano. Hutchinson (1957) definiría el nicho ecológico como el espacio n dimensional en el que cada dimensión representa la respuesta de un ser vivo a la variación de un determinado factor, siendo, por tanto, el producto de todos los factores que influyen sobre un organismo. Hutchinson (1978) llega a sugerir la idea de que el nicho de una especie puede incluir un rango de condiciones sin correspondencia física en el espacio geográfico.

En el plano de los hechos, nicho y distribución son realidades estrechamente vinculadas (Pulliam, 2000). En un sentido más amplio, biogeográfico, contemplando la distribución de las áreas, los procesos de dispersión, las interacciones bióticas/abióticas y, en conjunto, la historia natural, las especies están presentes de forma sostenida en aquellos espacios en los que las condiciones ambientales son compatibles con su supervivencia y reproducción y que, además, han podido ser colonizados de una manera u otra (Peterson et al., 2011).

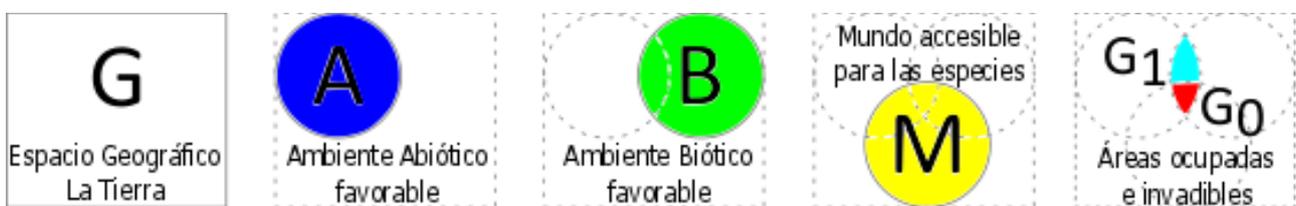
Atendiendo a la Web of Science (Thomsons Reuters, 2016), el término “modelo de distribución de especies” (en inglés “Species Distribution Models”, SDMs) es el más frecuentemente empleado para definir los modelos empíricos que, en base a la teoría del nicho ecológico, abordan la distribución de organismos o comunidades de seres vivos basándose en la relación entre registros de presencia, abundancia o ausencia y determinadas variables ambientales explicativas. Junto con esta denominación, coexisten otras tales como “modelos de envueltas ambientales”, “modelos predictivos del hábitat”, “modelos de idoneidad del hábitat”, “modelos correlativos” u otra muy fecunda como “modelos de nicho ecológico” (en inglés “Ecological Niche Models”, ENMs). La coexistencia de todas estas definiciones produce cierta confusión terminológica, porque, aunque cada una de las denominaciones no se refiere exactamente a lo mismo, a menudo se emplean como sinónimas, cuando en realidad no lo son. Se da el caso de que la definición más utilizada (SDM) es, en la práctica, la más confusa de todas, dado que en la mayor parte de estos modelos las predicciones se aproximan más a la idea de nicho ecológico y a la distribución potencial (ENM), que a un auténtico modelo de distribución de las especies, pues se requiere una serie de ajustes para lograrlo (Saupe et al., 2012).

La modelización basada en la teoría del nicho ecológico gravita en torno a un supuesto cardinal: la asunción de que cada especie —en tanto persiste— está en equilibrio o pseudoequilibrio con las condiciones ambientales que determinan su existencia a lo largo del tiempo (Austin, 2007). Esto significa que los modelos de nicho ecológico proyectan sus predicciones según una “foto fija”: los requisitos del nicho de cada organismo. Aunque la distribución espacial del nicho de una especie pueda estar sujeta a todo tipo de modificaciones, ocasionadas por cambios climáticos, la acción de agentes bióticos o por actividades humanas, se entiende que el nicho ecológico de un taxón, mientras éste existe como tal, es una cualidad constante en el tiempo. Aunque este aspecto podría suponer una limitación metodológica cuando se pretende proyectar los modelos en el tiempo, hacia el pasado o hacia el futuro, esta es una de las aplicaciones más comunes en Macroecología, ya que en la práctica las proyecciones se abordan planteando como hipótesis de partida un rango ecológico específico.

Si bien los modelos predictivos del nicho y la distribución de los seres vivos se nutren de datos ambientales que son capturados y proyectados en el espacio geográfico, necesitamos conceptualizar qué parte del nicho ecológico estamos modelizando realmente (Soberón, 2007). La

Figura 1 presenta el Diagrama de BAM (Biotic-Abiotic-Movement), una representación abstracta del espacio geográfico (G) en el que se distinguen tres regiones (Soberón & Townsend Peterson, 2005). “A” es la región donde se dan las condiciones ambientales adecuadas para la supervivencia y reproducción de la especie, lo que se conoce como nicho ecológico fundamental. “B” es la región que entrelaza el ambiente biótico adecuado para la especie, considerando todas las relaciones interespecíficas. Hasta este punto, estaríamos dentro de la visión del nicho multidimensional de Hutchinson, que no contemplaba la dimensión geográfica de la accesibilidad. Es entonces cuando entra en juego el punto de vista biogeográfico de Alfred Wallace con la inclusión de “M”, la región que representa aquellos lugares que la especie ha tenido tiempo de ocupar a partir de un momento dado. La superposición de estas regiones abstractas permite hablar de “G₀”: el área ocupada por la especie, esto es, el lugar con las condiciones ambientales y bióticas adecuadas que además ha sido accesible por la especie, que sería el espacio donde confluyen todas las regiones ecológicas descritas anteriormente; y, “G₁”: el área invadible, que es el espacio en el que se dan las condiciones bióticas y abióticas propicias, pero que no está ocupado. En todo lo anterior subyace la idea de que cada punto en el espacio geográfico se corresponde con un único punto en el espacio ecológico de una especie dada, pero cada punto en el espacio ecológico de la misma especie puede estar representado por más de un punto en el espacio geográfico.

Figura 1. Diagrama de BAM (biotic-abiotic-movement)



Fuente: elaboración propia a partir de Soberón y Townsend (2005)

1.2 Integración de modelos predictivos en Biogeografía

La Biogeografía estudia la distribución de los seres vivos en el espacio geográfico a través del tiempo. Se trata de una disciplina con una larga tradición, cuyos antecedentes más directos podríamos entroncarlos junto con los orígenes de la propia geografía moderna (Humboldt, 1805), disciplina que eclosionaría al compás de las ideas del evolucionismo (Darwin, 1859; Wallace, 1876). Académicamente, se fueron diferenciando dos disciplinas, Biogeografía y Macroecología, a menudo coalescentes, casi fusionadas, aunque cultivadas desde distintas escuelas (Fisher, 2002; Kent, 2005).

En las últimas décadas, el estudio de los patrones de distribución de la biodiversidad ha ocupado, cada vez más, un papel relevante en la Biogeografía. Algunos autores consideran necesario

reintroducir los principales temas y métodos de la Macroecología para reconducir el rol de la Biogeografía en el contexto de la Geografía Física (Kent, 2007). Desde la primera edición de uno de los principales manuales de referencia sobre Biogeografía (Lomolino et al., 2016), publicada en 1983, se ha venido destacando la importancia del estudio macroecológico de los patrones de distribución, que en los últimos años ha cobrado un nuevo impulso motivado por la revolución GIS.

Una de las primeras revisiones científicas sobre modelos predictivos de vegetación fue publicada a mediados de la década de los noventa, precisamente en *Progress in Physical Geography* (Franklin, 1995). En España, con un desarrollo más tardío de la Biogeografía (Ferrerías Chasco y Fidalgo Hijano, 1991), la comunidad de geógrafos no ha sido muy permeable a la introducción de esta línea de investigación basada en la inferencia estadística (con base ecológica), exceptuando importantes figuras como David Nogués Bravo, quien también relata la misma situación (Nogués Bravo, 2003). Sin embargo, a nivel internacional estamos ante una corriente metodológica que constituye el principal “Research front” (*Essential Science Indicators*, 2016) para el ámbito de la ecología y medio ambiente. Dentro de las ramas “Physical Geography”, “Ecology” o “Environment” del Science Citation Index del Journal of Citation Report, las principales revistas publican con frecuencia todo tipo de artículos sobre modelos de distribución del nicho ecológico de las especies, trabajos que se encuentran entre los más citados (Barbosa & Schneck, 2015).

Desde el punto de vista operativo, los modelos de nicho ecológico y distribución de las especies analizan las relaciones causales existentes entre registros de presencia, abundancia, riqueza o ausencia de una o varias especies y los factores del nicho / distribución: biofísicos, geográficos, históricos, etc. La componente espacial está presente a través de los SIG, que permiten capturar las localizaciones de las especies o comunidades, superponer los factores ambientales o proyectar las predicciones de los modelos en el espacio y en el tiempo. Finalmente, la cartografía constituye uno de los productos finales de estos modelos. No cabe duda de que la Geografía envuelve todas las fases de la modelización: captura de datos, análisis causal y la predicción espacio-temporal.

Las aplicaciones de los modelos predictivos en Biogeografía son muchas y variadas: cartografía de vegetación y comunidades (Box, 1981); simulación de migraciones e impactos relacionados con el cambio climático (Guisan & Thuiller, 2005); localización de nuevas poblaciones de especies poco conocidas (Pearson et al. 2007); estudio de paleo distribuciones de seres vivos en el pasado (Nogués-Bravo, 2009); modelización de invasiones biológicas y riesgos asociados a las mismas en los ecosistemas y agrosistemas (Gallien et al., 2010); diseño de espacios naturales protegidos y la planificación de estrategias de conservación (Tulloch et al., 2016); etc. Realmente las posibilidades solo están limitadas por la naturaleza de nuestros datos y aproximaciones conceptuales.

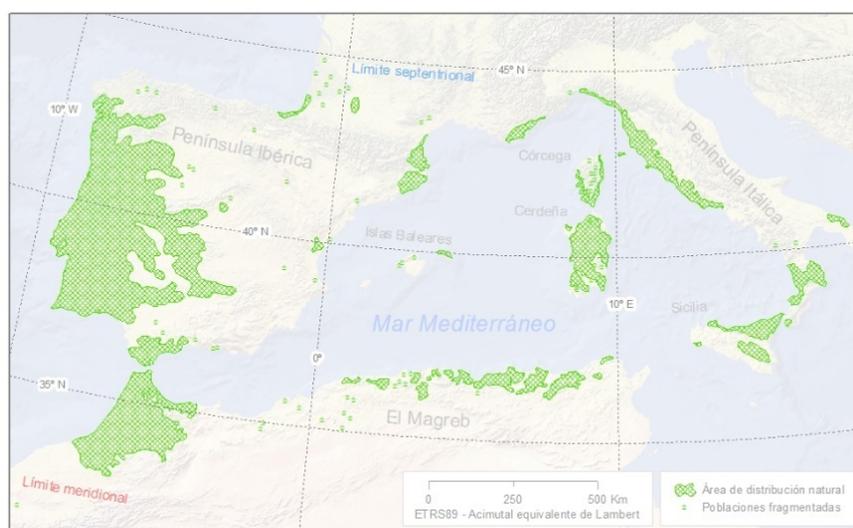
En resumen, por su linaje, entroncado con los orígenes de la Geografía moderna; por sus contenidos, relacionados con la Biogeografía y la Geografía Física; por tratarse de una propuesta

innovadora, impulsada por las Tecnologías de la Información Geográfica y la estadística inferencial; por su actualidad, ligada a la importancia del estudio de la biodiversidad, la conservación de la naturaleza y el impacto del cambio global en la distribución de los seres vivos; y, en conjunto, por la gran cantidad de aplicaciones de interés que comportan para los geógrafos en el campo de la investigación, la planificación y la gestión del territorio y el medio ambiente, consideramos que los modelos predictivos del nicho y distribución de las especies –en su acepción más amplia– deben formar parte del núcleo teórico y metodológico de la Biogeografía, como parte de la Geografía Física.

1.3 El alcornoque, especie objeto de estudio

El alcornoque (*Quercus suber* L.) es una especie endémica del Mediterráneo Occidental (Figura 2), que es especialmente abundante en el cuadrante sudoeste de la península Ibérica (Houston Durrant et al. 2016), donde ha sido objeto de un aprovechamiento amplio y antiguo de sus productos, especialmente del corcho, por lo que tanto la especie como su área de distribución se hallan fuertemente influenciados por la acción del hombre, que no sólo ha favorecido determinadas poblaciones y eliminado otras tantas, sino que además ha seleccionado individuos y recursos genéticos según los usos y aprovechamientos (Toumi & Lumaret, 1998; Urbietta et al. 2008).

