



Universidad de Jaén

Centro de Estudios de Postgrado

Trabajo Fin de Máster

Modelos de Distribución Potencial de Especies Forestales en las Sierras de Jaén. Aplicaciones de alta resolución

Alumno/a: **María Guzmán Yera**

Tutor/a: Prof. D. Julio M. Alcántara

Prof. D. Antonio J. Manzaneda

Dpto: Biología Animal, Biología Vegetal y Ecología

INDICE

1.	INTRODUCCIÓN.....	3
2.	OBJETIVOS	7
3.	MATERIAL Y METODOS.....	7
	3.1 Área de trabajo	7
	3.2 Aspectos básicos sobre SDM	9
	3.3 Variables ambientales	11
	3.4 Procedencia de los datos de presencia/ausencia.....	18
	3.5 Elaboración de los SDMs y proyección de las áreas potenciales	21
	3.6 Calibración del modelo	24
4	RESULTADOS.....	27
5	DISCUSIÓN.....	39
6	CONCLUSIONES	43
7	REFERENCIAS.....	45

RESUMEN

El reciente desarrollo en los Sistemas de Información Geográfica (SIG) en las últimas décadas, ha promovido la expansión del uso de un método que permite modelar el nicho ecológico de una especie o comunidad en el espacio mediante distintos procedimientos matemáticos, los llamados Modelos de Distribución de Especies ("*Species distribution models*" SDMs). Estos modelos requieren de registros de presencia conocida de una especie que bien puede obtenerse de bancos de datos de biodiversidad o de la observación *in situ* de la especie en el área de estudio. En el presente trabajo se han empleado 6 métodos de ajuste de SDMs: BioCLim, DOMAIN, Distancia de Mahalanobis, LGM (General Linear Models), Random Forest y SVM (Support Vector Machines). Estos métodos se han aplicado a modelar la distribución potencial de 23 especies arbóreas nativas de las Sierras Béticas, empleándose 17 variables ambientales que caracterizan el clima, topografía y suelo. Los algoritmos SDM produjeron resultados satisfactorios para la práctica totalidad de las especies (valor de AUC >0.9). Los mejores métodos de modelación en este estudio fueron Random Forest y Mahalanobis, pero todos los métodos resultaron ser el mejor para alguna de las especies, por lo que no se puede descartar ninguno de ellos *a priori*. Los mapas de alta resolución obtenidos permiten identificar zonas de mayor riqueza potencial de especies, lo cual facilitaría la planificación de la gestión de especies a pequeña escala espacial, típica de los trabajos de carácter forestal.

Palabras clave: algoritmo, Modelo de Distribución de Especies, variables predictoras, nicho ecológico, idoneidad del hábitat, colinealidad, riqueza de especies, curva ROC.

ABSTRACT

The recent development of Geographic Information Systems (GIS) over the last decades has promoted the use of a technique that allows modeling the ecological niches of species and communities in space through the use of several mathematical procedures, the so called Species Distribution Models (SDMs). These models require the use of extensive datasets on the observed presence of species, which can be obtained nowadays from public, open data repositories or through planned surveys of biodiversity designed *ex profeso* for modeling purposes. In this study we have used 6 SDM techniques: BioCLim, DOMAIN, Distancia de Mahalanobis, LGM (General Linear Models), Random Forest y SVM (Support Vector Machines). We have applied these techniques to model the distribution of 23 tree species native to the Sierras Béticas, using data on 17 environmental variables that synthesize the climate, topography and soil properties of the study área. The SDM

algorithms rendered satisfactory results for most species ($AUC > 0.9$). The best ones proved to be Random Forests and Mahalanobis Distance, although all the algorithms proved the best for at least one of the modeled species. Therefore, it is recommended that all the algorithms should be used in order to optimize SDM studies. The high resolution maps obtained allow identifying areas of higher potential species richness, what can facilitate the planning of species management practices at small spatial scales typical of forestry activities.

Key words: algorithm, Species Distribution Models, predictors variables, ecological niche, habitat suitability, collinearity, species richness, ROC curve.

1. INTRODUCCIÓN

Establecer los patrones espaciales y temporales de la distribución de los seres vivos sobre el planeta y sus causas, ha sido desde sus comienzos el centro del estudio de la biogeografía. Alexander von Humboldt (1769-1859) es considerado el "padre de la biogeografía". La aparición de esta ciencia permitió instaurar una relación causal entre la distribución geográfica de las especies vegetales y la distribución espacial del clima (Pliscoff y Fuentes-Castillo, 2011). Posteriormente un ecólogo, Joseph Grinnell en 1917 fue el primero en desarrollar el concepto de "*nicho ecológico*", como el tipo de hábitats a los que una especie está especialmente adaptada, y su relación con la distribución geográfica de las especies. Pero, tras un movimiento revolucionario en la etapa evolutiva de dicho concepto (por el uso de terminología matemática más que la mera semántica descriptiva), George Evelyn Hutchinson define entre 1944 y 1958 al nicho como la suma de todas las variables ambientales que actúan sobre un organismo. El nicho es descrito entonces como un hipervolumen de la región del espacio n-dimensional de todos los factores que actúan sobre una especie (Figura 1). Hutchinson además establece el principio de exclusión competitiva y define el *nicho fundamental* como el intervalo de factores ambientales que le posibilita a una especie subsistir indefinidamente, es decir, todas las condiciones del hipervolumen n-dimensional en las que podría vivir la especie. Por su parte, el *nicho realizado* de una especie es aquella fracción del nicho fundamental en la cual la especie se halla en realidad, por tanto, la zona del hipervolumen al que se restringe la especie debido a interacciones bióticas. Estas definiciones han sido fundamentales en el desarrollo del concepto de Modelos de Distribución de Especies (SDMs por sus siglas en Ingles).

En la segunda mitad del s. XX, los rápidos avances metodológicos en geografía física y capacidad de computación proporcionaron nuevos datos e información de sistemas, permitiendo un desarrollo sólido y detallado de modelos digitales de elevación de la superficie terrestre, interpolación de parámetros climáticos y detección remota de las condiciones de la superficie de la Tierra. El desarrollo de Sistemas de Información Geográfica (SIGs) proporcionó herramientas para el almacenamiento y manipulación de registros de especies y datos ambientales. Uno de los pioneros en aplicar un enfoque integrado del uso de sistemas SIG para el modelado de la distribución de especies fue Ferrier (Ferrier, 1984), quien aplicó regresión logística para predecir la distribución del *Atrichornis rufescens* (un ave passeriforme endémica de Australia) utilizando la población conocida, registros antiguos de la especie y variables ambientales mapeadas y modeladas remotamente.

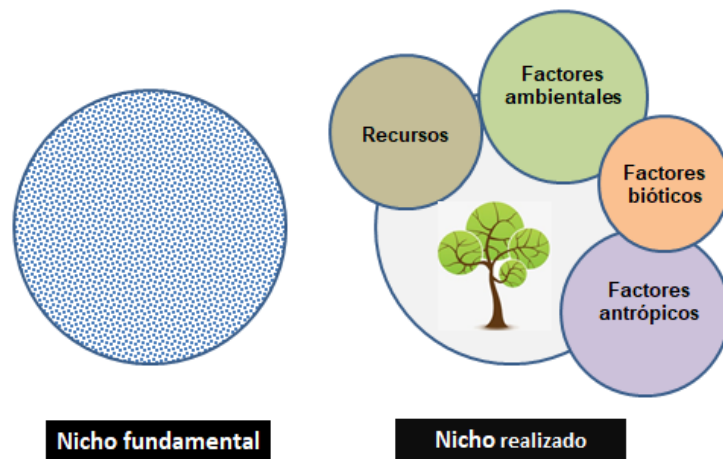


Figura 1. Representación gráfica de los modelos teóricos de nicho fundamental y nicho realizado. El nicho real o realizado describe la dimensión a la cual una especie se encuentra limitada por las interacciones bióticas, imposiciones antrópicas y disponibilidad de recursos que permiten a esta mantener una población viable. Fuente: elaboración propia.

Definimos un modelo como una representación parcial de la realidad que refleja algunas de sus propiedades. Los modelos son, por tanto, simplificaciones, debidas tanto a la necesidad de reducir la complejidad del objeto real como a nuestro desconocimiento de muchas de sus propiedades (Mateo *et al.*, 2011). Los SDMs, también conocidos como *modelos del nicho ambiental* (“*Environmental Niche Models*” ENM; entre otros muchos términos similares), indican la idoneidad de un hábitat para el desarrollo de las poblaciones de una especie o de una comunidad (Ferrier y Guisan, 2006). Para el cálculo de dicha idoneidad se emplean algoritmos o métodos estadísticos computarizados que definen las relaciones entre la presencia o abundancia conocida de la especie y un conjunto de variables ambientales (i.e. se hace una aproximación al nicho realizado). A partir de estas relaciones se pueden generar mapas predictivos sobre la distribución potencial de la especie en el espacio geográfico.

Los SDMs se han convertido en una de las áreas de investigación con mayor desarrollo en el campo de la biogeografía de la conservación (Richardson y Whittaker, 2010). Como tal, los SDM han permitido a los investigadores un instrumento innovador para explorar diversas cuestiones en ecología, evolución y conservación. Además de su uso para el estudio de las posibles relaciones entre los parámetros ambientales y la distribución de una especie, también se emplean en el análisis de la riqueza de especies (MacNally y Fleishman, 2004), las características y la configuración espacial de los hábitats que permiten la persistencia de especies en los paisajes (Araújo y Williams, 2000; Ferrier *et al.*, 2002;

Scotts y Drielsma, 2003), el potencial invasivo de especies no autóctonas (Peterson, 2003), distribuciones de especies en el pasado (Hugall et al., 2002) o bajo climas futuros (Thomas et al., 2004) y la diferenciación ecológica y geográfica de las distribuciones de especies estrechamente relacionadas (Graham et al., 2004b).

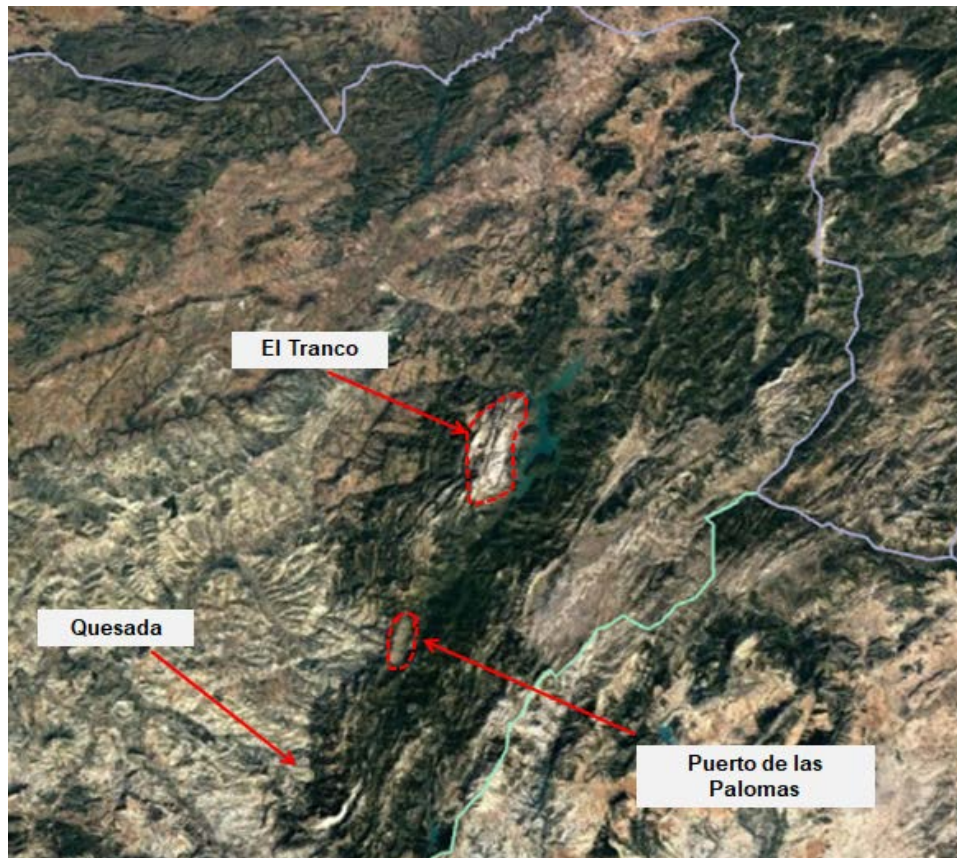
Los bosques ibéricos son contemplados como hábitats altamente vulnerables al cambio global y al crecimiento demográfico con su correspondiente actividad humana, especialmente en la región Mediterránea, lo que podría alterar el suministro de servicios ecosistémicos claves (Lindner et al., 2010). En las Cordilleras Béticas, donde se centra nuestra área de estudio, el estado actual de conservación de la diversidad vegetal es consecuencia directa de dos factores de amenaza básicos. Por un lado su posición geográfica y heterogeneidad ecológica, que dentro de un marco fundamentalmente mediterráneo, con una marcada sequía estival, viene determinada por variaciones edáficas y climáticas. Por otro lado, la humanización del territorio, que desde hace ya mucho tiempo se traduce en una fuerte utilización de la ganadería doméstica, una progresiva expansión de terrenos destinados a la agricultura (como el olivar y el cultivo bajo invernadero), y un elevado grado de degradación y contaminación de hábitats (eutrofización, residuos industriales, plaguicidas, residuos urbanos, pastoreo, lluvia ácida, etc.) (Cabezudo, 2005).