Figura 2. Distribución mundial del alcornoque (*Quercus Suber* L.)



Fuente: elaboración propia a partir del European Forest Genetic Resources Programme (EUFORGEN, 2008)

En general, las masas de quercíneas, especialmente *Quercus ilex* y *Quercus suber*, suministran un variado elenco de servicios ecosistémicos que se traducen en grandes beneficios ambientales, sociales y económicos en los espacios rurales (Bugalho et al. 2011; Parsons, 1962). La dehesa (y su homólogo portugués, el montado) constituye un sistema agrosilvopastoral, entre lo natural y lo

cultural, producto de la actividad humana sobre el bosque originario de encinas y alcornoques, donde los árboles aparecen dispersos dentro de una matriz de pastizal, matorral o cultivos herbáceos (Joffre et al. 1999; Silva Pérez, 2010). Según el Tercer Inventario Forestal Nacional (IFN3), la superficie de bosque adehesado en el país alcanza las 2.384.372 hectáreas, cifras similares a las que registra el proyecto sobre ocupación del suelo Corine Land Cover. La dehesa es el pilar fundamental en la economía rural de muchas comarcas extremeñas y andaluzas, como piedra angular de una importante actividad ganadera, cinegética y forestal. Los alcornoques aparecen como formaciones abiertas en la dehesa, pero también como formaciones boscosas cerradas y suponen otra importante fuente de recursos debido a la importancia económica de la producción del corcho (Carrero Carrero, 2015; Coelho et al., 2012). Estos sistemas forestales también representan importantes reservorios para la biodiversidad.

En Andalucía, el alcornoque ocupa una superficie muy importante, que alcanza las 438 000 hectáreas, de las cuales 91 816 constituyen masas puras incluidas como Hábitat de Interés Comunitario 9330 (Díaz Esteban et al. 2009). Se trata de una especie poco xerófila, que requiere cierta humedad ambiental, de carácter termófilo y calcífugo, se presenta sobre suelos pobres y áreas con oscilaciones térmicas moderadas (Urbieto et al., 2011).

En las últimas décadas, el decaimiento de las masas de *Quercus suber* y *Quercus ilex* se está convirtiendo en un asunto de extremada gravedad (de la Hoz Rodríguez, 2009). En estos términos se expresa el Pacto Andaluz por la Dehesa (BOJA, 78/2006), cuando expone “la conservación de las dehesas está hoy día gravemente amenazada por una serie de causas que pueden alterar de modo irreversible el equilibrio de sus recursos y el sistema de explotación”. Ante este panorama, que prevé un futuro incierto para el alcornoque en el sur de la península Ibérica, es necesario un estudio exhaustivo sobre del nicho ecológico fundamental esta especie usando modelos predictivos, tanto para conocer el nicho y distribución de la especie en la región como para discutir las posibilidades metodológicas de este enfoque en trabajos de Biogeografía dentro de la Geografía Física.

1.4 Objetivos

El presente trabajo tiene un doble objetivo:

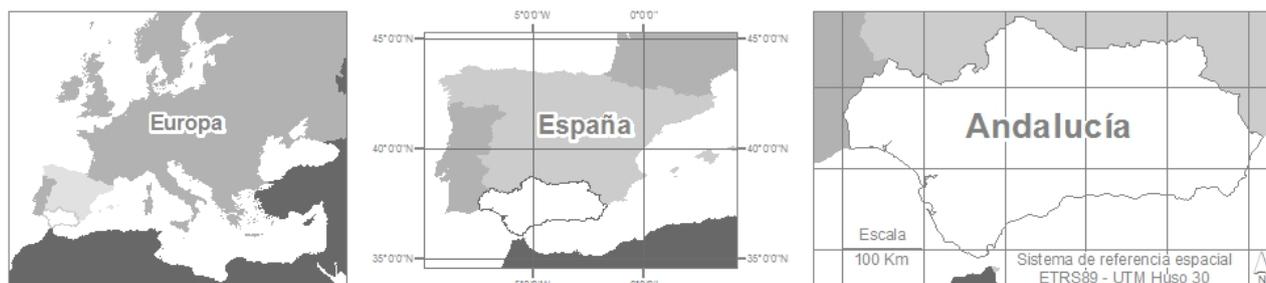
1. Metodológico. Integración de los modelos predictivos de nicho ecológico en Biogeografía.
2. Estudio de caso. Predicción del nicho ecológico fundamental del alcornoque en Andalucía.

2 Material y métodos

2.1 Área de estudio

El área de estudio abarca el conjunto de la Comunidad Autónoma de Andalucía (Figura 3), una superficie de 87 268 km² en el sur de la península Ibérica, dentro de la región biogeográfica Mediterránea, límite meridional del área de distribución del alcornoque en el continente europeo.

Figura 3. Área de estudio



Fuente: elaboración propia a partir de los *Datos Espaciales de Referencia de Andalucía* (Instituto de Estadística y Cartografía de Andalucía, 2017)

Desde el punto de vista biogeográfico, Andalucía está situada en el extremo sur-occidental de la región Mediterránea, en el contexto florístico del Reino Holártico, donde las especies del género *Quercus*, integrantes de la gran familia de las fagáceas, se hallan muy extendidas. Dentro de Andalucía, el alcornoque está presente en casi todas las provincias biogeográficas; muy abundante en las occidentales, especialmente en las provincias Lusitano-Andaluzas Litoral y Luso-Extremadurenses; presente con áreas de distribución muy dispersas dentro de la provincia Bética; y escasa, rara o prácticamente extinta en las provincias más orientales.

2.2 Estrategia de muestreo

Generalmente, la modelización del nicho ecológico se realiza través de métodos estadísticos no espaciales donde se requiere el cumplimiento del supuesto estadístico de independencia de las observaciones entre los sujetos muestrales (Peterson et al., 2011). Pero en realidad, sólo podemos controlar en parte la dependencia propia de los sucesos que ocurren en el espacio geográfico. Por definición, los datos espaciales violan en alguna medida el principio de independencia, de acuerdo con la Ley de Tobler (Tobler, 1970). Los procesos ecológicos suceden en el espacio geográfico y dependen de las distancias, y ello se manifiesta ya en múltiples funciones biológicas (nutrición, reproducción, dispersión, etc.) que proyectan una estructura espacial. Dado que nuestros datos son ecológicos y suceden en el espacio geográfico, diseñamos una estrategia de muestreo para adecuar los datos a la inferencia estadística.

Los registros de presencia son las localizaciones donde consta que la especie está presente. Obtuvimos estos registros a partir del Mapa de Vegetación de Andalucía a Escala de Detalle 1:10 000 (REDIAM) en formato vectorial (polígonos). Rasterizamos los mapas originales y extrajimos una nube de puntos como unidades de observación originales. Redujimos el tamaño de la muestra con una primera selección aleatoria para así disminuir el efecto de la autocorrelación espacial. Debido a que, posteriormente, fijaríamos la resolución de los predictores (mapas raster) en una malla de 250 m, establecimos un máximo de un punto (presencia/ausencia) dentro de cada píxel (o factor), para evitar los efectos negativos que sobre los procedimientos estadísticos tiene la pseudorreplicación entre variables independientes y variables dependientes (Guisan & Zimmermann, 2000). Seguidamente, fijamos una distancia mínima idéntica para todos los puntos de 2 Km, para reducir los efectos relacionados con la agregación de los puntos (*point clustering*). Finalmente, realizamos una última selección aleatoria sobre los puntos restantes, y establecimos 400 registros de presencia.

Los registros de ausencia son las localizaciones donde consideramos que la especie no está presente. Así como un registro de presencia es un hecho basado en una o varias observaciones; un registro de ausencia tiene una naturaleza más compleja que puede expresar diferentes significados: una ausencia podría estar causada por cuestiones biogeográficas y/o requisitos ambientales, pero también por un deficiente esfuerzo de muestreo (Lobo et al. 2010). En nuestro caso, generamos registros de “ausencias ambientales”, asumiendo que la especie ha podido colonizar toda la región, siguiendo los mismos criterios de muestreo descritos anteriormente para las presencias, con algunos matices diferenciales: en lo específico, generamos las ausencias allí donde no se ha documentado presencia de alcornoque y donde, a priori, es poco probable que se desarrolle la especie; pero, establecimos un filtro de 5 Km en torno a las presencias donde fijamos áreas de no muestreo de ausencias, para minimizar el riesgo de falsas ausencias dentro del espectro ambiental de la vecindad de las presencias. Finalmente, también realizamos una selección aleatoria final sobre los puntos que cumplieron todos los requisitos estableciendo 400 registros de ausencia.

De este modo, diseñamos una muestra balanceada con idéntico número de presencias (400) y ausencias (400), contabilizando un total de $n = 800$ puntos o unidades de observación de la variable dependiente (binaria), codificados como “presencia” ($p = 1$) y como ausencia ($p = 0$). Trabajamos una red de puntos lo más representativa y uniforme posible, estableciendo una muestra suficiente para abordar con garantías la inferencia estadística, sin que esta muestra llegara a ser de un tamaño excesivo, con objeto de evidenciar, además, la potencialidad de los modelos para realizar buenas predicciones usando un número limitado de registros y una superficie de muestreo relativamente pequeña.

En nuestro esquema metodológico, los registros de presencia y ausencia componen la variable dependiente (y) que se relaciona con las variables independientes o predictoras (x). Esta relación de $y=f(x)$ es posible efectuarla gracias a una superposición de datos georreferenciados que permite asignar correspondencias. Como en este trabajo empleamos fuentes cartográficas con distintas escalas de captura, como el mapa de vegetación, de donde obtuvimos las presencias, o el mapa litológico, fundamental dado que las propiedades del sustrato se consideran uno de los principales factores discriminantes en la distribución del alcornoque (Serrasolses et al., 2009; Urbietta et al., 2011), fue necesario realizar una última corrección basada en un análisis de superposición SIG para armonizar entre sí los datos de vegetación (presencia/ausencia de alcornoque) y de litología. Así, para garantizar una correspondencia posicional coherente entre registros/factores, realizamos una revisión exhaustiva eliminando puntos incongruentes (en relación con los predictores) y sustituyéndolos por nuevos puntos coherentes, manteniendo los criterios del muestreo y el mismo balance de observaciones.

2.3 Selección de variables predictoras

Los predictores son las variables ambientales que evaluamos para explicar el nicho ecológico de un organismo. Comúnmente, la selección de las variables ambientales está condicionada por el conocimiento que se tiene de la especie objeto de estudio. Los investigadores seleccionan aquellos factores que, según su criterio, pueden ejercer alguna una influencia explicativa sobre la distribución del organismo. En nuestro caso, la selección se ha realizado a partir del conocimiento del alcornoque y algunos trabajos publicados sobre ecofisiología, ecología y distribución de la especie (Díaz Esteban et al., 2009; Houston Durrant et al., 2016; Montero González & Cañellas, 2002; Pérez-Latorre, 1996; Serrasolses et al., 2009; Urbietta et al., 2011). Inicialmente, generamos 18 variables candidatas registradas espacialmente en formato raster a una resolución de 250 m de malla (Tabla 1).