A estas amenazas se suma el papel de los incendios forestales en nuestra región. En nuestra zona de estudio se encuentra el Parque Natural de las Sierras de Cazorla, Segura y las Villas, el cual ha sido escenario de múltiples incendios importantes en las últimas décadas (Figura 2). Por ejemplo, según datos de la Red de Información Ambiental de la Junta de Andalucía (REDIAM: <http://www.juntadeandalucia.es/medioambiente/site/rediam>), en agosto de 2005 en la Sierra de las Villas se produjo un incendio que consumió un total de 5.116 hectáreas, el incendio del Puerto de las Palomas de agosto de 2001 devastó más de 800 hectáreas y más recientemente, en 2015, el incendio forestal de Quesada afectó a más de 10.000 hectáreas de alto valor ecológico. Esto suma un total de 15.916 hectáreas, sin contar otros pequeños incendios declarados en estos años intermedios.

Cuando es imposible la regeneración natural del medio, la reforestación del ecosistema debe hacerse por implantación artificial de una cubierta vegetal. Esto se hace aún más indispensable cuando la frecuencia de incendios es elevada, por lo que es necesario crear sistemas que tengan una gran capacidad de respuesta a estas perturbaciones, mediante su adaptación a las mismas (Plan INFOCA: La restauración de las áreas incendiadas XVII). Con lo cual disponer de material que nos proporcione información objetiva acerca de los patrones espaciales de presencia de organismos, inclusive el análisis de especies aisladas, para establecer el área potencial que podría ocupar, permitiría por

ejemplo un herramienta más de apoyo para poder planificar técnicas forestales de gestión de la conservación.

Figura 2. Imagen recogida de GoogleMap donde se manifiesta el área de proyección. Sobre la imagen se han señalado las zonas donde transcurrieron los 3 incendios más importantes de las últimas décadas en las Sierras de Cazorla, Segura y las Villas. Se pueden advertir zonas de clareado aledañas a zonas forestales, lo que da un visión de las extensiones de los incendios



En relación al control de los efectos del fuego relativo a la biodiversidad forestal se pueden plantear dos tipos de usos o aplicaciones de los SDMs: (1) Preventivo. Existen 3 elementos básicos que constituyen un buen plan de prevención: selvicultura preventiva, sensibilización y vigilancia. En este último elemento, el uso de SDMs permitiría la planificación de áreas especialmente sensibles o estratégicas, con un alto grado de riqueza en especies o que puedan ser especialmente idóneas para el fortalecimiento poblacional de especies amenazadas, y que para su conservación sea necesario asignar zonas especiales de protección que requieran un mayor esfuerzo de prevención. (2) Recuperación de terrenos quemados. Para acelerar la recuperación del terreno, una de las mejores formas de hacerlo

es mediante la reforestación, sobre esta cuestión un SDMs contribuiría en la selección de especies que mejor se adaptarían a las características del terreno y particularidades del clima en la zona.

2. OBJETIVOS

En la actualidad existen varios trabajos que han modelado la distribución potencial de especies forestales en la Península Ibérica (Garzón, 2007; Alba-Sanchez, 2010), pero los objetivos de estos estudios no requieren proyecciones de alta resolución, por lo que trabajan con píxeles de 200 metros o mayores. Sin embargo, la planificación de trabajos de restauración de la vegetación suele realizarse sobre superficies de 100 x 100 metros o inferiores. El objetivo general de este trabajo es desarrollar un modelado predictivo de distribución potencial de especies de alta resolución (100 x 100 metros) que pueda servir como herramienta de apoyo en la toma de decisiones para una planificación estratégica de tareas de restauración post-incendio y conservación de biodiversidad forestal. Para ello, se han abordado dos objetivos específicos:

1. Comprobar si es posible la elaboración de SDMs de alta resolución con un nivel de validación suficientemente alto para un conjunto representativo de las especies de árboles del P. N. de las Sierras de Cazorla, Segura y Las Villas.
2. Identificar las zonas de mayor riqueza forestal potencial para planificar áreas de especial vigilancia contra incendios.

3. MATERIAL Y METODOS

3.1 Área de trabajo

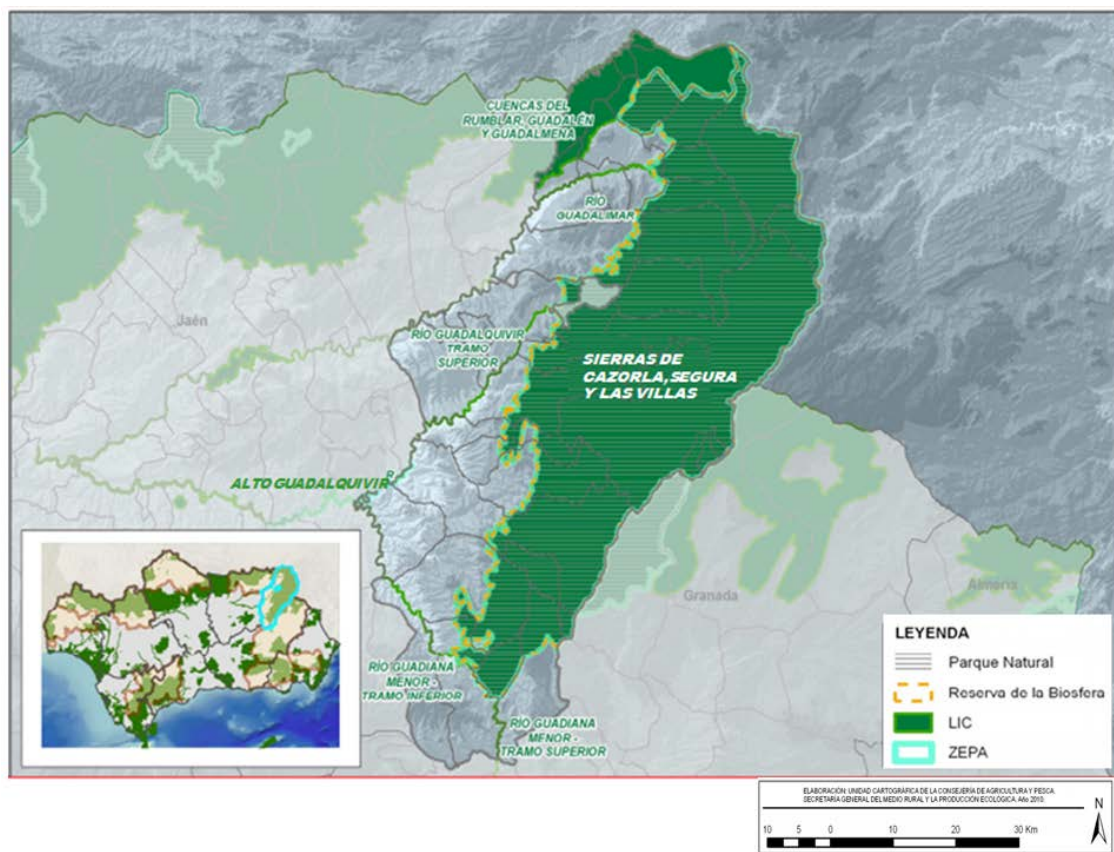
Para la elaboración de un SDM es fundamental la selección de una zona de estudio, que sirva como fuente de información para el ajuste del modelo, que presenten unas características ambientales y biogeográficas que incluyan aquellas correspondientes a las de la zona donde se va a realizar la proyección. En el presente trabajo la zona de estudio seleccionada corresponde a las provincias andaluzas en las que se circunscribe la mayor parte del Sistema Bético (Subbético y Penibético): Cádiz, Málaga, Jaén, Granada, Almería y Córdoba. Este conjunto de sistemas montañosos se extiende desde el Golfo de Cádiz hasta Alicante e incluso Baleares, sin embargo las capas de información ambiental correspondientes a territorios externos a la Comunidad andaluza no poseían la misma

resolución ni el mismo sistema de coordenadas de referencia (Datum), por lo que su empleo habría requerido la aplicación de algún método de “Downscaling” y recalcu de la proyección, todo lo cual puede tener efectos perjudiciales a la hora de ajustar y proyectar SDMs.

La zona de proyección escogida abarca gran parte del Parque Natural de las Sierras de Cazorla, Segura y Las Villas (Jaén). Concretamente el territorio enmarcado por las coordenadas: 495802 O; 538568 E; 4175872 S; 4236396 N. El sistema de referencia de coordenadas empleado fue UTM, referido al uso 30 y elipsoide GRS80, utilizando el datum WGS84. La resolución del trabajo es de 100 x 100 metros, con un total de 258.949 hectáreas analizadas. Este espacio protegido es el mayor de España y el segundo en Europa, declarada como Reserva de la Biosfera por la UNESCO desde 1983, como Parque Natural desde 1986 así como también Zona de Especial Protección para las Aves (ZEPA) desde 1987. Su complejidad orográfica, formada por un relieve quebrado y abrupto compuesto por materiales calizos, dolomíticos y margosos, así como la heterogeneidad de las precipitaciones, que da lugar a la existencia de zonas de elevada pluviometría (da origen a dos de los grandes ríos del sur de España; Guadalquivir y Segura) y zonas de clima semiárido, han posibilitado que estas sierras exhiban una de las mayores diversidades vegetales de Europa.

Es por ello que allí se dejan entrever sus más de 1800 especies catalogadas, más del 16% de la flora europea y del 22% de la ibérica y con 24 endemismos del territorio (Fernández de Trespalacios, 2007). La principal amenaza del parque se debe a la frecuencia de incendios forestales, por lo que es importante disponer de herramientas que permitan llevar a cabo una posible gestión de reforestación acorde a las particularidades del entorno afectado.

Figura 3. Espacios Naturales protegidos en las Sierras de Cazorla, Segura y las Villas. La superficie total de la Red de espacios protegidos son de unas 214.300 ha. Como se observa en el mapa la mayoría se hallan continuos, concentrándose principalmente en la zona oriental bajo figuras de LIC (Lugares de Interés Comunitario), ZEPA y Reserva de la Biosfera. La Reserva de la Biosfera, declarada el 21 de abril de 1983, está delimitada por líneas amarillas señala una zona que ocupa una superficie de 190.000 ha. Fuente: Plan de zona rural a revitalizar comunidad autónoma de Andalucía. Junta de Andalucía.



3.2 Aspectos básicos sobre SDM

La generación de SDMs requiere 2 tipos básicos de datos:

- Datos de localización de la especie (Variable dependiente): habitualmente se utilizan registros de presencia, que se suelen obtener de atlas temáticos y repositorios on-line como GBIF (<https://www.gbif.org/>), o registros de presencia-ausencia.
- Variables predictoras (Variables independientes): datos ambientales que describen el marco geográfico en el que se encuentra la especie. Teóricamente, se deben seleccionar aquellas variables ambientales que se consideren las causantes directas de la distribución de la especie. Sin embargo, rara vez se dispone de conocimiento científico suficiente como para realizar esta selección sobre criterios objetivos. Alternativamente, es habitual que se utilice el conjunto de variables ambientales disponibles, como las derivadas de modelos digitales de elevaciones o las obtenidas mediante interpolación espacial a partir de datos de estaciones climáticas (Benito, 2009). Los datos ambientales más empleados se encuentran disponibles en repositorios de información geográfica online, cuya información suele estar almacenada en formato raster.

Otra cuestión crítica es el método estadístico o algoritmo utilizado. Según Mateo et al. (2011), las diferentes técnicas pueden clasificarse en los siguientes grupos:

Técnicas descriptivas. Sólo precisan datos de presencia. Entre las más frecuentemente empleadas se encuentran:

- Modelos de envoltorio ambiental: **Bioclim** (Busby, 1986), **Anuclim** (Houlder et al., 1999), **Biomap** (Nix and Switzer, 1991) o **Hábitat** (Walker & Cocks, 1991);
- Modelos de distancias: **DOMAIN**, que utiliza la distancia de Gower (Walker & Cocks, 1991; Carpenter et al., 1993) y **MADIFA**, que se basan en la distancia de Mahalanobis (Hirzel et al., 2002; Calenge et al. 2008).

Técnicas discriminantes. Utilizan registros de presencia y ausencias.

- Técnicas de clasificación: **CART** –“*classification and regression trees*”– (Breiman et al., 1984), Random Forest (Breiman, 2001).
- Análisis de regresión: RLM, **GLMs** –“*generalized linear models*”– considerado como el método pionero en regresión logística (McCullagh & Nelder, 1989) y **GAM** –“*generalized additive models*”– (Hastie & Tibshirani, 1986).
- Máxima entropía: **MaxEnt**, genera sus propios casos de pseudo-ausencias (denominadas eufemísticamente “*background*”) (Phillips et al., 2006).

Técnicas mixtas: **GARP** -algoritmos genéticos- (Stockwell & Peters, 1999), **SVM** –“*support vector machines*”- (Vapnik, 1995).

Técnicas multi-respuesta: **MARS-Multiresponde** (Friedman, 1991).

Sea cual sea el método elegido, los modelos deben pasar por un proceso de validación para estimar con qué frecuencia las predicciones del modelo aciertan sobre la presencia/ausencia de la especie en puntos para los cuales conocemos si la especie está o no en realidad. La aplicación del modelo tendrá poco valor si no hemos evaluado la precisión de sus predicciones. Lo ideal para este paso es servirse de algún método estadístico. Para probar el rendimiento predictivo, es necesario tener datos con los que comparar las predicciones del modelo. Podemos referirnos a estos como datos de prueba (a veces llamados datos de evaluación) para distinguirlos de los datos de calibración (a veces llamados datos de entrenamiento) que se utilizan para construir el modelo. El rendimiento se

puede medir por una serie de criterios (Morrison et al., 1998), los cuales pueden aplicarse en diferentes etapas del modelo.