Tabla 1. Variables ambientales iniciales

Variable	Código	Unidades	Fuente > Procesos
Modelo Digital de Elevaciones (MDT)	elev	Metros, m	IECA / REDIAM
Radiación solar	solar	Nº horas, n	MDT > Modelo de radiación
Pendientes	slope	Porcentaje, %	MDT > Análisis del terreno
Litología	lito	Clases	IGME geológico > rasterización
pH	ph	Unid. de pH x10	ISRIC - World Soil Info.
Temperatura media anual	tmean_anual	Grado Celsius, C °	AEMET > Regresión múltiple
Temperatura media del mes más cálido	tmean_warm	Grado Celsius, C °	AEMET > Regresión múltiple
Temperatura media del mes más frío	tmean_cold	Grado Celsius, C °	AEMET > Regresión múltiple
Amplitud térmica media	tmean_range	Grado Celsius, C °	AEMET > Regresión múltiple
Temperatura media de las máximas del mes más cálido	tmax_warm	Grado Celsius, C °	AEMET > Regresión múltiple
Temperatura media de las mínimas del mes más frío	tmin_cold	Grado Celsius, C °	AEMET > Regresión múltiple
Amplitud térmica absoluta	tmaxmin_range	Grado Celsius, C °	AEMET > Regresión múltiple

Tabla 1. Continuación

Precipitación anual	p_tot	Milímetros, mm	AEMET > Regresión-Kriging
Precipitación en otoño	p_autumn	Milímetros, mm	AEMET > Regresión-Kriging
Precipitación en invierno	p_winter	Milímetros, mm	AEMET > Regresión-Kriging
Precipitación en primavera	p_spring	Milímetros, mm	AEMET > Regresión-Kriging
Precipitación en verano	p_summer	Milímetros, mm	AEMET > Regresión-Kriging
Evapotranspiración potencial	pet_tot	Milímetros, mm	AEMET > Regresión-Kriging

Fuente: elaboración propia

Algunas variables se obtuvieron directamente a partir de proveedores de datos, caso del modelo digital de elevaciones (IECA–REDIAM), el litológico (IGME) o el mapa de *pH* (ISRIC). Las variables climáticas se generaron a partir de las estaciones meteorológicas de la AEMET mediante modelos estadísticos de Regresión Múltiple (Benito et al., 2014) y Regresión-Kriging (Peña-Angulo et al., 2016).

Aunque se emplearon todas las variables para abordar un primer análisis causal, con métodos de regresión, para la modelización predictiva, sólo tuvimos en cuenta las variables con mayor capacidad explicativa y, a la vez, menos correlacionadas entre sí. El segundo es un aspecto crucial que no siempre se tiene en cuenta en modelización, y es lo que se conoce como multicolinealidad, hecho que reduce tanto la potencia estadística como la estabilidad de los modelos estadísticos (Graham, 2003).

Para estudiar la multicolinealidad, lejos de automatizar el proceso, nos basamos en una metodología que combina el rigor estadístico y un criterio (dúctil) experto (Benito et al., 2016): 1º, realizamos un análisis de correlación entre las variables predictoras y graficamos los resultados en un correlograma, esto nos permitió conocer en qué grado y dirección estaban correlacionadas entre sí; 2º, convertimos los valores de correlación (+/-) en distancias positivas de similitud entre predictores para componer un clúster que clasificara los factores en ramas; 3º, representamos el análisis clúster proyectando un dendrograma con los predictores agrupados en ramas de parentesco y escogimos una variable por cada rama cortada en el umbral de 0,4 de distancia; 4º, realizamos un nuevo filtro estadístico para detectar si cada variable seleccionada en el paso anterior constituía una combinación lineal de otras variables, y, para ello, se calculó el VIF (Variance Inflation Factors) siguiendo un procedimiento secuencial donde excluimos las variables con un VIF superior a 5 (James et al., 2013).

2.4 Métodos estadísticos y minería de datos

Los modelos lineales se emplean para explicar la relación de una variable respuesta o dependiente en función de una variable explicativa o independiente. En matemáticas se emplea la expresión $y = f(x)$ para expresar la relación de “y” en función de “x” y ajustar la denominada ecuación de regresión:

$$y = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Donde β_0 es el intercepto, o valor de la función cuando $y=0$; β_1 expresa el cambio de la variable y por cada unidad de cambio de cada variable x_i , es decir, proyecta la pendiente de la recta de regresión; y ε_i representa el error. β_0 y β_1 son valores constantes y ε_i es un valor aleatorio ajustado a una distribución normal. La relación no se refiere a una simple correlación, sino que presupone una conexión causal, por lo que es necesario asumir una serie de supuestos:

1. Relación lineal. La respuesta de y en función de x es lineal.
2. Independencia de las observaciones. Los sujetos de la muestra son independientes entre sí.
3. Normalidad. Los errores se distribuyen siguiendo una distribución normal o gaussiana.
4. Homocedasticidad. La varianza es constante, especialmente la de los errores.

Sin embargo, no siempre se cumplen estos supuestos. En los apartados anteriores, describimos una serie de procedimientos para minimizar la dependencia de las observaciones, tanto en los predictores como en los registros de presencia. Por otro lado, otro de los ejemplos en que no se cumple la relación lineal es cuando la variable respuesta es binaria, como es nuestro caso: presencia (1) versus ausencia (0). Por ello, recurrimos a una extensión de los modelos lineales clásicos.

Los Modelos Lineales Generalizados (GLM, Generalized Linear Models) (Nelder & Wedderburn, 1972) permiten trabajar con diferentes distribuciones de la variable dependiente (Poisson, Binomial, Gamma, Exponencial) la cual se relaciona con una combinación lineal de variables independientes a través de una función de vínculo (*link function*) que permite linealizar la relación entre ambas mediante una transformación de la variable respuesta, lo que evita recurrir a transformaciones (e.g. logarítmica, donde $y = \beta_0 + \beta_1 \ln x_i$) que a la postre son difíciles de interpretar. En este sentido, los GLM son una alternativa a la clásica transformación de los datos. En nuestro caso, generamos una curva logística para ajustar la distribución de la variable respuesta. La regresión logística se encuadra dentro de los Modelos Lineales Generalizados. En ocasiones, esperamos respuestas no lineales en la variable dependiente, más allá de la estructura binomial de la distribución de la mencionada variable. En este supuesto, los Modelos Aditivos Generalizados – GAM (Generalized Additive Models) (Hastie y Tibshirani, 1986) permiten modelar directamente respuestas no lineales mediante un suavizado (*smoothing*) de las variables predictivas. Es un método de regresión no lineal y no paramétrico, porque trabaja con valores suavizados de cada variable que en conjunto pueden adoptar una respuesta no lineal. En otras palabras, es el propio modelo el que se ajusta empíricamente generando múltiples curvas denominadas *splines*. Tanto los Modelos Lineales Generalizados como los Modelos Aditivos Generalizados se han empleado con éxito para modelizar la distribución de especies (Austin, 2002; Guisan & Edwards, 2002).

Los métodos de regresión se emplearon para buscar las conexiones causales y combinaciones óptimas entre variables, pero también para realizar las predicciones definitivas. Adicionalmente, para contrastar dichas predicciones, utilizamos métodos basados en algoritmos de inteligencia artificial y aprendizaje automático (*Machine Learning*) usando los mismos registros de presencia y las mismas variables predictoras; si bien, en lugar de emplear registros de ausencia, utilizamos el *background*, es decir, una selección aleatoria de datos que permite caracterizar el rango de condiciones del área de estudio. Se trata de una aproximación complementaria que entiende que, esencialmente, registros de presencia y ausencia tienen distinta naturaleza, y no siempre representan una relación de antagonismo debido a que las ausencias, como explicábamos anteriormente, pueden tener distintos significados. Existe un tercer grupo de algoritmos que opera sólo con presencias, estimando envolventes o distancias ecológicas, pero descartamos su aplicación debido a su menor capacidad discriminante para ajustar las predicciones (Elith & Graham, 2009). Esto no significa que estos métodos sean peores, sino menos adecuados para alcanzar los objetivos de este trabajo.

Dentro de la aproximación complementaria para generar las predicciones basadas en presencia versus datos de fondo, utilizamos dos algoritmos diferentes: Maxent (Phillips & Dudi, 2008) y Random Forest (Breiman, 2001). Maxent es uno de los algoritmos más populares en modelización de nichos ecológicos, tanto por los buenos resultados que ofrece, como por la facilidad de uso de su interfaz gráfica más extendida. Como su nombre indica se trata de un algoritmo que maximiza la entropía bajo las restricciones de los factores considerados buscando la distribución más uniforme posible, por lo que resulta especialmente útil cuando se dispone de pocos registros de presencia (Elith et al. 2011); si bien, este método también ha sido descrito o re-interpretado como un tipo de regresión logística basada en presencia/pseudo-ausencia que sigue un proceso puntual de Poisson y que, por tanto, no puede determinar probabilidad de presencia (Renner & Warton, 2013). Random Forest es un algoritmo de aprendizaje automático muy potente que puede trabajar con grandes volúmenes de datos, tanto con registros de presencia/background como de presencia/ausencia, a diferencia de Maxent. Se basa en la construcción de múltiples árboles de regresión aleatorios independientes y no correlacionados entre sí, cuyos resultados son promediados, minimizando el ruido en las predicciones.

En los métodos de regresión se utilizaron 400 registros para caracterizar el rango de condiciones de las presencias y 400 registros para hacer lo propio con el rango de condiciones de las ausencias; con esta aproximación complementaria, contrastamos los mismos registros de presencias respecto a 400 puntos aleatorios que describen el rango de condiciones de toda el área de estudio. Para minimizar un posible sesgo muestral, realizamos diez modelos independientes por cada algoritmo (GLM, GAM, Maxent, Random Forest) seleccionando los registros aleatoriamente

mediante el procedimiento de remuestreo conocido como Bootstrapping (Efron, 1981), y en cada iteración escogimos un 50 % de los registros para entrenar el modelo y un 50 % para evaluarlo.

Resumiendo, utilizamos los métodos de regresión para estudiar los vínculos causales entre los factores ambientales y los registros de presencia y ausencia de la especie, y ajustar el mejor modelo: basado en la combinación óptima de variables significativas siguiendo el principio de parsimonia (Goloboff, 2003). En este sentido, consideramos tanto relaciones aditivas (expresadas con el operador "+") e interacciones (expresadas con el operador "*") entre variables ($\beta_1 x_i$) del modelo. Junto con los métodos de regresión (GLM y GAM) basados en datos de presencia/ausencia, también incorporamos los métodos de inteligencia artificial (Maxent y Random Forest) aplicados en datos de presencia/background, usando la misma combinación óptima de predictores, para generar las predicciones definitivas (espaciales) sobre la distribución potencial del alcornoque.

2.5 Evaluación de modelos

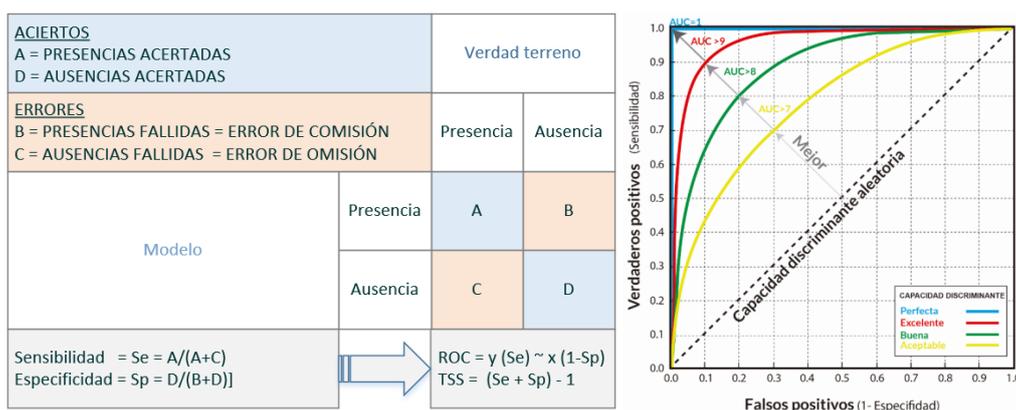
Siguiendo el esquema anterior, abordamos una doble evaluación de los modelos.

En primer lugar, recurrimos a la inferencia estadística para calibrar los modelos, empleando la significación estadística (p-valores) y los coeficientes de regresión de cada variable explicativa ($\beta_1 x_i$) en el marco de los métodos de regresión, así como las principales medidas de tendencia central y dispersión para describir los datos. A continuación, evaluamos la combinación óptima de variables para construir el mejor modelo en función de la máxima devianza explicada (D2), es decir, la cantidad de varianza explicada por el modelo. Conjuntamente, siguiendo el principio de parsimonia, descartamos los modelos más complejos y escogimos según el criterio de información de Akaike o AIC (Akaike, 1974) el modelo más simple, aquel que tiene un AIC más bajo. Finalmente, examinamos gráficamente la distribución de los errores a través del histograma de los residuos y el gráfico Q-Q. Una buena predicción es aquella que proyecta los errores siguiendo una distribución normal de acuerdo con el Teorema del Límite Central. El mejor modelo estadístico sería aquel que incluyera variables estadísticamente significativas y explicase una mayor cantidad de varianza con la menor complejidad posible y cuyas predicciones y residuos se ajustaran a una distribución normal.

En segundo lugar, evaluamos la capacidad predictiva de todos los algoritmos a partir del modelo seleccionado en el paso anterior. Los modelos de nicho ecológico proyectan los valores de idoneidad (o probabilidad de presencia) con guarismos que oscilan entre 0 y 1. Un buen modelo debe maximizar el acierto de presencias y el acierto de ausencias. De este modo, podemos encontrar dos fuentes de error fundamentales. Por un lado, que el modelo genere falsos positivos, esto es, valores de presencia fallidos, incurriendo en error de comisión; y, falsos negativos, o sea, valores de ausencia falsos, incurriendo en error de omisión. Esta relación se ilustra en la

denominada matriz de confusión, por medio de la cual calculamos la sensibilidad y la especificidad (Fielding & Bell, 1997). A través de estos parámetros trazamos la curva ROC (Receiver Operating Characteristic) (Hanley & McNeil, 1982), que es el resultado de representar en el eje Y (1-sensibilidad) y en el eje X (1-especificidad). El resultado es el estadístico AUC (Area Under the Curve), es decir, la probabilidad de que, escogiendo aleatoriamente una presencia o una ausencia, el modelo otorgue al positivo un mayor valor. Un valor de 0,5 indica que el modelo no tiene capacidad discriminante. Y un valor de 1 indica que el modelo tiene máxima capacidad discriminante. En general, se asume que un modelo con buena capacidad discriminante tiene un AUC por encima de 0,80, si bien debemos calibrar el verdadero significado de este parámetro, pues también depende de la representatividad del dominio ambiental (Jiménez-Valverde et al., 2013). Además, no siempre interesa ponderar equilibradamente sensibilidad y especificidad, no en vano, desde el punto de vista teórico el error de omisión se considera más grave que el error de comisión. La Figura 4 muestra la matriz de confusión y la curva ROC.

Figura 4. Matriz de confusión y curva ROC



Fuente: elaboración propia a partir de Fielding y Bell (1997)

En la práctica, el AUC se ha convertido en un estándar de evaluación de la capacidad discriminante de los modelos y es la métrica discriminante más utilizada. Sin embargo, algunos autores han puesto en duda su uso indiscriminado como estándar por diferentes limitaciones relacionadas con la naturaleza de los datos (presencia/ausencia, presencia/background) y la extensión del área de estudio (Lobo et al. 2008). Por ello, decidimos implementar otros estadísticos complementarios. Siguiendo la matriz de confusión, también calculamos el True Skill Statistic ($TSS = Sensibilidad + Especificidad - 1$). Este estadístico es una medida de ajuste similar a *Kappa*, pero mejorada, pues evita el efecto de la prevalencia ($p = A+C/n$, esto es, proporción de presencias en la muestra total), manteniendo sus ventajas para estimar conjuntamente los errores de omisión (B) y comisión (C) (Allouche et al., 2006). Además, calculamos el coeficiente de correlación (COR) como medida paramétrica de correlación biseriada entre los valores observados frente a los valores predichos

(Zheng & Agresti, 2000), la cual aporta un extra de información con respecto al AUC, dado que varía linealmente con el error. Por último, computamos la reducción la devianza como medida de ajuste del modelo en relación con los parámetros del mismo, donde un menor valor indica mayor parsimonia.

Todas las métricas descritas son globales y no espaciales. Por este motivo, a partir de las distintas réplicas de los diferentes modelos, proyectamos la distribución espacial del error ($SE\bar{x}$) en los valores predichos, como una aproximación a la distribución geográfica de la incertidumbre.

2.6 Especificaciones técnicas y software empleado

Realizamos el pre-procesado de datos raster con el Sistema de Información Geográfica Grass 7.2. Con este SIG preparamos las variables ambientales codificadas en formato raster a una resolución de 250 m y definimos el sistema de referencia espacial: ETRS89-UTM Huso 30. Dentro del entorno y lenguaje de programación R, importamos los datos espaciales, desarrollamos los análisis estadísticos, los trabajos de modelización, así como la producción gráfica y cartográfica. Como sería muy prolijo citar todos los paquetes (o librerías) utilizados en este sistema, mencionaremos aquellos que han tenido una mayor relevancia. Así, usamos “dismo” (Hijmans et al., 2011) para el ajuste y evaluación de modelos estadísticos; y, empleamos “sdm” (Naimi & Araújo, 2016), para la construcción y ensamblado de los modelos de nicho ecológico proyectados en el espacio geográfico. Esta última librería facilita la construcción de modelos basados en múltiples algoritmos (incluido Maxent) usando los mismos registros, por lo que la evaluación de modelos se hizo incorporando exactamente los mismos datos, hecho que no es posible cuando se aplican diversos algoritmos desde diferentes plataformas informáticas no integradas con distintos datos remuestreados aleatoriamente.

3 Resultados y discusión

3.1 Contraste de los registros de presencia / ausencia

Trabajamos con un total de 800 puntos, entre registros de presencia y ausencia. Dado que dentro de cada píxel se estableció una única unidad de observación, consecuentemente, sólo utilizamos una superficie de muestreo acumulada de 5000 hectáreas (cada punto computa 6,25 hectáreas): la mitad para entrenar el modelo y la otra mitad para evaluarlo. Por lo tanto, los modelos calcularon la distribución potencial para toda el área de estudio utilizando el 0,05 % de la superficie de la misma, y el 1,14 % del área con respecto a la superficie ocupada por alcornoques o fuente de presencias.

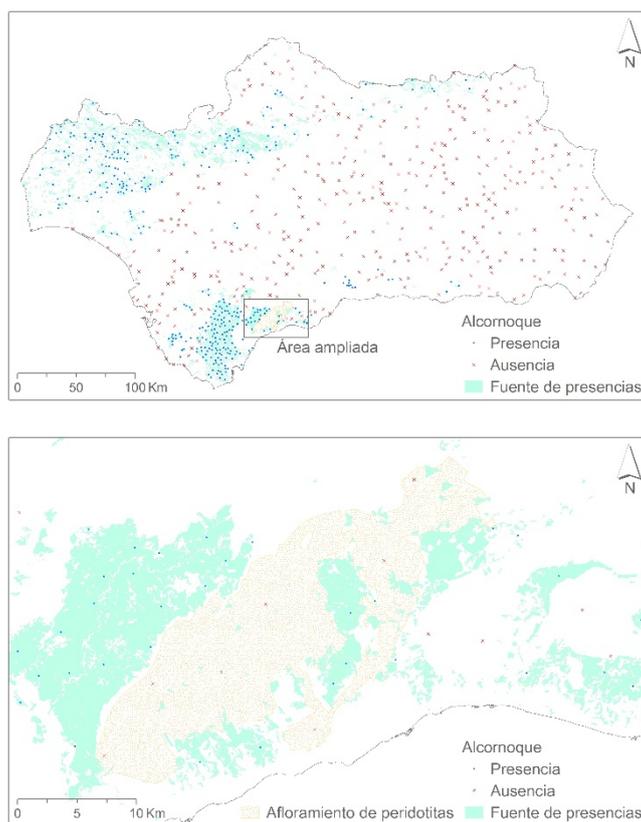
Con el procedimiento de muestreo conseguimos una distribución de los puntos en toda el área de estudio (Figura 5), con los registros de presencias más concentrados que los de ausencia, debido a que empleamos la misma cantidad de puntos para caracterizar presencias y ausencias. Esta

densidad espacial variable produjo una mayor autocorrelación en los registros de presencia, algo que es normal en datos ecológicos (Dormann et al., 2007). Con el muestreo, redujimos este efecto, especialmente en los registros de ausencias, hecho que logramos porque estas ocuparon toda el área de estudio y así logramos balancear las observaciones considerando que se analizarán con métodos no espaciales. Cabe añadir que, posteriormente, con la aplicación de técnicas de remuestreo aleatorio (*bootstrap*) en cada una de las 10 iteraciones ejecutadas por cada algoritmo en la fase de construcción de modelos, redujimos aún más el efecto de la autocorrelación espacial al seleccionar un 50 % de los puntos para entrenar el modelo. En el siguiente apartado, evaluamos la efectividad de los pasos llevados a cabo para el control de la normalidad analizando la distribución de los residuos del modelo. Si hubiera falta de normalidad, podría ser por la influencia de la autocorrelación espacial (Naimi et al., 2011).

Como resultado de la revisión supervisada de las observaciones basada en la superposición de cartografía y fotointerpretación, corregimos 12 registros de presencia y 49 registros de ausencia. Eliminamos presencias en áreas bien muestreadas y las reubicamos en algunos enclaves representativos y aislados donde tenemos registros actuales de la especie, pero que fueron omitidos por los resultados de la aleatorización. Se trata de poblaciones disjuntas situadas en La Vera de Doñana (Huelva), La Mocheta-El Arahal (Sevilla), Montes de Málaga (Málaga) y Haza del Lino (Granada). Aunque todavía permanecen algunos pies de alcornoques en la sierra de Cabrera (Almería), descartamos ubicar presencia (ni ausencia) alguna para usarla en el modelo, ya que se trata de individuos centenarios y hace mucho que no se produce reclutamiento de nuevas poblaciones.

La corrección de las ausencias fue más laboriosa debido a que estas se distribuyeron por toda el área de estudio, y no fueron suficientes las restricciones establecidas en el muestreo. Quedaron muchos enclaves calizos sin registros de ausencia. El alcornoque no está presente en suelos calizos o dolomíticos, excepto cuando estos, localmente, han sido descarbonatados, como ocurre en arcillas de descalcificación (*terra rossa*) donde sí pueden medrar los alcornoques (Serrasolses et al., 2009). Por consiguiente, reubicamos ausencias en zonas calizo-dolomíticas no descarbonatadas. Asimismo, definimos registros de ausencia sobre determinados espacios arenosos en ombrotipos secos sin acceso a un nivel freático cercano. Por último, también asignamos registros de ausencia sobre afloramientos de peridotitas, basándonos en el mapa de vegetación actual y trabajos publicados sobre vegetación en peridotitas (Gómez Zotano, 2004). En este sentido, en el libro *Introducción a la historia natural y a la geografía física de España* de Guillermo Bowles, se hace referencia explícita a la ausencia de alcornoques sobre peridotitas en sierra Bermeja (Málaga), “La blanca produce sólo Alcornoques y Encinas, y la roxa (en referencia al color bermejo de las peridotitas meteorizadas) no tiene ninguno de estos árboles y está cubierta de Abetes” (Bowles, 1782; p 76).

Figura 5. Muestreo general de registros de presencia / ausencia y ejemplo de área revisada



Fuente: elaboración propia

Teóricamente, la aleatorización de los registros sujeta a restricciones espaciales propicia una sistematización del muestreo que prepara los datos para un tratamiento estadístico óptimo (Jiménez-Valverde & Lobo, 2006). El conocimiento de la especie y del área de estudio es fundamental para corregir errores generados por la selección automática. Recomendamos la consulta de la bibliografía y documentos históricos para establecer con mayor fundamento los registros de presencia y ausencia, especialmente los de ausencia, cuya naturaleza es más problemática. En conjunto, un enfoque geográfico con apoyo del SIG contribuye a la mejora en la calidad de los datos, porque es fundamental tener datos representativos para desarrollar buenos modelos (Lobo, 2008).

3.2 Variables explicativas del nicho del nicho ecológico fundamental. Autoecología

El resultado de la exploración estadística y el modelo aditivo generalizado aplicado (GAM) para conocer la importancia de cada variable, mostró que la mayor parte de las variables consideradas ejercieron una influencia significativa sobre la distribución de la especie (Tabla 2).