Por lo general todos los modelos tienen errores de predicción porque los modelos son simplificaciones de la realidad. Hay un acceso electrónico cada vez mayor a un considerable conjunto de registros de presencias u ocurrencias en museos y herbarios, pero existe una escasa orientación útil sobre la mejor manera de utilizar esta información en el contexto de como modelar distribuciones. Para satisfacer esta necesidad comparamos 6 modelos de distribución de especies: BIOCLIM, DOMAIN, MAHAL (distancia de Mahalanobis), Random Forest, GLM (General Linear Models) y el SVM (Support Vector Machine), con 23 especies arbóreas nativas de la Península Ibérica.

3.3 Variables ambientales

El siguiente paso es la selección de las variables predictoras antes de la aplicación de las técnicas de modelación. Para cualquier estudio en particular, todas las capas deben de tener la misma extensión espacial, resolución, proyección y a poder ser, la misma fuente. Se escogieron 17 variables ecológico-ambientales, que vienen recogidas en la siguiente tabla:

Tipología	Variable	Atributo
Climáticas	Evapotranspiración	Etp
	Precipitación	Pp
	Temperatura mínima anual	Tmin
	Temperatura media anual	Tmed
	Temperatura máxima anual	Tmax
	Índice de aridez	GAI
Topográficas	Altitud	Altitud
	Escabrosidad	Rugosidad
	Pendiente	Slope
	Coseno	CosenoAspect
	Seno	SenoAspect
Edáficas	Contenido hídrico del suelo	AWC
	Densidad del suelo	BulkDensity
	Fragmentos gruesos	CoarseFragments
	Arcilla	Clay
	Arena	Sand
	Limo	Silt

Tabla 1: Listado de variables predictoras escogidas para la modelación de mapas predictivos de posibles áreas potenciales para la presencia de nuestras especies. Contiguamente a la lista de

variables los atributos correspondientes con los que se le ha indicado al programa, así como su tipología.

Las capas de estas variables climáticas fueron extraídas de REDIAM (Red de Información Ambiental de Andalucía) con una resolución de 100 m. Las capas de las variables topográficas se obtuvieron del “modelo digital de elevaciones” disponible en el Instituto Geográfico Nacional (IGN: <http://www.ign.es/web/ign/portal>), adscrito al Ministerio de Fomento del Gobierno de España. Inicialmente se seleccionaron capas con una resolución de 25 m, pero estas se transformaron mediante el empleo del método de *up-scaling* a una resolución de 100 m, permitiendo una mejora en la relación entre la resolución y la extensión de área de estudio, para que posteriormente éstas se acoplen correctamente con las capas de las variables climáticas.

Por último las capas de las variables edáficas se adquirieron del *European Soil Data Centre* (ESDAC: <https://esdac.jrc.ec.europa.eu>), cuyos datasets se encuentran incluidos en el trabajo *Topsoil physical properties for Europe* (based on LUCAS topsoil data). Dichas capas tenían una resolución de unos 500 m, por lo que en este caso se realizó un *downscaling* para modificar las capas a una resolución de 100 m, la misma que el conjunto total de capas con las que se trabajó posteriormente en los modelos.

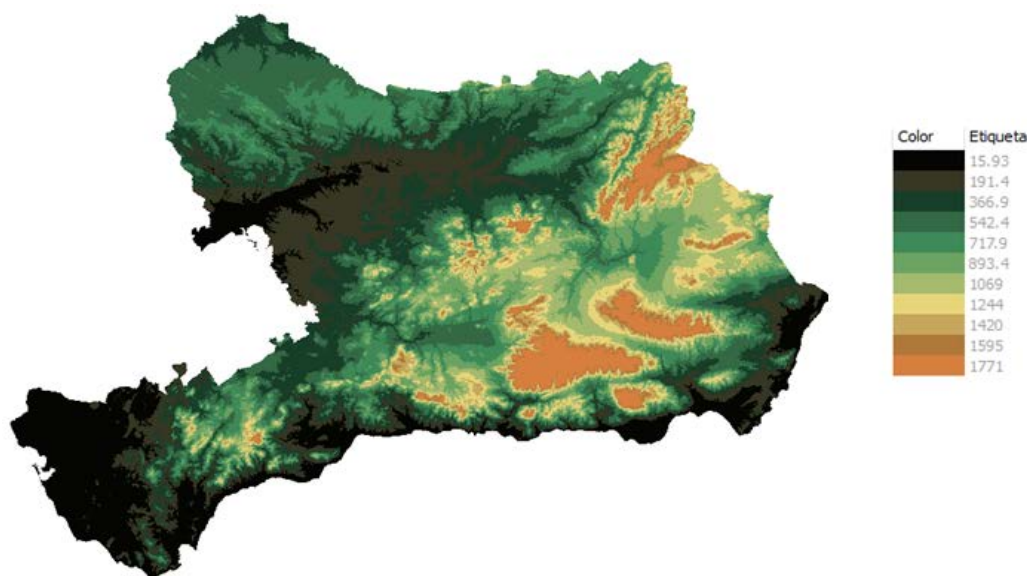


Figura 3.A. Mapa de altitud.

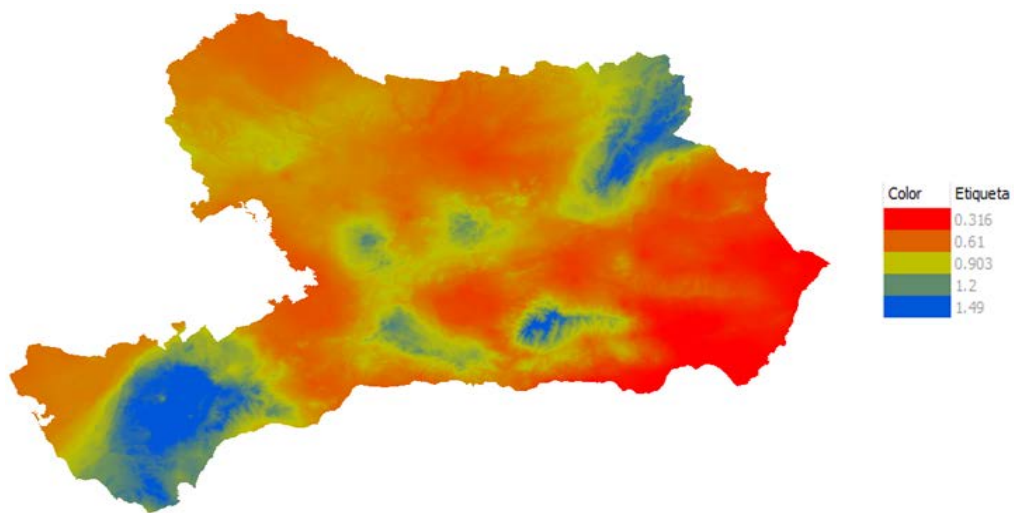


Figura 3.B. Mapa de Índice de aridez (GAI).



Figura 3.C. Mapa de densidad aparente del suelo

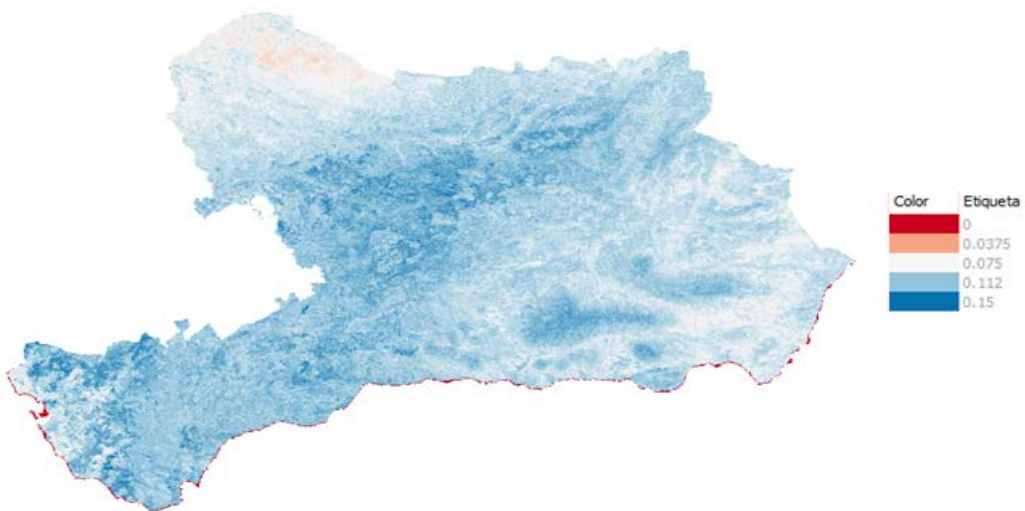


Figura 3.D. Mapa del contenido hídrico del suelo



Figura 3.E. Mapa de contenido de arcilla en suelos

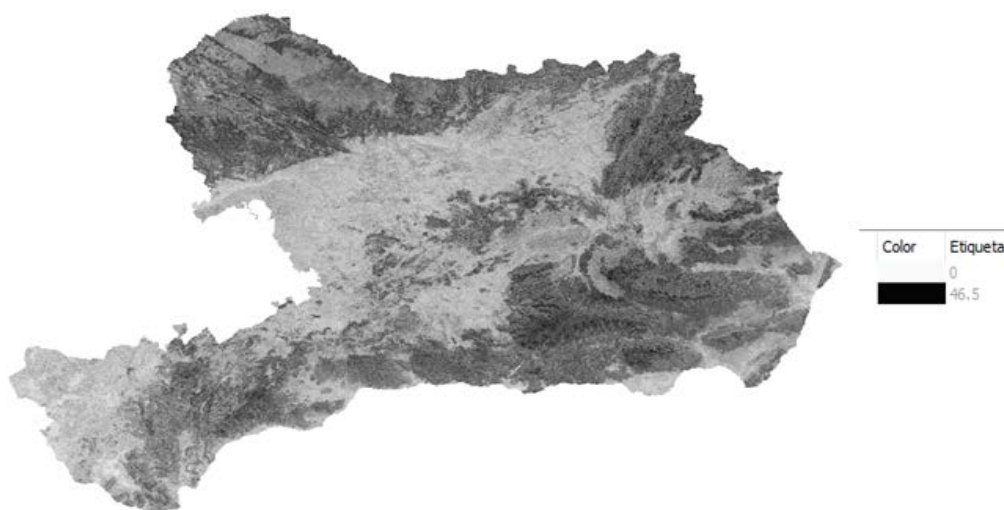


Figura 3.F. Mapa de fragmentos gruesos



Figura 3.G. Mapa del coseno de la orientación de la ladera



Figura 3.H. Mapa de evapotranspiración

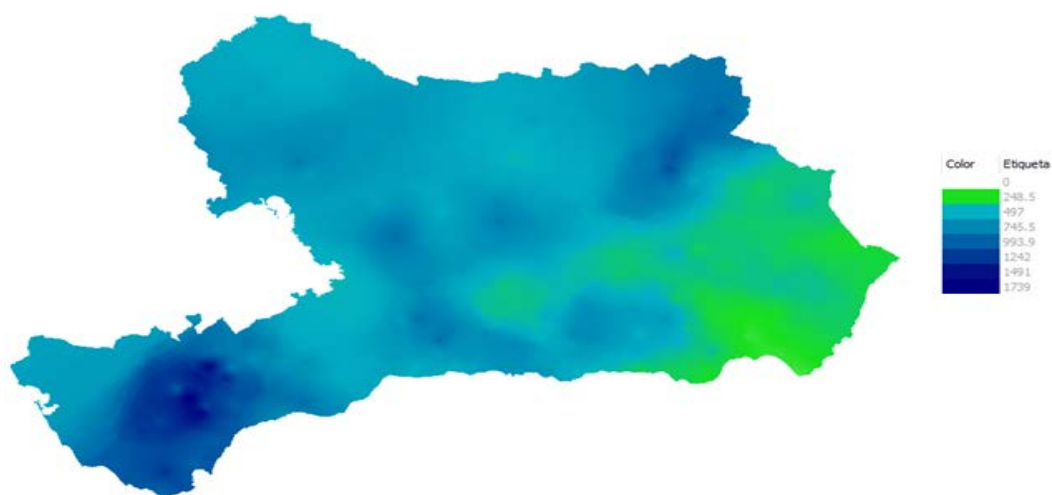


Figura 3.I. Mapa de las precipitaciones.