Tabla 2. Regresión

Variable	Estadísticos descriptivos				Regresión	
	\bar{x}	σ	Min	Máx	β	p-valor
elev	368	243	4	1264	-0.0018	<0.001***
solar	2009	258	737	2431	-0.0002	0.426
slope	15.7	11	0.3	52.2	0.0211	<0.001***
pH	6.5	3.4	5.8	7.7	-0.4238	<0.001***
tmean_anual	16.5	1.1	13	18.3	0.3520	<0.001***
tmean_warm	25.1	1	22	27.4	0.0064	0.890
tmean_cold	9.7	1.6	5.5	12.6	0.4669	<0.001***
tmean_range	15.4	1.9	11	20.5	-0.5188	<0.001***
tmax_warm	32.4	2.4	26	36.4	-0.1055	<0.001***
tmin_cold	5.2	2.1	0.8	9.58	0.5005	<0.001***
tmaxmin_range	27.2	4.1	18	35	-0.2370	<0.001***
p_tot	881	181	573	1395	0.0062	<0.001***
p_autumn	228	40	150	337	0.0281	<0.001***
p_winter	389	101	222	714	0.0136	<0.001***
p_spring	228	43	147	348	0.0190	<0.001***
p_summer	33.9	9.1	17	57.6	-0.0170	0.005**
pet_tot	1279	164	782	1617	-0.0025	<0.001***

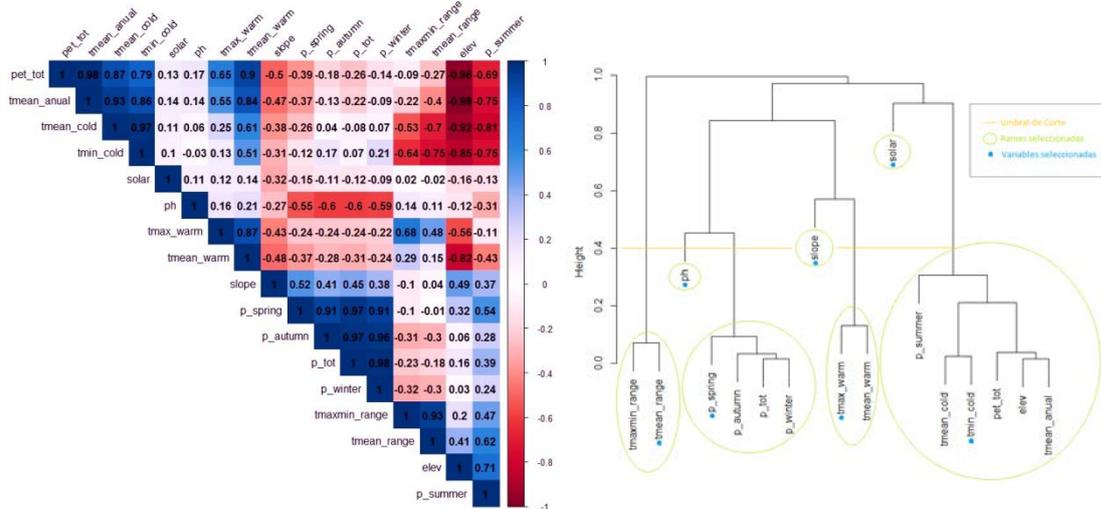
Fuente: elaboración propia

La temperatura media del mes más frío, la amplitud térmica y el pH mostraron una gran influencia, demostrando cómo la especie evita los lugares fríos y con grandes amplitudes térmicas y prefiere suelos entre ligeramente ácidos y moderadamente ácidos. Con respecto a las precipitaciones, se podría interpretar que las ausencias del alcornoque responden a ambientes áridos, y la presencia ocurre dentro de ambientes subhúmedos a húmedos, con un gran espectro de indefinición en los ambientes secos, donde los requisitos de agua interactúan con factores favorecedores o limitantes; también se advierte cómo la especie no está bien representada en zonas hiperhúmedas, que suelen coincidir con las áreas montañosas más elevadas, dado que la ocurrencia de la especie es escasa por encima de los 1000 m en estas latitudes. Resultados similares han sido encontrados en publicaciones recientes (López-Tirado & Hidalgo, 2016). Sin embargo, hemos de interpretar el análisis de los factores sobre el control de la distribución de la especie en el marco de una escala regional de trabajo a una resolución de 250 m. Localmente, algunos factores pueden tener una mayor influencia, operar en sentido inverso o interactuar con otros, pero ya no para caracterizar el nicho fundamental de la especie en sí, sino la presencia, diversidad, abundancia o ausencia de determinadas poblaciones.

3.3 Selección de los predictores y determinación de los parámetros del modelo general

En principio, detectamos una correlación elevada entre las variables relacionadas con las elevaciones y temperaturas. En efecto, en relación con el gradiente término, la elevación tuvo una correlación muy alta con la temperatura media de las mínimas invernales, y moderada-alta con la temperatura media de las máximas veraniegas y la amplitud térmica. La correlación fue más elevada entre la amplitud térmica y la temperatura media de las máximas veraniegas. El resto de variables mostraron una correlación muy baja. Los resultados se ilustran en la Figura 6.

Figura 6. Matriz de correlación y dendrograma de factores



Fuente: elaboración propia

El análisis clúster, representado como un dendrograma (Figura 6), confirmó el alto grado de asociación entre la temperatura media de las mínimas invernales y la altitud, y entre la amplitud térmica y la temperatura media de las máximas, conformando dos familias de predictores emparentados. Entre estos predictores similares, escogimos “tmin_cold” y “tmax_warm”, “tmean_range” y “p_spring”, debido a que son las variables que menos correlación mostraron entre las variables correlacionadas entre sí, y porque forman parte de dos ramas diferentes, aunque relativamente próximas de la familia de predictores. Y aunque algo correlacionadas con la precipitación, también escogimos el “pH”. El resto de predictores no correlacionados se situaron en ramas con una gran distancia de similitud. Finalmente, para disipar todo indicio de multicolinealidad que, como indicábamos en el apartado de metodología, también podría estar provocada porque una variable fuera una combinación lineal de varias variables, calculamos el VIF para las variables seleccionadas en el paso anterior (Tabla 3).

Tabla 3. Análisis VIF de factores y selección de variables definitivas

Variable	VIF	Diagnóstico
<i>pH</i>	1,42	Variable significativa sin colinealidad / Seleccionada
solar	1,12	No tiene colinealidad, pero no es significativa / Excluida
slope	1,12	No tiene colinealidad, pero no es significativa / Excluida
tmean_range	10,3	Problemas de multicolinealidad / Excluida
tmax_warm	1,25	Variable significativa sin colinealidad / Seleccionada
tmin_cold	1,14	Variable significativa sin colinealidad / Seleccionada
p_spring	1,84	Variable significativa sin colinealidad / Seleccionada

Fuente: elaboración propia

Los valores de VIF sólo detectaron multicolinealidad en “tmean_range”. Excluida esta variable, usamos todas las demás variables no correlacionadas para construir los modelos candidatos. Se trata de que cada variable no correlacionada muestre una faceta diferente (mínimamente redundante) de la respuesta de la variable dependiente que intentamos modelizar.

En el siguiente paso, desarrollamos diferentes Modelos Aditivos Generalizados (GAM) usando todas las variables sin problemas de multicolinealidad para seleccionar el mejor modelo, y explicitamos, además, distintas relaciones entre factores. Añadimos una variable categórica, la litología, para ofrecer información agregada sobre el sustrato más allá del factor *pH*, que en realidad es una variable en parte relacionada con las precipitaciones y la vegetación, junto con la existencia de determinadas clases litológicas. Los resultados (Tabla 4) mostraron cómo el mejor modelo es al mismo tiempo el más simple. El modelo “gam5” explicó el 79 % de la varianza con el AIC más bajo, usando cinco variables donde mediaron dos interacciones.

Tabla 4. Modelos de regresión GAM

Modelos Aditivos Generalizados (GAMs)	D ² explicada	AIC
gam1 = qsuber ~ p_spring+tmax_warm+tmin_cold+ph+factor(lito)+solar+slope	73,1	202
gam2 = qsuber ~ p_spring+tmax_warm+tmin_cold+ph+factor(lito)	73	201
gam3 = qsuber ~ p_spring*tmax_warm+tmin_cold+ph+factor(lito)	73,7	183
gam4 = qsuber ~ p_spring+tmax_warm+tmin_cold+ph*factor(lito)	78,6	86
gam5 = qsuber ~ p_spring*tmax_warm+tmin_cold+ph*factor(lito)	79	74

Nota: las interacciones del modelo se expresan con el operador (*)

Fuente: elaboración propia

Por último, la distribución de los errores del mejor modelo se ajustó a la normalidad, como se aprecia en el histograma de residuos y el Gráfico QQ (Figura 7). Por tanto, empleamos el mejor modelo general para ejecutarlo con todos los algoritmos: GLM, GAM, Maxent y Random Forest.

Figura 7. Distribución de los residuos del modelo. Histograma y QQPLOT

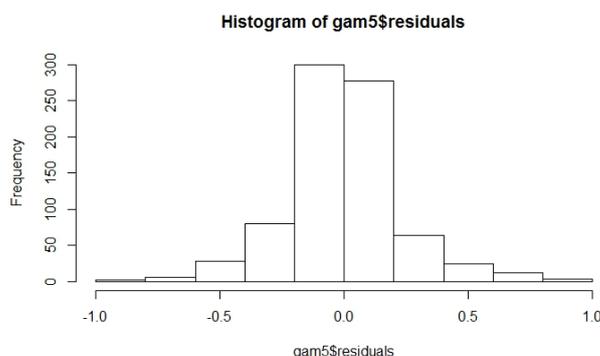
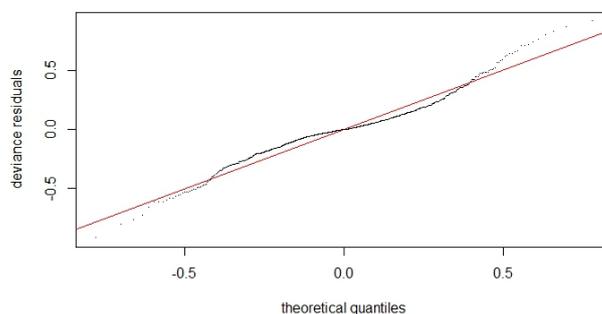


Figura 7. Continuación



Fuente: elaboración propia

3.4 Capacidad discriminante de los modelos finales

Todos los algoritmos ofrecieron una buena capacidad discriminante (Tabla 5). Según el AUC, obtuvimos modelos excelentes, con escasas diferencias entre sí. En conjunto, los mejores resultados se obtuvieron con GAM y Random Forest, y los peores resultados con Maxent, un algoritmo que suele destacar sobre el resto cuando el muestreo es incompleto y sesgado, no como el nuestro, que es sistemático y exhaustivo. Pero el AUC es realmente informativo con ausencias reales y se considera inadecuado cuando el objetivo no es el nicho realizado (Jiménez-Valverde, 2012). Cuando ejecutamos los algoritmos en las distintas iteraciones, como también se aprecia en el dibujo de la curva ROC (Figura 8) con los trazos (en azul) de cada réplica, se proyectó un intervalo de confianza muy ajustado al promedio. En este sentido, la replicación apenas alteró la capacidad discriminante de los modelos según las otras métricas empleadas, algo más en Random Forest según COR y TSS. Por tanto, estos resultados apoyan el acierto de la estrategia de muestreo y selección de las variables predictoras. Los coeficientes de correlación obtenidos por los mejores modelos fueron altos ($>0,8$), rozando la correlación muy alta con GAM y Random Forest. Los resultados obtenidos con el TSS siguieron el mismo patrón. En contraste con los valores de AUC, que fueron excelentes en todos los algoritmos, se obtuvieron valores proporcionalmente peores de COR y TSS con Maxent. Finalmente, los modelos con la devianza más reducida fueron los ejecutados con Random Forest y GLM.