Figura 3.J. Mapa de la rugosidad del terreno



Figura 3.K. Mapa de arena en suelos



Figura 3.L. Mapa del seno de la orientación de la ladera



Figura 3.M. Mapa de limo en suelo

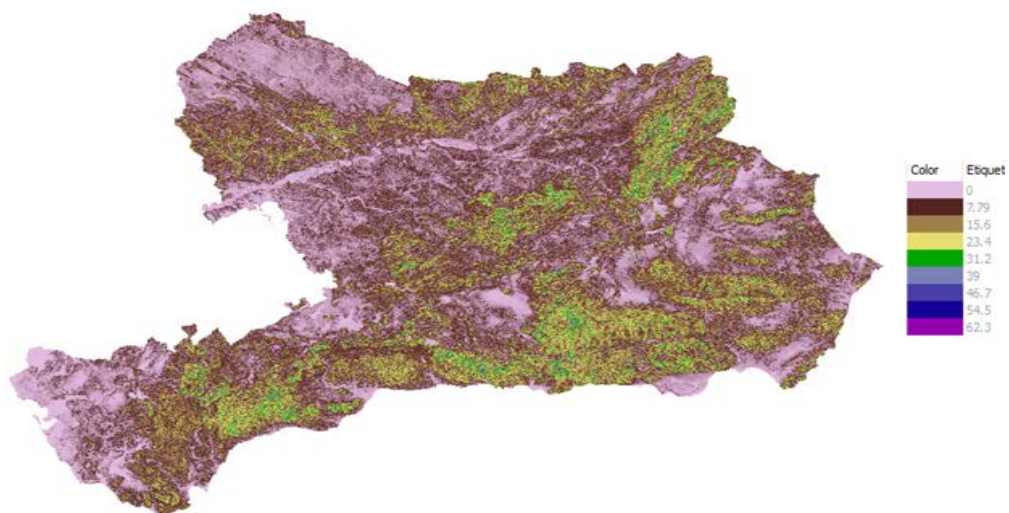


Figura 3. N. Mapa de la pendiente del suelo

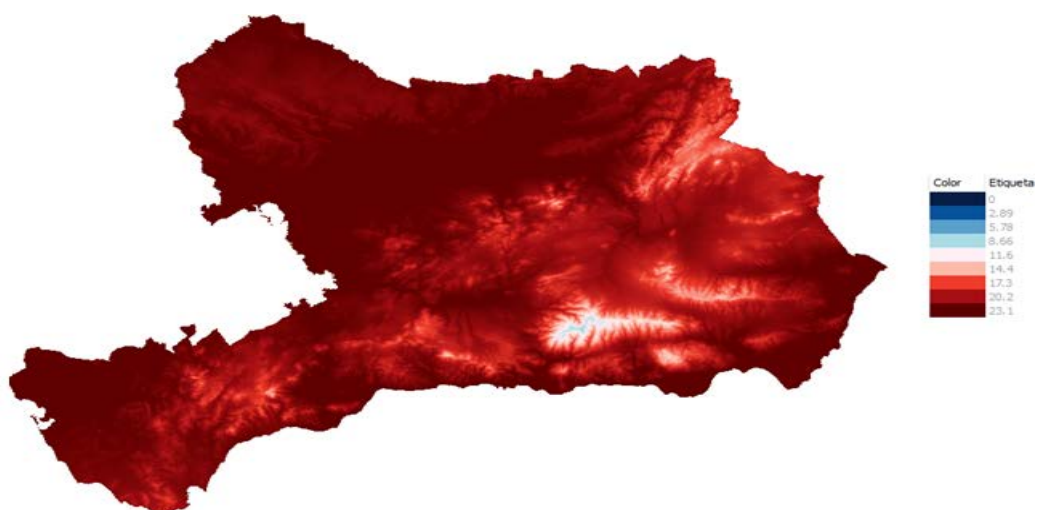


Figura 3.O. Mapa de Temperaturas máximas

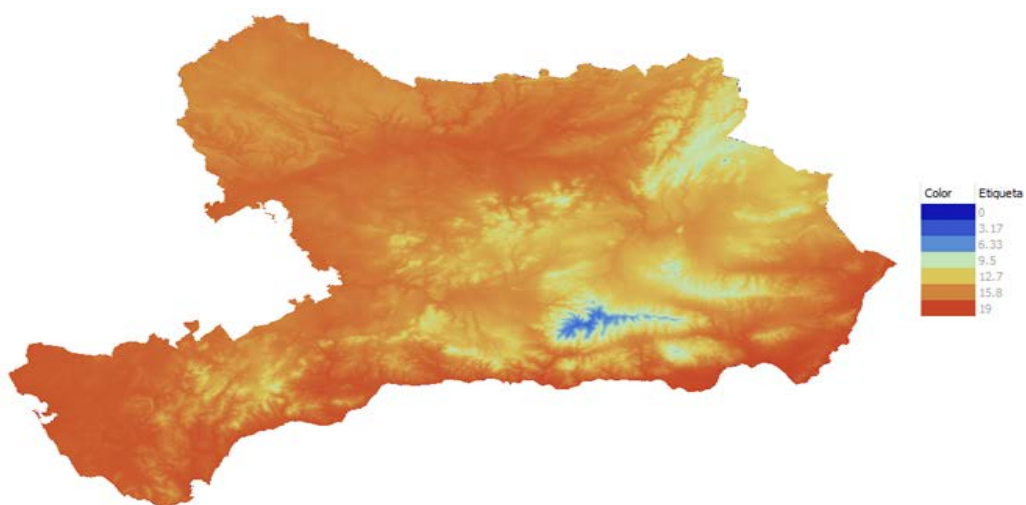


Figura 3.P. Mapa de Temperaturas medias

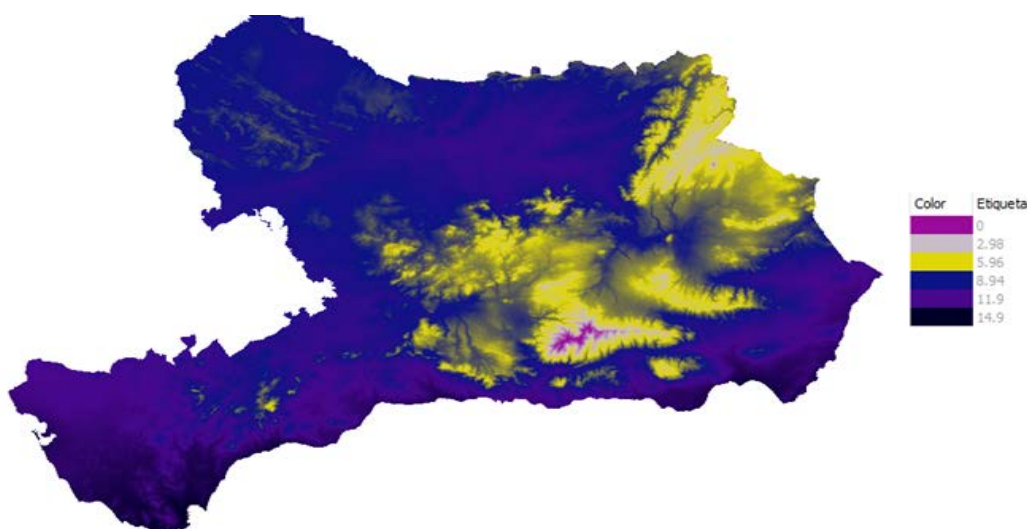


Figura 3.q. Mapa de Temperaturas mínimas

3.4 Procedencia de los datos de presencia/ausencia

El ajuste de los modelos seleccionados para este trabajo se realizó con datos del Tercer ciclo del Inventario Forestal Nacional (IFN3), realizado entre los años 1997 – 2007. En él se recogen las especies arbóreas y arbustivas que habitan en los distintos montes forestales a nivel provincial. Este trabajo queda disponible en forma de bases de datos descargables a través de los servicios del Ministerio de Agricultura, Pesca y Alimentación (Link:https://www.mapama.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/ifn3_base_datos_1_25.aspx).

De este banco de datos de acceso público se obtuvieron los ficheros con las bases de datos SIG y de CAMPO, realizados a una escala 1:50.000, de las provincias elegidas para este estudio. Esta base de datos arroja un total de 8693 puntos de muestreo con los que poder ajustar nuestros modelos. Cada uno de estos puntos corresponde a una parcela fija circular de 25 metros de radio, cuya localización quedó georreferenciada y, entre otras muchas anotaciones, se tomaron datos de las especies forestales presentes (véase ejemplo en la figura 4). El listado de especies que hemos empleado en este trabajo se muestra en la tabla 2. La "limpieza" de la información obtenida de bases de datos, como el Inventario Forestal Nacional (IFN3) en nuestro caso, GBIF o cualquier otra fuente, es particularmente un paso fundamental para obtener un buen modelo de distribución de una especie. Hay que revisar estos datos para evitar o corregir errores comunes como por ejemplo:

- Errores de omisión; celdas vacías (*nulo*), celdas con datos NA (*No Aplica*), palabras incomprensibles por falta de letras.
- Errores de redundancia: mismo dato capturado más de una vez.
- Errores de convención: signos de exclamación, interrogación, comillas, comillas simples, etcétera.

Se ha encontrado el caso en los que una especie presentaba más de un dato de ocurrencia en el mar y ha sido debido a un error en la georreferenciación. Luego es importante verificar también las coordenadas por medios visuales o de otro tipo, como por ejemplo su proyección en el mapa.

Especie	Valor de incidencia
<i>Acer monspessulanum</i> (NT)	54
<i>Acer opalus</i> (NT)	157
<i>Amelanchier ovalis</i> (NT)	242
<i>Arbutus unedo</i>	926
<i>Celtis australis</i> (NT)	35
<i>Crataegus laciniata</i> (VU)	62
<i>Crataegus monogyna</i>	1914
<i>Ilex aquifolium</i> (VU)	19
<i>Juniperus communis</i> (NT)	263
<i>Juniperus oxycedrus</i>	2509
<i>Juniperus phoenicea</i> (VU)	610
<i>Olea europea</i>	2099
<i>Phillyrea latifolia</i>	1005
<i>Pinus halepensis</i>	2171
<i>Pinus nigra</i>	1114
<i>Pistacia terebinthus</i>	859
<i>Prunus spinosa</i>	170
<i>Quercus faginea</i>	1135
<i>Quercus ilex</i>	5219

<i>Quercus pyrenaica</i>	(NT)	43
<i>Rhamnus alaternus</i>		884
<i>Sorbus aria</i>	(NT)	44
<i>Sorbus torminalis</i>	(EN)	29

Tabla 2: Recoge el listado de especies seleccionadas para el estudio. El valor de incidencia hace referencia el número de parcelas es los que se ha detectado la presencia de la especie. EN= Especies en peligro; VU=Especies vulnerables; NT= Especies casi amenazadas

Hay que resaltar que algunas de estas especies escogidas para el presente trabajo están incluidas en el catálogo de especies amenazadas o casi amenazadas, bajo la aplicación regional de criterios y categorías UICN (La Unión Internacional para la Conservación de la Naturaleza) (2001), de la Lista Roja de la Flora Vascular de Andalucía (Cabezudo *et al.*, 2005). *Sorbus torminalis* se encuentra dentro de la categoría de especies en peligro (EN), la cual estaría en una situación de riesgo de extinción muy alto. Sus principales amenazas son, según este libro el urbanismo, la silvicultura y los incendios. Otras 10 especies de las 23 seleccionadas, se encuentran en unas circunstancias vulnerables (VU) o casi amenazadas (NT) (Ver tabla 2).

Por tanto los mapas de modelado de idoneidad del hábitat para nuestras 23 especies elegidas, pueden alzarse como una herramienta más que pueda estar incluida dentro del Plan Forestal Andaluz.

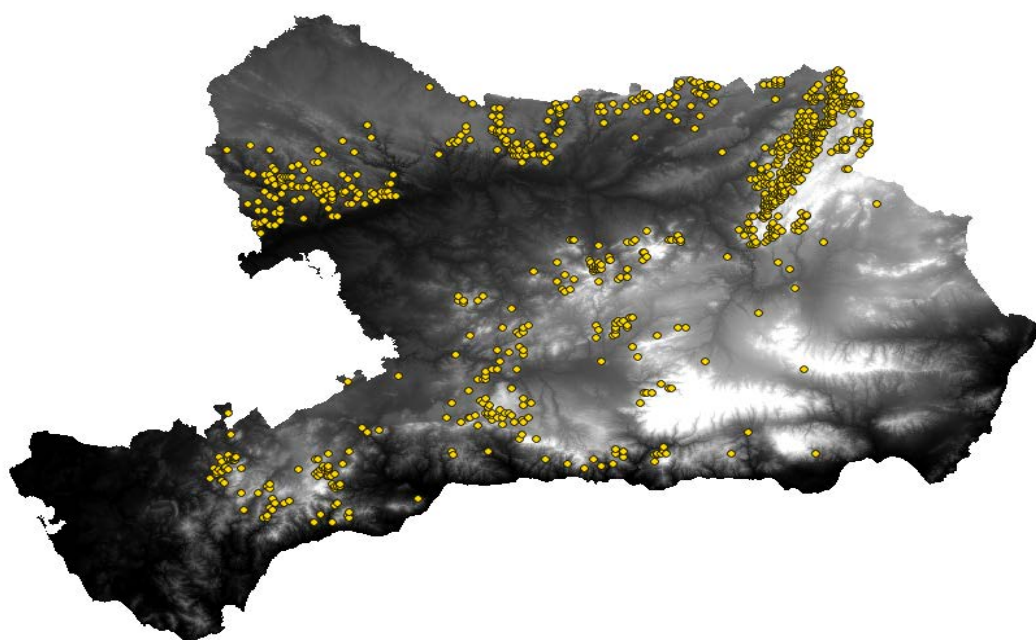


Figura 4. Mapa de altitud al cual se le incorpora una capa de texto con las coordenadas de las parcelas donde consta la presencia de *Pistacia terebinthus*, también conocida como “cornicabra”

3.5 Elaboración de los SDMs y proyección de las áreas potenciales

Los SDMs indican la idoneidad del hábitat para el desarrollo de poblaciones de una especie en concreto o de una comunidad (Ferrier y Guisan, 2006). En varios estudios se ha podido constatar que los diferentes algoritmos presentan distinta precisión a la hora de estimar la distribución del hábitat idóneo para una especie (Manel et al., 1999; Thuiller, 2003; Muñoz y Felicísimo, 2004; Elith et al., 2006). Por esta razón, en este trabajo, se emplean múltiples modelos calculados con distintos algoritmos, mediante los cuales posteriormente, se seleccionarán aquellos modelos más óptimos para cada especie.