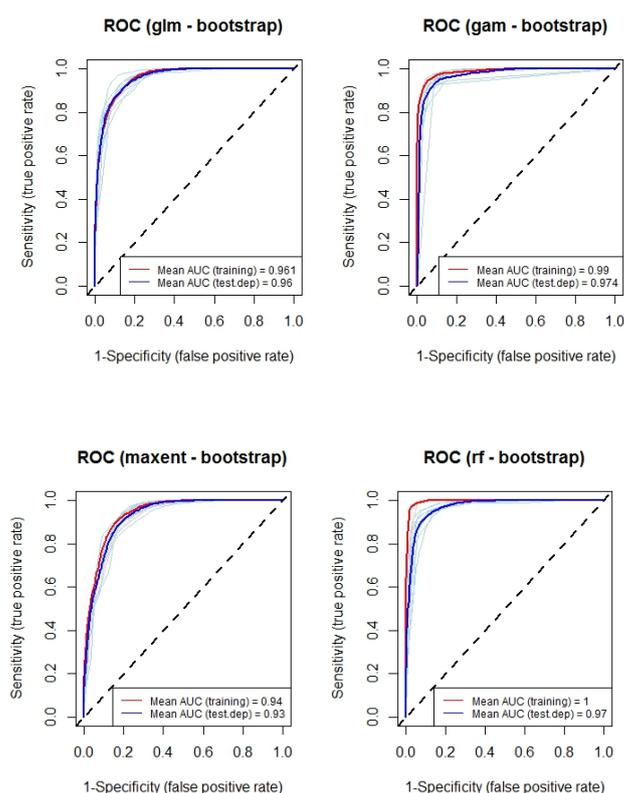
Tabla 5. Métricas de evaluación de los modelos

Modelo	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
Algoritmo / Parámetro	<i>AUC</i>	<i>AUC</i>	<i>AUC</i>	<i>COR</i>	<i>COR</i>	<i>COR</i>	<i>TSS</i>	<i>TSS</i>	<i>TSS</i>	<i>D</i>	<i>D</i>	<i>D</i>
GLM	0,96	0,95	0,96	0,82	0,80	0,83	0,77	0,75	0,80	0,52	0,58	0,51
GAM	0,99	0,96	0,97	0,90	0,83	0,89	0,88	0,81	0,88	0,31	0,67	1,07
Maxent	0,93	0,93	0,93	0,76	0,74	0,76	0,74	0,74	0,74	0,91	0,93	0,93
Random Forest	1	0,94	0,97	0,95	0,79	0,86	0,99	0,74	0,86	0,27	0,64	0,48
\bar{x}	0,97	0,95	0,96	0,86	0,79	0,84	0,85	0,76	0,82	0,50	0,70	0,75
σ	0,03	0,01	0,02	0,08	0,04	0,06	0,11	0,03	0,06	0,29	0,15	0,30

Nota: M1: modelo evaluado con datos de entrenamiento, sin particiones; M2: modelo evaluado con una partición, 50 % datos de entrenamiento y 50 % datos para evaluación; M3: promediado de 10 modelos evaluados con particiones, 50 % datos de entrenamiento y 50 % datos para evaluación, en cada iteración se extrajeron los datos aleatoriamente (*bootstrap*).

Fuente: elaboración propia

Figura 8. Curva ROC de los modelos



Fuente: elaboración propia

En este punto, hemos de llamar la atención sobre el término validación. A menudo se emplea con un uso que no es el estrictamente correcto, porque es imposible validar integralmente el modelo numérico de un sistema natural (Oreskes et al., 1994). Los modelos son siempre aproximaciones y su valor es eminentemente heurístico. Siguiendo a Oreskes (2010), nosotros abogamos por el uso

de la expresión “evaluación de modelos”, porque, a diferencia de la lógica del lenguaje implícita en el concepto de validación, que cierra el debate en torno a la incertidumbre, la discusión científica debe estar abierta hacia el campo de la medición y predicción sistemática del error (Oreskes, 2010).

3.5 Cartografías de distribución potencial

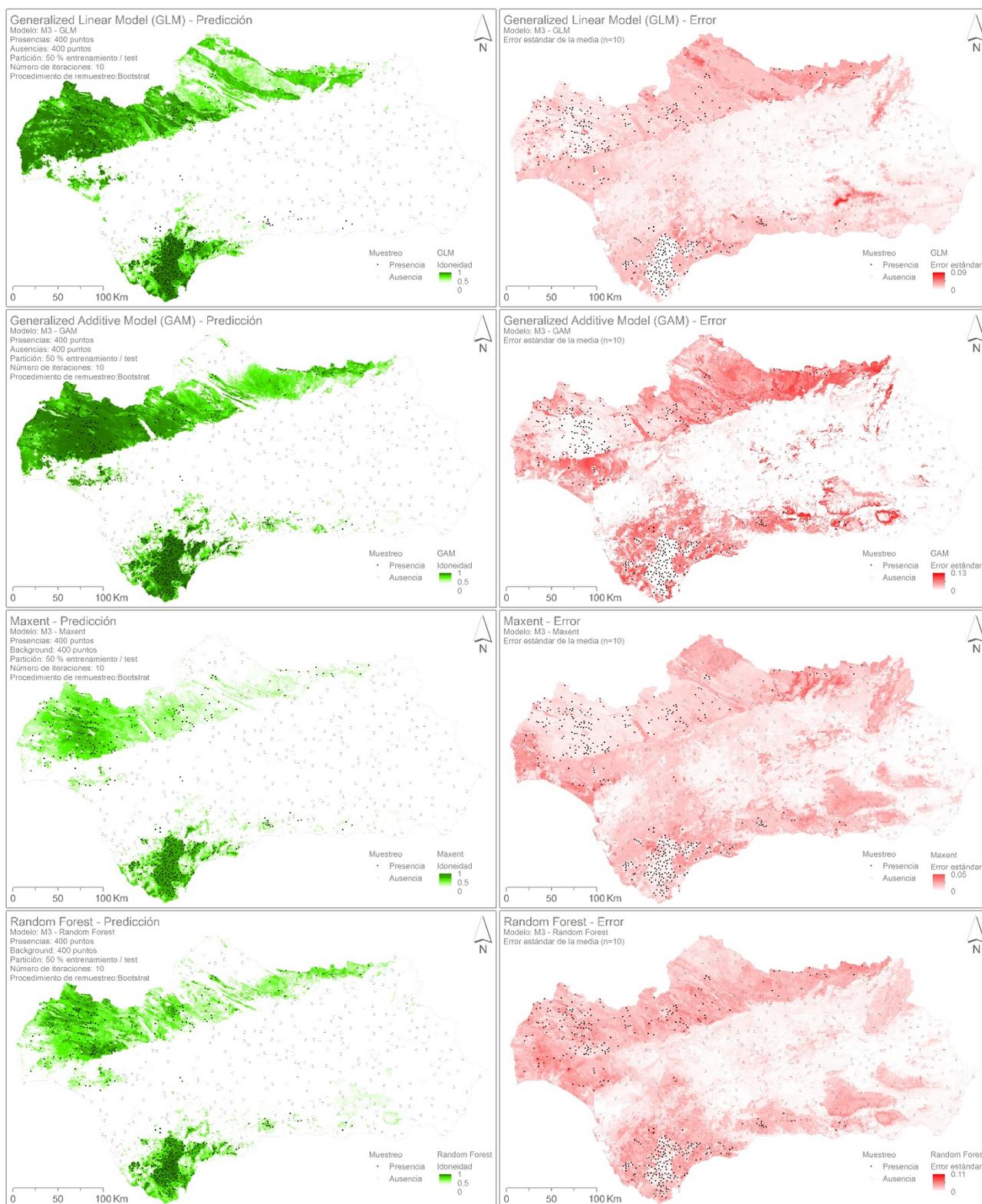
En el presente trabajo, las predicciones que incluyen los mapas representan la distribución potencial del alcornoque en términos de idoneidad (*suitability*), parámetro que expresa el grado en que cada localización (píxel) es más (1) o menos (0) idónea para la especie desde el punto de vista de su nicho ecológico fundamental. Es importante aclarar este aspecto porque en muchas publicaciones las predicciones se refieren a erróneamente a “probabilidad de presencia” de la especie objeto de estudio.

Realmente, las predicciones que no analizan registros con ausencias reales, no informan *sensu stricto* sobre probabilidad de presencia. No es posible predecir dicha probabilidad sin conocer la prevalencia de la especie. A decir verdad, cuando hablamos de probabilidad, hemos de hacerlo siempre en términos de aproximación, porque no controlamos todos los parámetros que la determinan. Y toda buena aproximación debe partir de un muestreo sistemático y balanceado, si es posible con ausencias fiables, puesto que tanto un muestreo sesgado como una mala calidad de los datos de ausencia —ya de por sí problemáticos—, nos alejarían de la idea de probabilidad de presencia (Lobo et al. 2010). Aunque preferimos usar el término “idoneidad” en los mapas, optimizamos el muestreo para que este parámetro se acercara lo más posible a la idea de probabilidad de presencia.

La Figura 9 muestra la predicción espacial proyectada por cada algoritmo aplicado, tras promediar el resultado de diez iteraciones basadas en el remuestreo aleatorio de los datos (*bootstrap*), así como la distribución espacial del error estándar de la media de cada predicción.

En el presente apartado, razonaremos acerca del modo en que se ha proyectado la distribución de acuerdo con la naturaleza de los datos y los algoritmos empleados. Seguidamente, discutiremos acerca de los matices encontrados en nuestras predicciones y también por las realizadas por otros autores sobre la misma especie, el alcornoque. Finalmente, nos centraremos en la explicación de la distribución del nicho ecológico fundamental del alcornoque en Andalucía, en contraste con el nicho realizado.

Figura 9. Modelos de distribución potencial del alcornoque. Predicción y error



Fuente: elaboración propia

Al analizar las predicciones en función de los datos y métodos empleados, detectamos –en primer lugar– una sobrepredicción del área de distribución potencial de la especie cuando se utilizaron métodos de regresión basados en presencia/ausencia con respecto a los métodos de aprendizaje

automático basados en presencia/background. Por el contrario, los métodos basados en presencia/background y, en especial, Maxent, infraestimaron tanto la distribución del nicho como los valores máximos de idoneidad en el núcleo de las principales poblaciones de *Quercus suber* existentes en la región. Este efecto ha sido detectado también por otros autores (Royle et al., 2012). GLM incurrió en más errores de comisión, influenciado por la escasez de registros de ausencia en determinadas clases de la única variable categórica utilizada, la litología, por lo que el desempeño de su predicción espacial resultó menos coherente con el conocimiento que se tiene sobre la distribución real de la especie. GAM y Random Forest mejoraron la predicción espacial, reduciendo los falsos positivos y ajustándose más a los cambios. Esto es así porque ambos algoritmos están diseñados para adaptarse empíricamente a los patrones observados, mientras que GLM es obsecuente respecto a la manera en que el usuario determina la función de vínculo y el polinomio asociado. En la depresión del Guadalquivir, casi todos los modelos –excepto Random Forest– incurrieron en un error de omisión, al no detectar potencialidad en el pequeño enclave de La Mocheta-El Arahal (Sevilla). Por el contrario, en el suroeste de la provincia de Huelva, entre las comarcas de la Costa Occidental y Andévalo, los modelos basados en presencia y ausencia, proyectaron más áreas favorables que los basados en presencia/background –en los que aparecieron más reducidas y fragmentadas–, incurriendo en notables errores de comisión, alguno de los cuales merece la pena detallar a continuación.

López-Tirado e Hidalgo (2016), trabajaron con registros de presencia y ausencia aplicando un modelo de regresión logística. Sus resultados proyectaron una elevada idoneidad para la especie en amplias zonas del manto eólico de Doñana (Huelva) y Sierra Bermeja (Málaga). Pero, con respecto al entorno de Doñana, el alcornoque no suele prosperar en zonas donde se produce un prolongado periodo de encharcamiento, pero tampoco en zonas secas y arenosas en las que el nivel freático está muy alejado de la superficie. Y en el caso de Sierra Bermeja, donde los autores también proyectaron amplias áreas favorables para la especie debido al uso como predictor del pH , el alcornoque sólo aparece en los sustratos ácidos situados en el perímetro de la aureola de metamorfismo del núcleo peridotítico, y dentro de éste, sólo en enclaves muy puntuales por la concurrencia de la presencia materiales de arrastre o la acción de procesos geomorfoedáficos. Estimamos que este error de comisión se produce porque los autores únicamente consideraron el pH como factor de control edáfico, pero no contemplaron el efecto general de la litología como variable categórica, que en el caso de Sierra Bermeja incluye un factor de fitotoxicidad que es incompatible con la supervivencia del alcornoque. Por el contrario, en nuestros modelos introdujimos una interacción entre las variables pH (continua) y litología (categórica), lo que mejoró la calidad y coherencia de las predicciones. No obstante, en nuestro caso, también se aprecia una tendencia hacia el error de comisión en Sierra Bermeja dentro del modelo GLM (que es una regresión logística). Esto se explica porque se ha usado un número muy reducido de ausencias

dentro de las peridotitas y porque el resto de factores empleados determinaban una alta idoneidad para la especie en esta extensa sierra malagueña. En estos casos, no recomendamos producir un desajuste del muestreo localmente (e.g. agregando más ausencias en Sierra Bermeja), sino considerar el uso de máscaras en el post-proceso del modelo.