Para la preparación de nuestros modelos resultó clave tener en cuenta una serie de factores que pueden afectar al resultado final. La calidad de los datos de localización de la especie, donde existen dos tipos de registros: de presencia y de presencia–ausencia. Los de presencia son incuestionables (sobre todo si son especies sésiles o se comprende que no se cometió ninguna confusión en su identificación), pero con los datos de ausencia no ocurre lo mismo (procesos como la fragmentación del hábitat, dinámica de poblaciones, presencia de especies invasoras, pueden afectar a las ausencias). Es complicado encontrar bases de datos donde tengan registros de ausencia-presencia, por lo que mediante una expresión binaria dimos valor 1 a aquellos puntos georreferenciados, donde nuestra fuente asegura la presencia de la especie, y valor 0 donde no existen datos de existencia. Una vez creado nuestro *dataset* seleccionamos primeramente algunas técnicas descriptivas que usan únicamente datos de presencia (fueron las primeras utilizadas en modelización):

- **BIOCLIM** (Busby, 1986): Este algoritmo genera para la especie un rango ecológico de n -dimensiones, siendo n el número de variables predictoras, es decir, genera un hipercubo delimitado por el mínimo y máximo de cada variable ambiental en los lugares ocupados por la especie. Predicen la idoneidad del hábitat de forma binaria (adecuado/no adecuado) en base a la ubicación de las condiciones ambientales dentro del hipercubo.
- **DOMAIN** (Carpenter et al., 1993): Este modelo consiste en un conjunto de hipervolúmenes contruidos en torno a cada punto ocupado por la especie. Predicen la idoneidad del hábitat como la distancia desde un punto al punto de presencia más

próximo, relativizada a la amplitud del rango de cada variable ambiental (distancia Gower).

- **Distancia de Mahalanobis** (Hirzel et al., 2002): Modelo de envoltorio ambiental que utiliza la distancia de Mahalanobis. Determina la similitud entre dos variables aleatorias multidimensionales teniendo en cuenta la covariación entre ellas.

El resto de modelos que se han empleado en este trabajo se basan en el uso de datos de presencia/ausencia. Se tratan de técnicas discriminantes que trabajan con funciones o algoritmos estadísticos que también han demostrado ser una herramienta útil dentro del campo de la biogeografía de la conservación. Concretamente los modelos empleados son:

- **GLM** (General Linear Models; Guisan et al., 2002): Se ajustó un modelo de regresión logística (presencia =1; ausencia = 0) mediante el uso de una distribución Binaria. Se ajustaron dos modelos de GLM distintos, uno con la función de enlace Logit y el otro con Probit. El modelo que alcance un menor valor de AIC (*Akaike Information Criteria*) será el que mejor explique la información contenida en la variable dependiente, es decir, será el que consiga una menor disparidad entre lo observado y lo predicho.
- **Random Forest** (Breiman et al. 1984): Es una extensión de los métodos de Árboles de Regresión y Clasificación (CART) en la que se producen multitud de árboles de regresión que luego son sintetizados en uno de consenso.
- **SVM** (Support Vector Machines). En español máquinas de vectores de soporte, el uso de este algoritmo se inicia con los trabajos de Vapnik, pero fue utilizados por primera vez en el modelado de distribución de especies por Guo et al. (2005). Trabaja bien en un sistema binario (presencia/ausencia), se encuentra dentro del grupo de los métodos de aprendizaje automático y requieren una cantidad mínima de ajuste del modelo (Joachims, 1998; Brown et al., 2000).

Como se ha comentado anteriormente, diversos estudios coinciden en que ninguno de estos métodos se puede considerar el más adecuado en todas las circunstancias. Ante esta situación algunos autores (Araujo y New, 2007) recomiendan la elaboración de modelos de consenso. Una vez ajustado todos los modelos y obtenido un valor AUC de validación (véase apartado 3.6) se realiza la media ponderada de todos los modelos. En este trabajo se ha ponderado por el valor de AUC al cuadrado para reforzar la importancia de una buena validación. Es decir, si un modelo tiene un valor de AUC de 0.9 esté tendrá más peso (0.81)

en el modelo de consenso final que un modelo que tenga un valor de AUC de 0.7 (con un peso de 0.49). Este modelo de consenso, al que hemos llamado “**Combinado**”, es a su vez validado para obtener un valor propio de AUC.

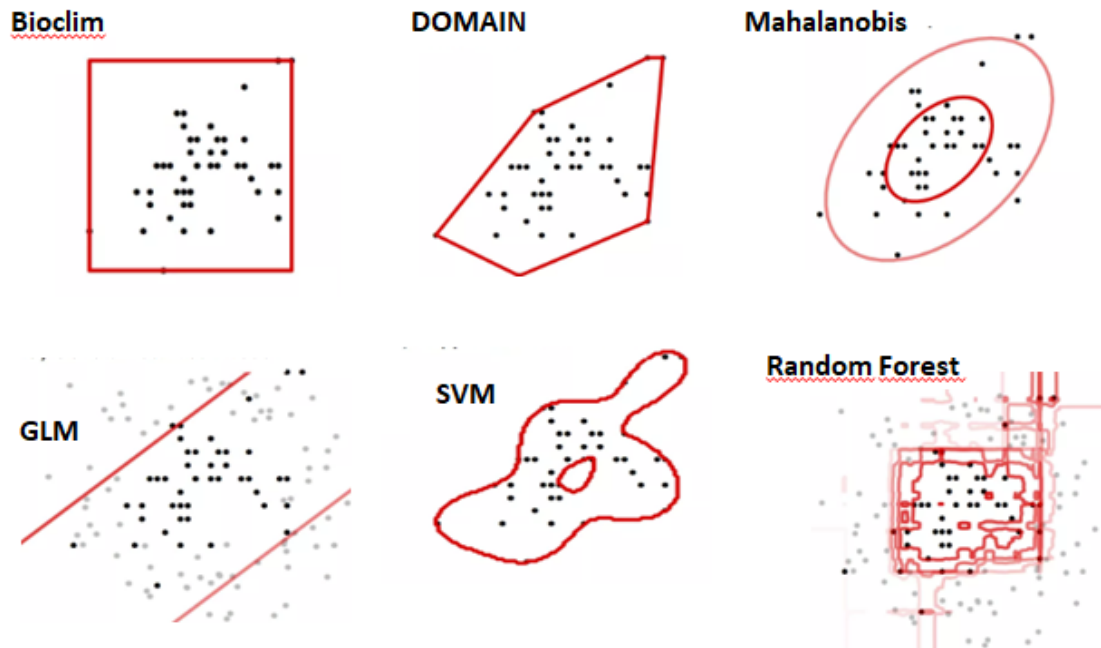


Figura 5. Comparación de los métodos de estimación de hipervolumen seleccionados. Los puntos de datos de presencia se muestran como círculos negros cerrados. Para aquellos métodos que precisan datos de ausencias los datos se muestran como círculos grises abiertos. Bioclim refleja cuadros de rango, encerrando los datos dentro de cada eje de forma independiente. DOMAIN encuentra el conjunto mínimo de restricciones de desigualdad lineales que encierran los datos. Mahalanobis forma elipses probabilísticas, tienen una distribución normal multivariada de los datos. GLM o modelos lineales generalizados tienen una superficie de regresión lineal para los datos. SVM o máquinas de vectores de soporte encuentran un clasificador de alta dimensión para los datos. Los modelos de bosques aleatorios como Random Forest usan simultáneamente árboles de regresión múltiple para votar los datos. Fuente: imagen recogida del trabajo de Blonder (2017).

3.6 Calibración del modelo

Tanto la calibración como la validación de los modelos expuestos anteriormente se realizó con el programa de RStudio Desktop 1.1.456. La figura 6 representa esquemáticamente los pasos necesarios para la elaboración de los modelos.

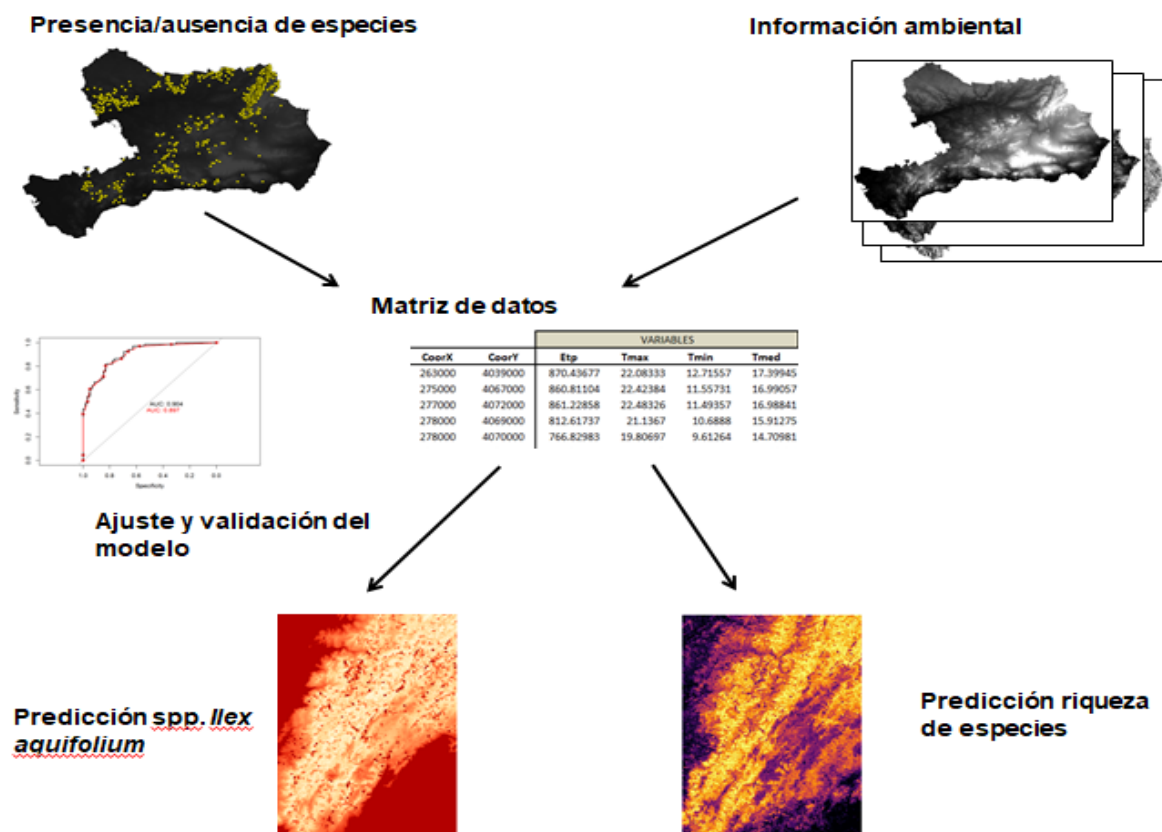


Figura 6. Esquema conceptual de la configuración de un Modelo de Distribución de Especies. Fuente: elaboración propia.

Preparación de los datos. Señalamos al programa donde se encuentra el archivo el cual recoge los datos de las coordenadas así como las presencias – ausencias de nuestras especies a modelar. Introducimos un “extent” con las coordenadas de nuestra área de proyección y cargamos las capas ambientales, cuya carpeta se redirigirá a nuestro directorio de trabajo.

Filtrado de las variables predictoras. Es probable la existencia de colinealidad, esto hace referencia a la posible situación en la que algunas de nuestras variables explicativas estén fuertemente correlacionadas, y por lo tanto, nos puede resultar difícil observar sus efectos individuales sobre la variable endógena. Para reducir los problemas de colinealidad empleamos una matriz de correlaciones entre todas las variables ambientales, a partir de la

cual se elaboró un grafo (véase ejemplo en figura 7) en el que se conectan aquellas variables cuyo valor de correlación de Pearson era $r \geq 0.7$. A la vista del grafo es fácil identificar aquellas variables cuya eliminación daría lugar a la desconexión completa del grafo; es decir, daría lugar a un conjunto de variables correlacionadas entre sí con $r < 0.7$. Si este conjunto presentaba un “reciprocal condition number” superior a 0.01, se puede asumir que no hay problemas significativos de colinealidad.

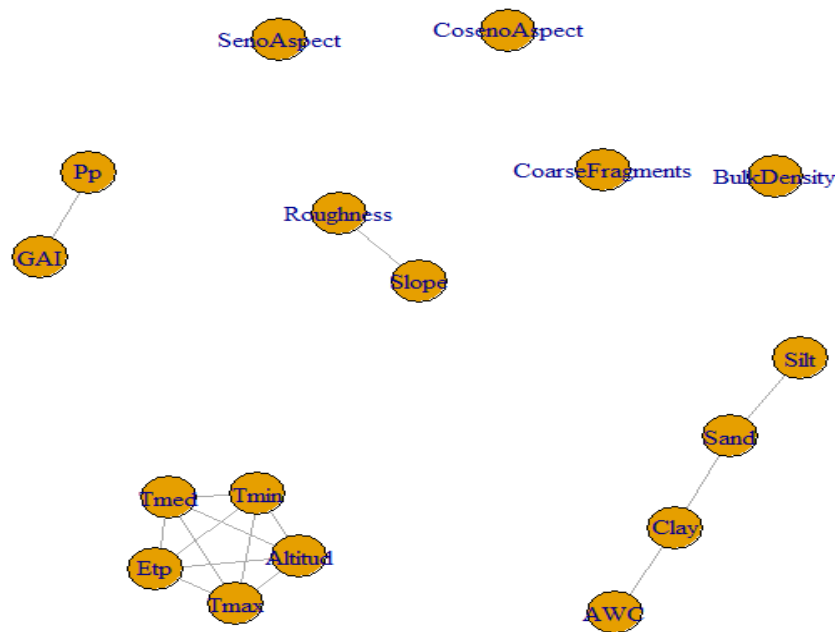


Figura 7. Grafo creado a partir de la matriz de correlación con los datos de *Amelanchier Ovalis*. En este caso se eliminaron las variables: Slope, Sand, Tmed, Tmin, Tmax, Etp, Clay y Pp. Su eliminación nos concede un grafo totalmente desconectado, descartando la existencia de fuertes correlaciones entre las variables restantes. En este ejemplo nos devolvió un valor de $R_{cond} = 0,108$, (superior a 0.1) con lo cual la ejecución del modelo puede proseguir de manera más fiable.