Más allá de la interpretación de las métricas globales, es importante hacer una lectura correcta de los mapas para descubrir posibles incongruencias geográficas en las predicciones. Aquí empleamos el término geográfico en una acepción más amplia que el sentido meramente espacial, como producto de la integración de factores que proyectan una distribución espacial. Y en este sentido, hemos de abordar conjuntamente predicción y error. Con respecto a los errores, en general, los mapas mostraron un error promedio muy bajo ($SE\bar{x} = 0,01$), con valores más elevados ($SE\bar{x} > 0,10$) en las zonas con menor idoneidad para la especie. Los modelos de regresión basados en presencia/ausencia minimizaron el error en la envolvente de los registros de presencia y en los espacios más áridos y continentalizados. Debemos considerar que tanto la predicción como el error están influenciados por la forma en que cada algoritmo interacciona con las variables, pero también por el error heredado de los registros de presencia/ausencia y los predictores. En nuestro caso, empleamos predictores procedentes de muy diversas escalas de captura (aunque homogeneizáramos a una malla de 250 m) y diferentes metodologías. En este sentido, es fundamental una aproximación interdisciplinar para lograr un desarrollo óptimo de estos modelos. Se requieren inventarios sistemáticos de flora y fauna para contar con buenos datos de presencias y ausencias. También se requiere un conocimiento de las bases ecofisiológicas (derivadas de estudios experimentales y observacionales) para fundamentar nuestras hipótesis sobre los factores que controlan el nicho y la distribución (Mod et al., 2016).

A grandes rasgos, encontramos las áreas más idóneas para *Quercus suber* entre el parque natural de Los Alcornocales (Cádiz-Málaga) y el valle del Genal (Málaga). Se trata de un espacio donde los alcornoques forman masas boscosas muy espesas, territorios donde, además, la especie ha tenido un intenso aprovechamiento derivado de uno de sus principales productos: el corcho. El hombre ha favorecido la densidad del bosque en los enclaves donde la especie encuentra las mejores condiciones para producir biomasa, y ello ocurre en las áreas más lluviosas y templadas. Por ello, debe interpretarse la presencia y distribución del alcornoque en estas zonas como un requerimiento ecológico, aunque su abundancia está muy condicionada por el manejo del hombre, el cual ha encontrado en estas áreas las condiciones bioclimáticas idóneas para incrementar la densidad de las masas, en aras de una mayor productividad.

Otros espacios en los que el alcornoque encuentra condiciones ambientales idóneas son las sierras situadas al norte de la provincia Huelva, particularmente en el Parque Natural de la Sierra de Aracena y Picos de Aroche. Se trata de una zona con unas características parecidas a las que encontramos en Los Alcornocales, también lluviosa, pero con una mayor continentalidad según

avanzamos hacia el norte. Esta continentalidad se hace extrema hacia el norte y este de la distribución de la especie en la región, por lo que el alcornoque aparece formando nichos ecológicos de pertinencia borrosa, ya que se encuentra con zonas más o menos favorables donde la competencia con la encina produce un avance de esta última con respecto al alcornoque. Nuevamente, si analizáramos la distribución real de la especie, estaríamos también ante una respuesta dual, ecológica y antrópica, pues en estas áreas no tan favorables para el alcornoque como las anteriormente mencionadas, el hombre ha favorecido a la encina (por la calidad de su bellota) en un contexto geográfico e histórico donde siguen primando las actividades ganaderas y cinegéticas.

También encontramos una importante área favorable en la comarca malagueña de la Axarquía donde la especie está reducida a pequeños bosquetes e individuos aislados debido a que este ha sido un territorio cultivado históricamente con olivos, vides y almendros. El modelo, que dispuso de pocos registros de presencia para esta zona, proyectó un área favorable para el establecimiento de la especie que podría recomendar una reintroducción del alcornoque en la zona. Hacia el extremo oriental de la región, hallamos valores de idoneidad para el alcornoque muy bajos, si bien los modelos también detectaron pequeños reductos con una favorabilidad media-baja en algunas sierras litorales granadinas y almerienses, especialmente en Lújar (donde persiste una población) y Gádor.

La distribución del nicho ecológico fundamental del alcornoque, como es normal en entornos humanizados, es mayor que la distribución actual de la especie, si bien, a diferencia de otras especies del género *Quercus*, como la encina, el nicho ecológico del alcornoque está mejor representado por su distribución actual en la Comunidad Autónoma de Andalucía. Esta circunstancia se debe a que la alcornoque medra en suelos muy pobres, que suelen ser los que no han sido convertidos históricamente por las actividades agrarias. La litología es un factor de capital importancia en la separación entre los usos del suelo agrarios y forestales (Gutiérrez-Hernández et al., 2016). En conjunto, es la combinación de factores bioclimáticos y litológicos la que explica la proyección espacial del nicho ecológico fundamental del alcornoque. Y estos factores, junto con la intervención de las actividades antrópicas (Urbieto et al., 2008) y otras interacciones bióticas (Ibáñez et al., 2014), en las que no hemos profundizado, determinan por qué el nicho ecológico fundamental coincide en buena medida con la distribución actual (nicho realizado) de los alcornocales andaluces.

4 Conclusiones

En este trabajo, hemos modelizado el nicho ecológico fundamental del alcornoque a partir de un muestreo sistemático de registros de presencia/ausencia y registros de presencia/background. Hemos examinado los datos con rigor estadístico y conocimiento experto, integrando diversas

fuentes desde la perspectiva de las ciencias geográficas. Preparamos las variables ambientales, exploramos las relaciones causales entre las mismas y los datos de ocurrencias, y calibramos modelos más explicativos y parsimoniosos. Evaluamos la capacidad discriminante de los modelos basados en datos remuestreados aleatoriamente en diez réplicas por cada algoritmo (GLM, GAM, Maxent, Random Forest). Y finalmente, proyectamos los resultados en el espacio geográfico y discutimos las bondades y limitaciones de las predicciones y sus errores asociados.

En el estudio de caso, revelamos que la distribución potencial del alcornoque está condicionada por la preferencia de la especie hacia suelos ligeros o moderadamente ácidos ($\bar{x}=6,5$ pH; $\sigma=3,3$ pH) sobre sustratos silíceos (areniscas, pizarras), un ombrotipo subhúmedo-húmedo ($\bar{x}=881$ mm; $\sigma=181$ mm), una amplitud térmica reducida ($\bar{x}=15,4$ °C; $\sigma=1,9$ °C), dentro de un clima templado con unas medias de las mínimas ($\bar{x}=5,2$ °C; $\sigma=2,1$ °C) por encima del umbral de las heladas. Esta distribución potencial del alcornoque contrasta —aunque no tanto como en otras especies— con la distribución actual, la cual es producto de un sistema de reciprocidad entre las condiciones ecológicas del medio y las actividades humanas de aprovechamiento de los recursos. He aquí un nuevo campo de estudio en el que la Biogeografía puede incorporar el enfoque holístico de la Geografía para modelizar los mecanismos causales que subyacen entre las variantes de los nichos y su dinámica espaciotemporal.

Hemos puesto de manifiesto la utilidad de la inferencia estadística y la modelización predictiva de nichos ecológicos para su aplicación en trabajos de Biogeografía en el contexto de la Geografía Física sobre los factores que controlan la distribución del nicho ecológico fundamental de las especies. A su vez, revelamos cómo la Geografía y su enfoque integrador puede contribuir a optimizar la metodología, en un campo donde no hay estándares cerrados ni “balas de plata” que solucionen todos los problemas desde la perspectiva de los modelos numéricos (Qiao et al., 2015). En el contexto más amplio de los modelos de distribución de especies, donde también se contemplan interacciones bióticas y factores biogeográficos para explicar el nicho ecológico y la distribución de las especies en un contexto de cambio global, la Geografía se halla en el meollo de la discusión, por su papel para desentrañar la complejidad de las actividades humanas en las redes de reciprocidad del planeta.

Agradecimientos: Todos los autores de este artículo tuvimos la ocasión de conocer a José Manuel Rubio Recio, precursor de la Biogeografía en España. Este trabajo, que pretende contribuir a una renovación metodológica de la Biogeografía, especialmente en España, está dedicado a su memoria.

Declaración responsable: Las/os autoras/es declaran que no existe ningún conflicto de interés en relación a la publicación de este artículo. Las tareas se han distribuido de la siguiente manera:

Oliver Gutiérrez Hernández, diseño de la investigación y redacción del manuscrito original; Luis V. García, Rafael Cámara Artigas y José María Senciales González: revisión crítica, correcciones y aportaciones sobre aspectos relacionados con el diseño, análisis de datos, redacción, biogeografía y ecología.

Bibliografía

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723. doi: <http://dx.doi.org/10.1109/TAC.1974.1100705>
- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the True Skill Statistic. *Journal of Applied Ecology*, 43, 1223–1232. doi: <http://dx.doi.org/10.1111/j.1365-2664.2006.01214.x>
- Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1–19. doi: <http://dx.doi.org/10.1016/j.ecolmodel.2006.07.005>
- Austin, M. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101–118. doi: [http://dx.doi.org/10.1016/S0304-3800\(02\)00205-3](http://dx.doi.org/10.1016/S0304-3800(02)00205-3)
- Barbosa, F. G., & Schneck, F. (2015). Characteristics of the top-cited papers in species distribution predictive models. *Ecological Modelling*, 313, 77–83. doi: <http://dx.doi.org/10.1016/j.ecolmodel.2015.06.014>
- Benito, B. M., Lorite, J., Pérez-Pérez, R., Gómez-Aparicio, L., & Peñas, J. (2014). Forecasting plant range collapse in a mediterranean hotspot: When dispersal uncertainties matter. *Diversity and Distributions*, 20, 72–83. doi: <http://dx.doi.org/10.1111/ddi.12148>
- Benito, B. M., Svenning, J. C., Kellberg-Nielsen, T., Riede, F., Gil-Romera, G., Mailund, T., Kjaergaard, P. C., & Sandel, B. S. (2016). The ecological niche and distribution of Neanderthals during the Last Interglacial. *Journal of Biogeography*, 44(1), 51–61. doi: <http://dx.doi.org/10.1111/jbi.12845>
- Bowles, G. (1782). *Introducción a la historia natural y a la geografía física de España*. Madrid: Imprenta Real.
- Box, E. O. (1981). Predicting physiognomic vegetation types with climate variables. *Vegetatio*, 45, 127–139. doi: <http://dx.doi.org/10.1007/BF00119222>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. doi: <http://dx.doi.org/10.1023/A:1010933404324>
- Bugalho, M. N., Caldeira, M.C., Pereira, J. S., Aronson, J., & Pausas, J. G. (2011). Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. *Frontiers in Ecology and the Environment*, 9, 278–286. doi: <http://dx.doi.org/10.1890/100084>
- Carrero Carrero, A. J. (2015). *Los aprovechamientos del monte en el Suroeste Peninsular: el corcho*. Huelva: Universidad de Huelva.