Creación de los datasets. Usando el mismo conjunto de datos para calibrar y evaluar los SDM, proceso al que también se le conoce como "resubstitución", se tiende a sobreestimar el rendimiento predictivo del modelo para predecir nuevas observaciones (Edwards et al., 2006). Algunos estudios tienen datos independientes disponibles para la validación (Franklin, 2002; Elith et al., 2006), sin embargo, frecuentemente no es factible recopilar nuevos datos. En esta situación existen varias estrategias para dividir datos, la más simple es una división única en la que los datos disponibles son asignados por un lado a un conjunto de datos de calibración y por otro un conjunto de prueba seleccionados aleatoriamente (Pearson et al., 2002). Otro método de partición de datos útil es realizar una validación Cruzada o k-fold Cross Validation, consiste en tomar los datos originales y crear a partir de ellos dos conjuntos separados: un primer conjunto de entrenamiento (*Training*), y

un segundo conjunto de validación (*Testing*). En este trabajo se indicó un $K\text{-fold} = 5$, de esta manera estaríamos asignando un 80% de los casos a training y 20% a testing.

Ejecución y validación de los modelos. Una vez ejecutado el modelo es indispensable su validación y verificación de su capacidad predictiva. Los modelos de predicciones binarias, como la presencia/ausencia de una especie, a menudo se evalúan utilizando la proporción de presencias reales que se predicen correctamente (Sensibilidad) y la proporción de las ausencias reales que se predicen correctamente (Especificidad) (Allouche et al., 2006). Para calcular estos estadísticos es necesario aplicar a la predicción de los modelos (que es una variable continua) un valor umbral de predicción a partir el cual se considerará que un lugar (i.e. un pixel del mapa de proyección) es adecuado para la especie. De esta forma transformamos la predicción en una variable binaria. Para determinar el valor umbral hemos empleado aquel que maximiza el estadístico TSS (True Skills Statistic), que combina los valores de sensibilidad y especificidad.

Para verificar la capacidad predictiva del modelo se suele emplear la curva de ROC (acrónimo de *Receiver Operating Characteristic*; Hanley McNeil, 1982). La curva ROC y su estadístico derivado, el AUC (figura 8), necesitan que los datos de entrada sean de presencia/ausencia, y en número mayor a las 15-20 presencias únicas, ya que con valores menores puede incrementar artificialmente la consistencia del modelo (Veloz 2009). Es una medida directa de la capacidad de discriminación del modelo, que toma valores próximos a 1 cuando existe un buen ajuste con los datos de evaluación (los casos se han clasificado correctamente); cercanos a 0.5 cuando el ajuste nos es el mejor que el obtenido por el azar y valores menores a 0.5 indican que el modelo es realmente malo, ya que clasifica erróneamente más casos que el azar.

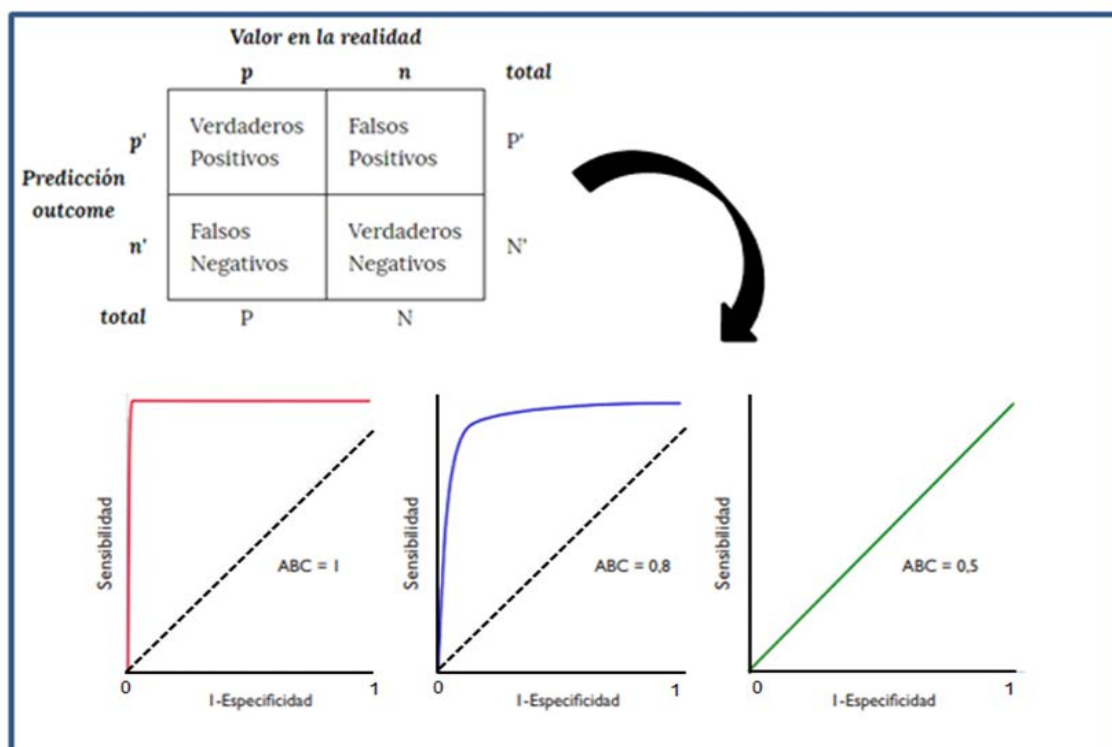


Figura 8. El cuadro superior de la figura refleja los 4 posibles resultados que se pueden formular en una Matriz de confusión. Debajo de este se encuentra una representación de diferentes Curvas ROC que podemos encontrarnos. Fuente: elaboración propia a partir de Molina Arias, M. (2007).

Cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC. La curva se obtiene dibujando los valores de sensibilidad en el eje Y, frente a los valores de 1-especificidad en el eje X para cada uno de los valores de idoneidad o probabilidad del modelo. La AUC es la integral definida por esta curva al variar los umbrales de corte (Benito, 2009) (Figura 8).

4 RESULTADOS

Los siete SDM fueron capaces de predecir la distribución potencial para nuestras especies seleccionadas, aunque los resultados de predicción fueron diferentes (Tabla 3, figura 9).

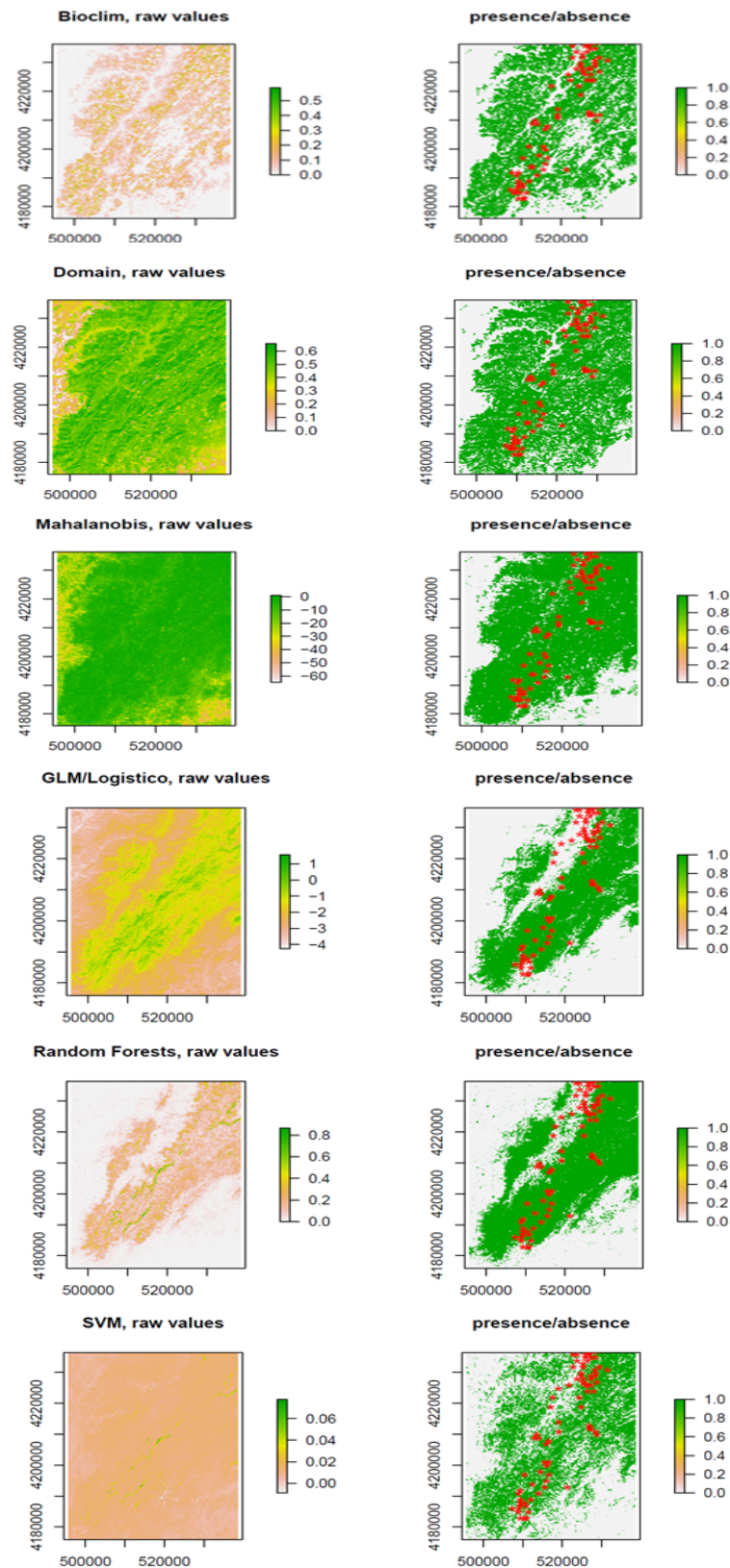


Figura 9. Representación de los 6 modelos escogidos y sus correspondientes mapas de presencia/ausencia predicha para la especie *Acer opalus*. Los puntos rojos indican lugares de presencia conocida de la especie. Se puede apreciar claramente como los resultados son diferentes con cada modelo a pesar de que en todos los casos se empleó el mismo set de datos.

Como se ha explicado previamente en el apartado 3.4, con la media ponderada de AUC² se realiza un séptimo modelo al que se le ha llamado Combinado. La tabla 3 recoge los valores de AUC que han arrojado la ejecución de los SDMs para cada especie. El valor de prevalencia señala la proporción de individuos de una especie que teóricamente podríamos encontrar ocupando un lugar dentro de la zona de proyección y que ha sido predicho por el mapa de presencia/ausencia del modelo que mayor confianza ha mostrado.

En general los valores de AUC registrados superan el 0,8 a partir del cual se considera por bueno el modelo. Solo podemos destacar como especies para las que se han conseguido modelos de menor calidad a *Rhamnus alaternus*, con un valor de AUC de 0.73 con el uso del modelo de Random Forest, y a *Celtis australis* 0.77 con Mahalanobis. Por la baja capacidad predictiva conseguida para estas dos especies, no resulta recomendable el extraer conclusiones respecto a su idoneidad en distintas zonas del área de proyección.

Tabla 4. Valor medio de AUC para cada modelo. SD es la abreviatura de desviación estándar. El rango de valores máximos y mínimos indican los resultados más bajos y más altos que ha revelado el estadístico AUC para todas las especies con la ejecución de cada uno de los modelos.

SDMs	AUC Media $\pm$ SD	Rango de valores Máx- Mín
BIOCLIM	0.76 $\pm$ 0.09	0.53 – 0.93
DOMAIN	0.75 $\pm$ 0.15	0.52 – 0.99
Mahalanobis	0.82 $\pm$ 0.11	0.64 – 0.98
GLM	0.79 $\pm$ 0.11	0.53 – 0.95
RANDOM FOREST	0.85 $\pm$ 0.09	0.52 – 0.97
SVM	0.80 $\pm$ 0.09	0.58 – 0.94
COMBINADO	0.79 $\pm$ 0.12	0.59 – 0.92

Especies	AUC							Prevalencia predicha
	BioClim	DOMAIN	Mahalanobis	Logístico	RandomForest	SVC	Combined	
<i>Acer monspessulanum</i>	0.688	0.897	0.925	0.829	0.857	0.798	0.928	0.08
<i>Acer opalus</i>	0.882	0.937	0.946	0.919	0.926	0.834	0.660	0.69
<i>Amelanchier ovalis</i>	0.901	0.829	0.928	0.900	0.894	0.802	0.641	0.73
<i>Arbutus unedo</i>	0.823	0.650	0.829	0.812	0.846	0.772	0.916	0.31
<i>Celtis australis</i>	0.738	0.757	0.776	0.530	0.527	0.582	0.647	0.37
<i>Crataegus laciniata</i>	0.735	0.977	0.984	0.877	0.872	0.885	0.761	0.38
<i>Crataegus monogyna</i>	0.708	0.596	0.691	0.816	0.840	0.805	0.800	0.2
<i>Ilex aquifolium</i>	0.622	0.976	0.959	0.721	0.748	0.628	0.929	0.43
<i>Juniperus communis</i>	0.883	0.915	0.940	0.959	0.970	0.941	0.885	0.41
<i>Juniperus oxycedrus</i>	0.768	0.655	0.755	0.688	0.861	0.798	0.655	0.66
<i>Juniperus phoenicea</i>	0.799	0.698	0.827	0.755	0.885	0.747	0.595	0.67
<i>Olea europaea</i>	0.716	0.552	0.648	0.875	0.887	0.859	0.869	0.19
<i>Phillyrea latifolia</i>	0.856	0.689	0.819	0.874	0.901	0.848	0.938	0.22
<i>Pinus halepensis</i>	0.794	0.654	0.759	0.716	0.895	0.825	0.893	0.53
<i>Pinus nigra</i>	0.930	0.895	0.938	0.958	0.957	0.928	0.778	0.57
<i>Pistacia terebinthus</i>	0.810	0.675	0.774	0.654	0.825	0.765	0.804	0.6
<i>Prunus spinosa</i>	0.821	0.838	0.835	0.866	0.870	0.795	0.682	0.5
<i>Quercus faginea</i>	0.821	0.707	0.764	0.682	0.804	0.736	0.718	0.72
<i>Quercus ilex</i>	0.716	0.600	0.666	0.701	0.846	0.807	0.858	0.48
<i>Quercus pyrenaica</i>	0.708	0.549	0.665	0.706	0.840	0.811	0.864	0.23
<i>Rhamnus alaternus</i>	0.687	0.520	0.650	0.707	0.737	0.664	0.627	0.57
<i>Sorbus aria</i>	0.532	0.788	0.853	0.891	0.864	0.854	0.949	0.18
<i>Sorbus torminalis</i>	0.667	0.996	0.980	0.820	0.898	0.896	0.841	0.48

Figura 10. SDMs utilizados con su correspondiente valor de AUC predicho. Se resaltan en rojo los valores conseguidos de AUC más altos. La prevalencia predicha expresa el porcentaje del área proyectada donde el modelo predice la presencia de la especie acorde a los datos de ocurrencia introducidos en la ejecución del dicho modelo.

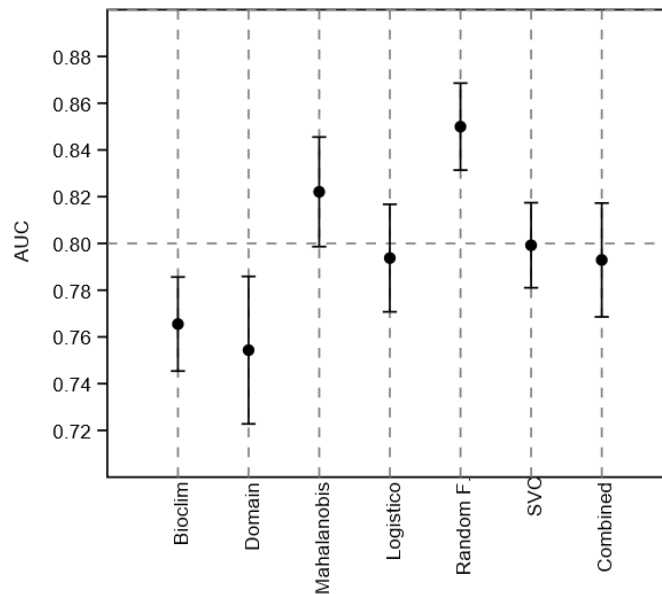


Figura 10. Diagrama de bigotes (MEDIA $\pm$ 1 S.D.) con los modelos usados frente a su valor promedio de AUC.

Una puntuación AUC alta revela que el modelo puede discriminar con precisión entre los lugares en los que la especie está presente o ausente. Como se aprecia en la tabla 4 o la figura 10, en general el modelo de Random Forest es por promedio el que mejores valores de AUC alcanza, seguido por Mahalanobis. Por el contrario DOMAIN es el modelo que peores datos de AUC acumula, con un promedio de 0.75 y por detrás de este BioClim. El análisis de algunas especies mediante el modelo de DOMAIN arrojaba valores próximos 0.5 (valor por referencia a partir del cual el modelo sería realmente tan malo como si se realizara la predicción de manera azarosa): *Rhamnus alaternus*, *Quercus pirenaica* y *Olea europea*. Por consiguiente se puede decir que DOMAIN ha sido el modelo que particularmente ha dado un peor rendimiento, un hecho que además se contempla cuando se examina la media de la desviación estándar de los valores de AUC, la cual ha sido la mayor sobre el resto de modelos (figura 10). Pero ha de destacarse que DOMAIN fue el mejor modelo para aquellas especies que presentaban un valor de incidencia bajo (*Ilex aquifolium* y *Sorbus torminalis*). Es decir, parece ser que DOMAIN tiene tendencia a funcionar bien para aquellas especies con datos de ocurrencia escasos.

Para exponer los resultados de la configuración de los mapas a partir de los modelos seleccionados, se han ilustrado en el trabajo aquellos cuyos valores de AUC se aproximaban más al valor 1 considerado como un valor del test excelente para cada especie. La figura 11 muestra la proyección de esos mapas.