- Coelho, M. B., Paulo, J. A., Palma, J. H. N., & Tomé, M. (2012). Contribution of cork oak plantations installed after 1990 in Portugal to the Kyoto commitments and to the landowners economy. *Forest Policy and Economics*, 17, 59–68. doi: <http://dx.doi.org/10.1016/j.forpol.2011.10.005>
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- De La Hoz Rodríguez, F. (2009). *Procesos de decaimiento forestal (la Seca): situación del conocimiento*. Consejería de Medio Ambiente. Sevilla: Junta de Andalucía.
- Díaz Esteban, M., Pulido Díaz, F. J., & Pausas, J. G. (2009). 9330 *Alcornocales de Quercus suber*. Madrid: MAGRAMA, Dirección General de Medio Natural y Política Forestal. Retrieved from <https://www.uv.es/jgpausas/papers/Diaz-2009-habitats.pdf>
- Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., ... Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30, 609–628. doi: <http://dx.doi.org/10.1111/j.2007.0906-7590.05171.x>
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68, 589–599. doi: <http://dx.doi.org/10.1093/biomet/68.3.589>
- Elith, J., & Graham, C.H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32, 66–77. doi: <http://dx.doi.org/10.1111/j.1600-0587.2008.05505.x>
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17, 43–57. doi: <http://dx.doi.org/10.1111/j.1472-4642.2010.00725.x>
- Elton, C. (1927). *Animal Ecology*. London: Sedgwick and Jackson.
- Ferreras Chasco, C., & Fidalgo Hijano, C. (1991). *Biogeografía y edafogeografía*. Madrid: Síntesis.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38–49. doi: <http://dx.doi.org/10.1017/S0376892997000088>
- Fisher, H. J. (2002). Macroecology: new, or biogeography revisited? *Nature*, 417, 787. doi: <http://dx.doi.org/10.1038/417787c>

- Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, *19*, 474–499. doi: <http://dx.doi.org/10.1177/030913339501900403>
- Gallien, L., Münkemüller, T., Albert, C. H., Boulangeat, I., & Thuiller, W. (2010). Predicting potential distributions of invasive species: where to go from here? *Diversity and Distributions*, *16*, 331–342. doi: <http://dx.doi.org/10.1111/j.1472-4642.2010.00652.x>
- Goloboff, P. A. (2003). Parsimony, likelihood, and simplicity. *Cladistics*, *19*, 91–103. doi: <http://dx.doi.org/10.1111/j.1096-0031.2003.tb00297.x>
- Gómez-Zotano, J. (2004). The broadleaved tree-conifer controversy at Sierra Bermeja, an ultramafic mountain in southern Spain. In R. S. Boyd, A. J. M. Baker, & J. Proctor (Eds.), *Ultramafic rocks: their soils, vegetation and fauna* (pp. 151–156). Herts: Science Reviews.
- Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*, 2809–2815. doi: <http://dx.doi.org/10.1890/02-3114>
- Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *Auk*, *34*, 427–433. doi: <http://dx.doi.org/10.2307/4072271>
- Guisan, A., & Edwards, T. C. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, *157*, 89–100. doi: [http://dx.doi.org/10.1016/S0304-3800\(02\)00204-1](http://dx.doi.org/10.1016/S0304-3800(02)00204-1)
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009. doi: <http://dx.doi.org/10.1111/j.1461-0248.2005.00792.x>
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186. doi: [http://dx.doi.org/10.1016/S0304-3800\(00\)00354-9](http://dx.doi.org/10.1016/S0304-3800(00)00354-9)
- Gutiérrez-Hernández, O., Senciales-González, J. M., & García, L. V. (2016). Evolución de la Superficie Forestal en Andalucía (1956–2007). Procesos y Factores. *Revista de Estudios Andaluces*, *33*, 111–148. doi: <http://dx.doi.org/10.12795/rea.2016.i33>
- Hanley, J. A., & Mcneil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29–36. doi: <http://dx.doi.org/10.1148/radiology.143.1.7063747>
- Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Stat. Scienc.*, *1*, 297–310. doi: <http://dx.doi.org/10.1214/ss/1177013604>
- Hijmans, R. J., Phillips, S., Leathwick, J. R., & Elith, J. (2011). dismo: Species Distribution Modeling [Functions package]. Retrieved from <https://cran.r-project.org/web/packages/dismo/index.html>

- Houston Durrant, T., De Rigo, D., & Caudullo, G. (2016). *Quercus suber* in Europe: distribution, habitat, usage and threats. In *European Atlas of Forest Tree Species* (pp. e01ff11+). Luxembourg: Publication Office of the European Union.
- Humboldt, A. V. (1805). *Essai sur le Géographie des Plantes*. Paris: Schoell, Lebrault & CO.
- Hutchinson, G. (1978). *An Introduction to Population Ecology*. New Haven: Yale University Press.
- Hutchinson, G. (1957). Concluding remarks. *Cold Spring Harb Symposia on Quantitative Biology*, 22, 415–427. doi: <http://dx.doi.org/10.1101/SQB.1957.022.01.039>
- Ibáñez, B., Ibáñez, I., Gómez-Aparicio, L., Ruiz-Benito, P., García, L. V., & Marañón, T. (2014). Contrasting effects of climate change along life stages of a dominant tree species: The importance of soil-climate interactions. *Diversity and Distributions*, 20, 872–883. doi: <http://dx.doi.org/10.1111/ddi.12193>
- Instituto de Estadística y Cartografía de Andalucía (2017). Datos Espaciales de Referencia de Andalucía (DERA) [Bases cartográficas]. Retrieved from <https://www.juntadeandalucia.es/institutodeestadisticaycartografia/DERA/>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning, with Applications in R*. Springer. New York: Springer. doi: <http://dx.doi.org/10.1007/978-1-4614-7138-7>
- Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21, 498–507. doi: <http://dx.doi.org/10.1111/j.1466-8238.2011.00683.x>
- Jiménez-Valverde, A., Acevedo, P., Barbosa, A. M., Lobo, J. M., & Real, R. (2013). Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecology and Biogeography*, 22, 508–516. doi: <http://dx.doi.org/10.1111/geb.12007>
- Jiménez-Valverde, A., & Lobo, J. M. (2006). The ghost of unbalanced species distribution data in geographical model predictions. *Diversity and Distributions*, 12, 521–524. doi: <https://doi.org/10.1111/j.1366-9516.2006.00267.x>
- Joffre, R., Rambal, S., & Ratte, J. P. (1999). The dehesa system of southern Spain and Portugal as a natural ecosystem mimic. *Agroforestry Systems*, 45, 57–79. doi: <https://doi.org/10.1023/A:1006259402496>
- Kent, M. (2007). Biogeography and macroecology: now a significant component of physical geography. *Progress in Physical Geography*, 31, 643–657. doi: <http://dx.doi.org/10.1177/0309133307087088>

- Kent, M. (2005). Biogeography and macroecology. *Progress in Physical Geography*, 29, 256–264. doi: <http://dx.doi.org/10.1191/0309133305pp447pr>
- Lobo, J. M. (2008). More complex distribution models or more representative data? *Biodiversity Informatics*, 5, 14–19. doi: <http://dx.doi.org/10.1093/molbev/msm137>
- Lobo, J. M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33, 103–114. doi: <https://doi.org/10.1111/j.1600-0587.2009.06039.x>
- Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 145–151. doi: <http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x>
- Lomolino, M. V., Riddle, B. A., & Whittaker, R. J. (2016). *Biogeography*. Sinauer Associates.
- López-Tirado, J., & Hidalgo, P. J. (2016). Predictive modelling of climax oak trees in southern Spain: insights in a scenario of global change. *Plant Ecology*, 217, 451–463. doi: <http://dx.doi.org/10.1007/s11258-016-0589-6>
- McInerney, G. J., & Etienne, R. S. (2012). Stitch the niche –a practical philosophy and visual schematic for the niche concept. *Journal of Biogeography*, 39, 2103–2111. doi: <http://dx.doi.org/10.1111/jbi.12032>
- Mod, H. K., Scherrer, D., Luoto, M., & Guisan, A. (2016). What we use is not what we know: Environmental predictors in plant distribution models. *Journal of Vegetation Science*, 27(6), 1308–1322. doi: <http://dx.doi.org/10.1111/jvs.12444>
- Montero González, G., & Cañellas, I. M. (2002). *El alcornoque (Quercus suber. L). Manual de reforestación y cultivo*. Madrid: Mundi–Prensa.
- Naimi, B., & Araújo, M.B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39, 368–375. doi: <http://dx.doi.org/10.1111/ecog.01881>
- Naimi, B., Skidmore, A. K., Groen, T. A., & Hamm, N.A.S. (2011). Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of Biogeography*, 38, 1497–1509. doi: <http://dx.doi.org/10.1111/j.1365-2699.2011.02523.x>
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, 135, 370–384. doi: <http://dx.doi.org/10.2307/2344614>
- Nogués-Bravo, D. (2009). Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, 18, 521–531. doi: <http://dx.doi.org/10.1111/j.1466-8238.2009.00476.x>

- Nogués Bravo, D. (2003). El estudio de la distribución espacial en la biodiversidad: concepto y métodos. *Cuadernos de Investigación Geográfica*, 29, 67–82.
- Oreskes, N. (2010). Evaluation (Not Validation) Models. *Quantitative Environment*, 106, 1453–1460.
- Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263, 641–646. doi: <http://dx.doi.org/10.1126/science.263.5147.641>
- Parsons, J. J. (1962). The acorn-hog economy of the oak woodlands of southwestern Spain. *Geographical Review*, 52, 211–235. doi: <http://dx.doi.org/10.1126/science.15.370.195>
- Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A.T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34, 102–117. doi: <http://dx.doi.org/10.1111/j.1365-2699.2006.01594.x>
- Peña-Angulo, D., Brunetti, M., Cortesi, N., & Gonzalez-Hidalgo, J. C. (2016). A new climatology of maximum and minimum temperature (1951–2010) in the Spanish mainland: a comparison between three different interpolation methods. *International Journal of Geographical Information Science*, 30, 1–24. doi: <http://dx.doi.org/10.1080/13658816.2016.1155712>
- Pérez-Latorre, A. V. (1996). Restauración del alcornocal en Andalucía. Elección y manejo de especies. *Ecología*, 10, 9–19.
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Bastos Araujo, M. (2011). *Ecological niches and geographic distributions*. Princeton, NJ: Princeton University Press.
- Phillips, S. J., & Dudi, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161–175. doi: <http://dx.doi.org/10.1111/j.0906-7590.2008.5203.x>
- Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters*, 3, 349–361. doi: <http://dx.doi.org/10.1046/j.1461-0248.2000.00143.x>
- Qiao, H., Soberón, J., & Peterson, A.T. (2015). No silver bullets in correlative ecological niche modelling: Insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution*, 6, 1126–1136. doi: <https://doi.org/10.1111/2041-210X.12397>
- Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics*, 69, 274–281. doi: <http://dx.doi.org/10.1111/j.1541-0420.2012.01824.x>

- Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. (2012). Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, 3, 545–554. doi: <http://dx.doi.org/10.1111/2041-210X.12397>
- Saupe, E., Barve, V., Myers, C., Soberón, J., Barve, N., Hensz, C., Peterson, A., Owens, H., & Lira-Noriega, A. (2012). Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling*, 237, 11–22. doi: <http://dx.doi.org/10.1016/j.ecolmodel.2012.04.001>
- Serrasolses, I., Pérez-Devesa, M., Vilagrosa, A., Pausas, J. G., Sauras, T., Cortina, J., & Vallejo, V. R. (2009). Cork Oak Distribution. In *Cork Oak Woodlands on the Edge: conservation, adaptive management, and restoration* (pp. 89–99). Island Press.
- Silva Pérez, R. (2010). La dehesa vista como paisaje cultural. Fisonomías, funcionalidades y dinámicas históricas. *Ería*, 82, 143–157.
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10, 1115–1123. doi: <http://dx.doi.org/10.1111/j.1461-0248.2007.01107.x>
- Soberón, J., & Townsend Peterson, A. (2005). Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodivers Informatics*, 2, 1–10. doi: <http://dx.doi.org/10.17161/bi.v2i0.4>
- Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Journal of Economic Geography*, 46, 234–240. doi: <http://dx.doi.org/10.2307/143141>
- Toumi, L., & Lumaret, R. (1998). Allozyme variation in cork oak (*Quercus suber* L.): The role of phylogeography and genetic introgression by other Mediterranean oak species and human activities. *Theoretical and Applied Genetics*, 97, 647–656. doi: <http://dx.doi.org/10.1007/s001220050941>
- Tulloch, A. I. T., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K.M.P.M.B., ... Rhodes, J. R. (2016). Conservation planners tend to ignore improved accuracy of modelled species distributions to focus on multiple threats and ecological processes. *Biological Conservation*, 199, 157–171. doi: <http://dx.doi.org/10.1016/j.biocon.2016.04.023>
- Urbieto, I. R., García, L. V., Zavala, M. A., & Marañón, T. (2011). Mediterranean pine and oak distribution in southern Spain: Is there a mismatch between regeneration and adult distribution? *Journal of Vegetation Science*, 22, 18–31. doi: <http://dx.doi.org/10.1111/j.1654-1103.2010.01222.x>
- Urbieto, I. R., Zavala, M. A., & Marañón, T. (2008). Human and non-human determinants of forest composition in southern Spain: Evidence of shifts towards cork oak dominance as a result of

management over the past century. *Journal of Biogeography*, 35, 1688–1700.
doi: <http://dx.doi.org/10.1111/j.1365-2699.2008.01914.x>

Wallace R., A. (1876). *The Geographical Distribution Of Animals*. Harper Brother.
doi: <http://dx.doi.org/10.1086/271871>

Zheng, B., & Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Statistics in Medicin*, 19, 1771–1781. doi: [http://dx.doi.org/10.1002/1097-0258\(20000715\)19:13<1771::AID-SIM485>3.3.CO;2-G](http://dx.doi.org/10.1002/1097-0258(20000715)19:13<1771::AID-SIM485>3.3.CO;2-G)