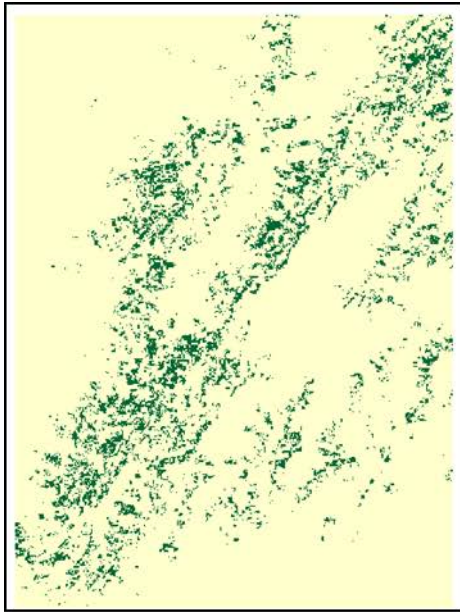


Figura 11.a Combinado:0.92
Acer monspessulanum

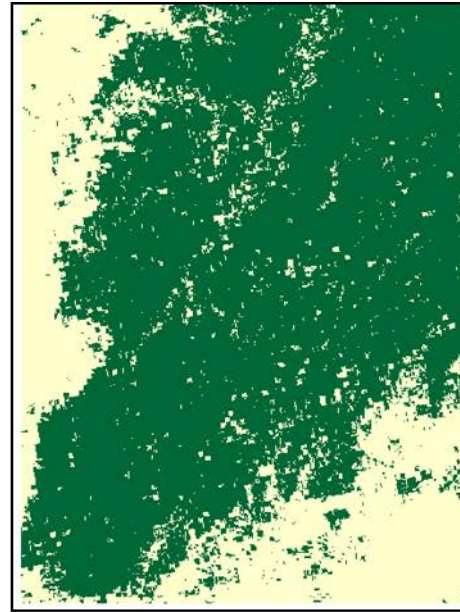


Figura 11.b D. Mahalanobis:0.94
Acer opalus

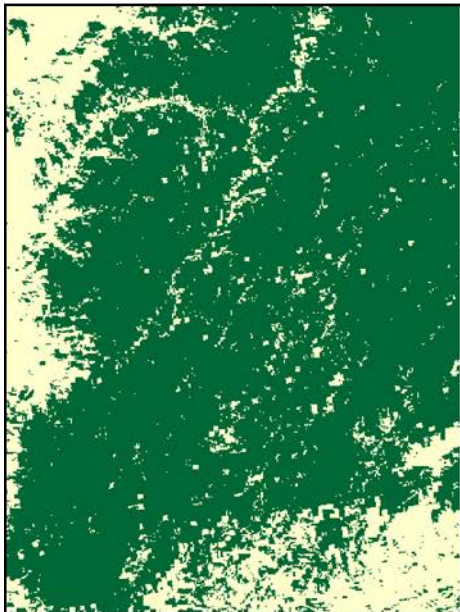


Figura 11.c D. Mahalanobis:0.93
Amelanchier ovalis

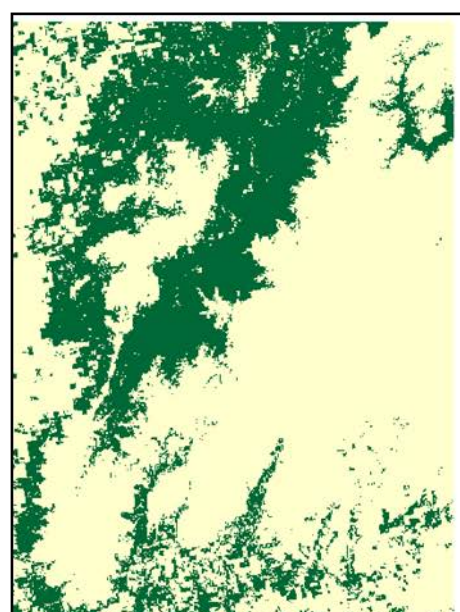


Figura 11.dD. Combinado:0.92
Arbutus unedo

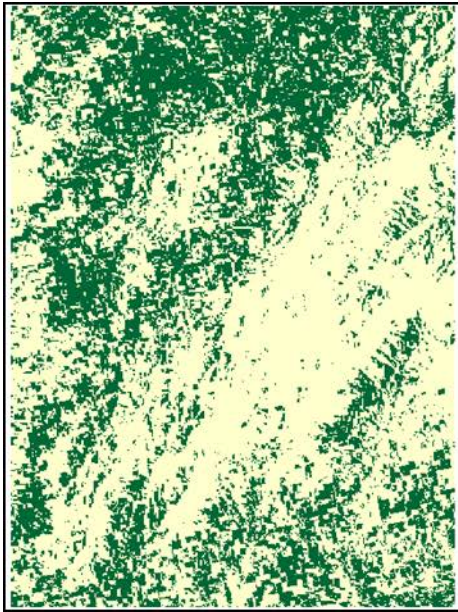


Figura 11.e. D.Mahalanobis:0.77
Celtis australis

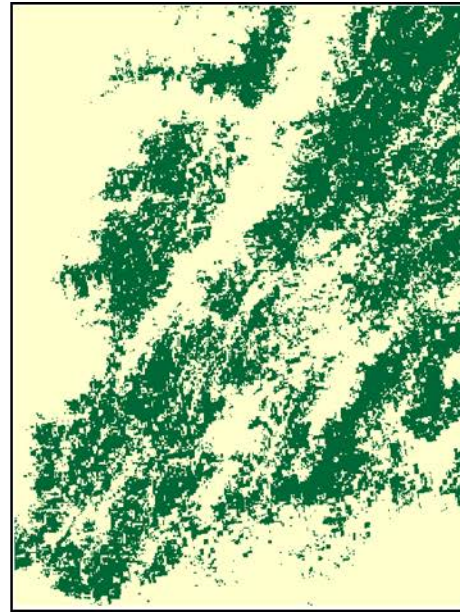


Figura 11.f. D.Mahalanobis:0.98
Crataegus lacinata

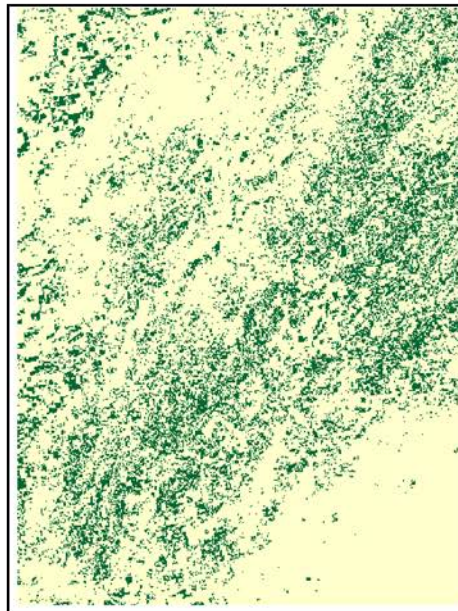


Figura 11.g R. Forests:0.84
Crataegus monogyna

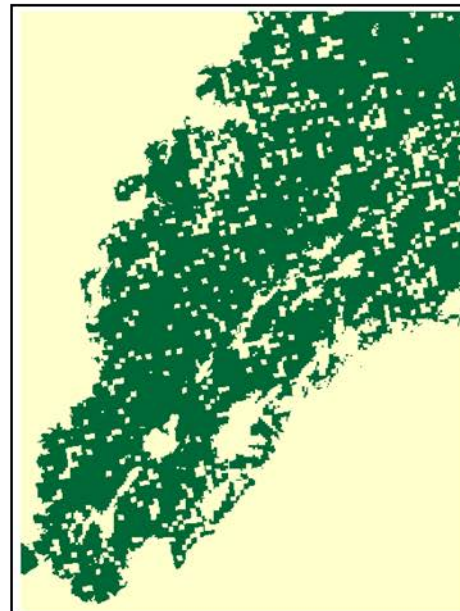


Figura 11.h DOMAIN:0.84
Ilex aquifolium

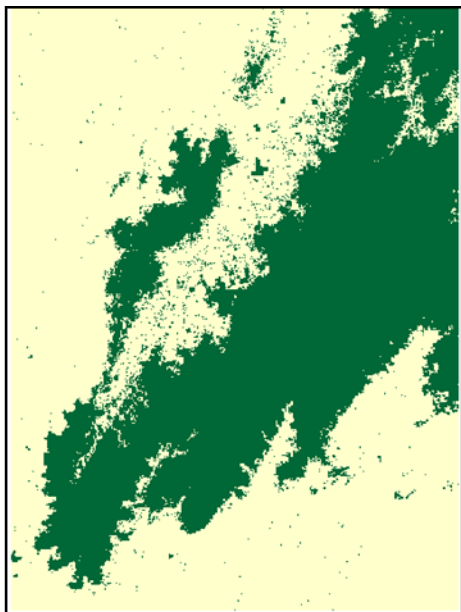


Figura 11.i. D.Mahalanobis:0.77
Juniperus communis

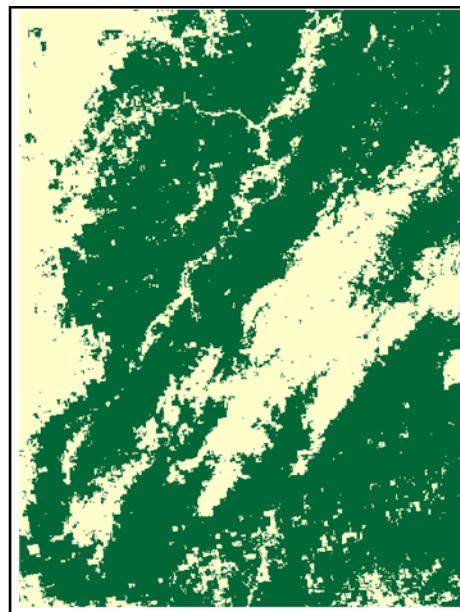


Figura 11.j. R. Forest:0.86
Juniperus oxycedrus

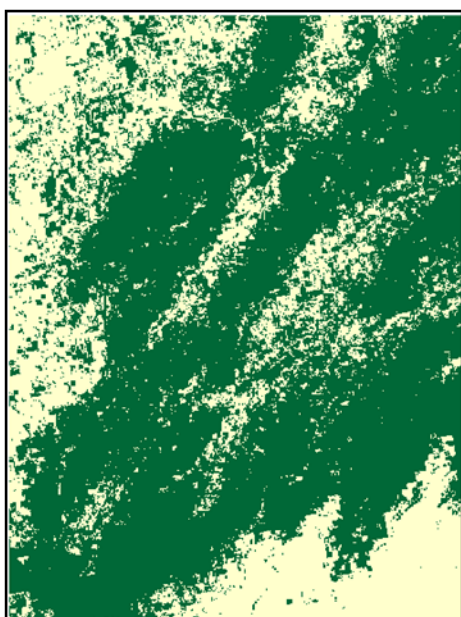


Figura 11.k. R. Forest:0.88
Juniperus phoenicea



Figura 11.L. R. Forest:0.89
Olea europaea



Figura 11.m Combined:0.94
Phillyrea latifolia

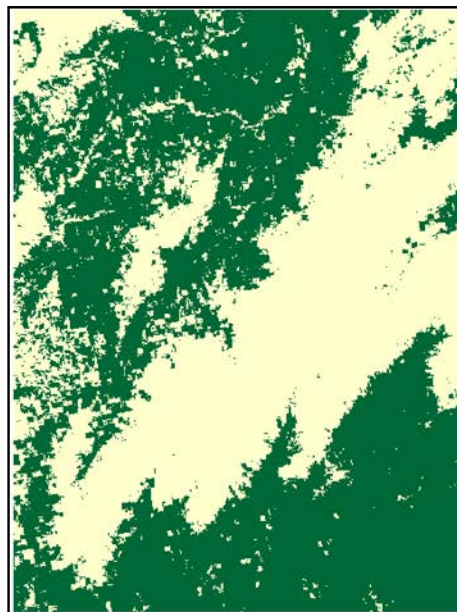


Figura 11.n. R. Forest:0.89
Pinus halepensis

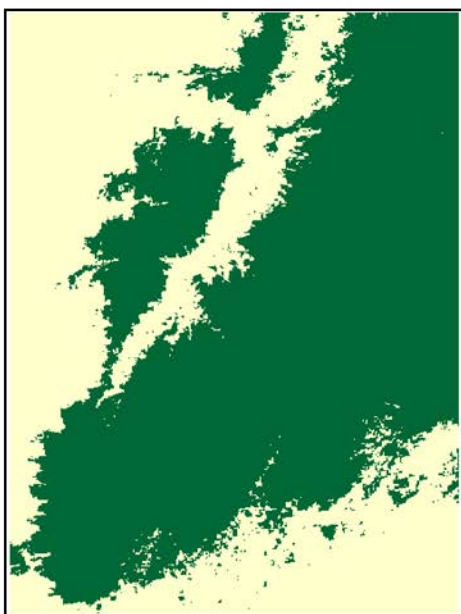


Figura 11.o. GLM:0.96
Pinus nigra

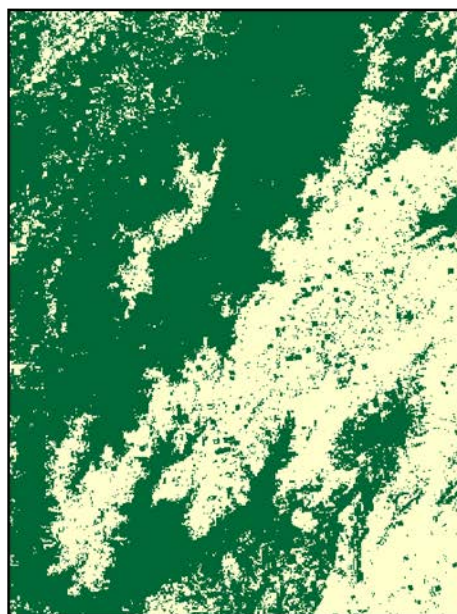


Figura 11.p. R. Forest:0.82
Pistacia terebinthus

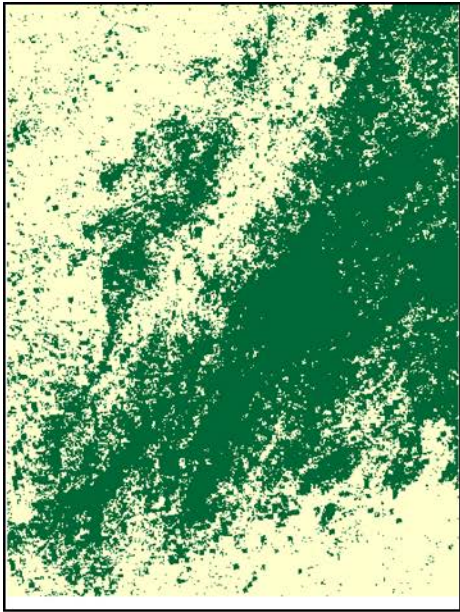


Figura 11.q. R. Forest:0.87
Prunus spinosa

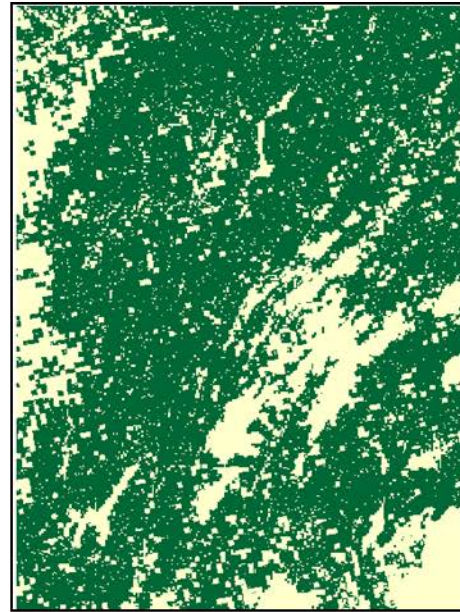


Figura 11.r. R. Forest:0.82
Quercus faginea

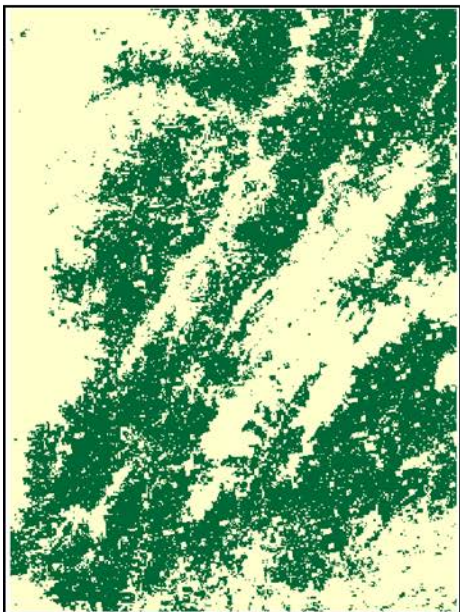


Figura 11.s. Combined::0.86
Quercus ilex

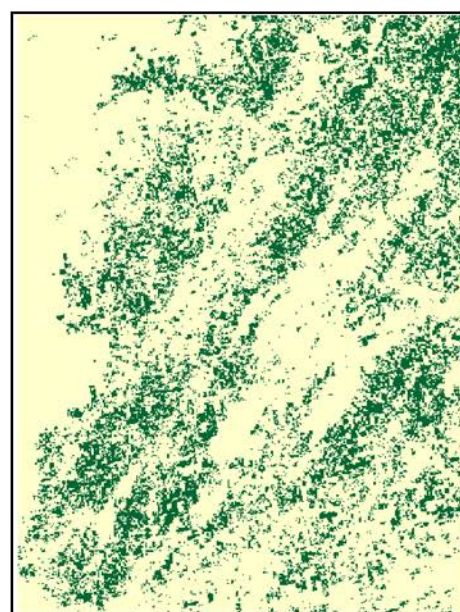


Figura 11.t. Combined::0.86
Quercus pyrenaica

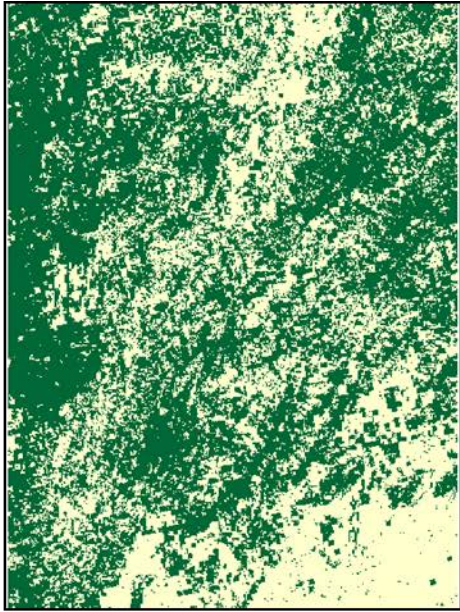


Figura 11.u. R. Forest:0.74
Rhamnus alaternus

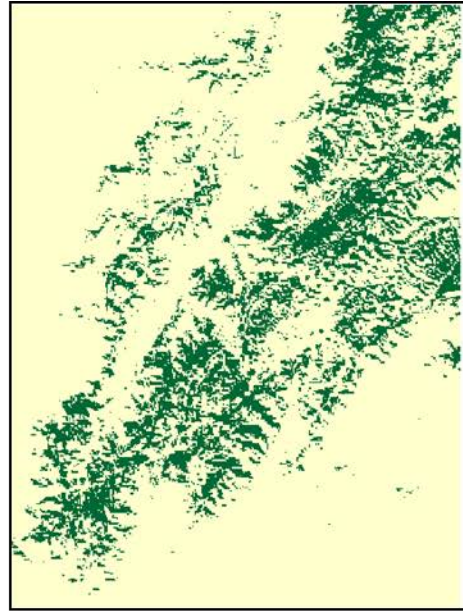


Figura 11.v. Combined:0.95
Sorbus aria

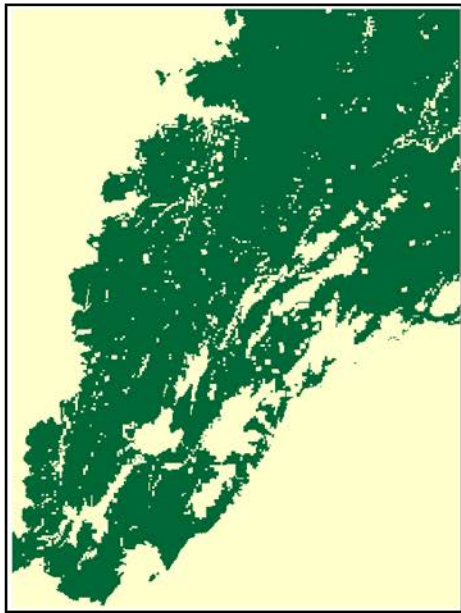


Figura 11.w. DOMAIN:0.99
Sorbus torminalis

Figura 11. Mapas de presencia/ausencia predicha para cada especie por su mejor SDM. Todos se encuentran a la misma escala. Se indica en verde los lugares identificados como aptos para la especie por el mejor modelo. Para cada especie se indica además cual fue el mejor método y su AUC.

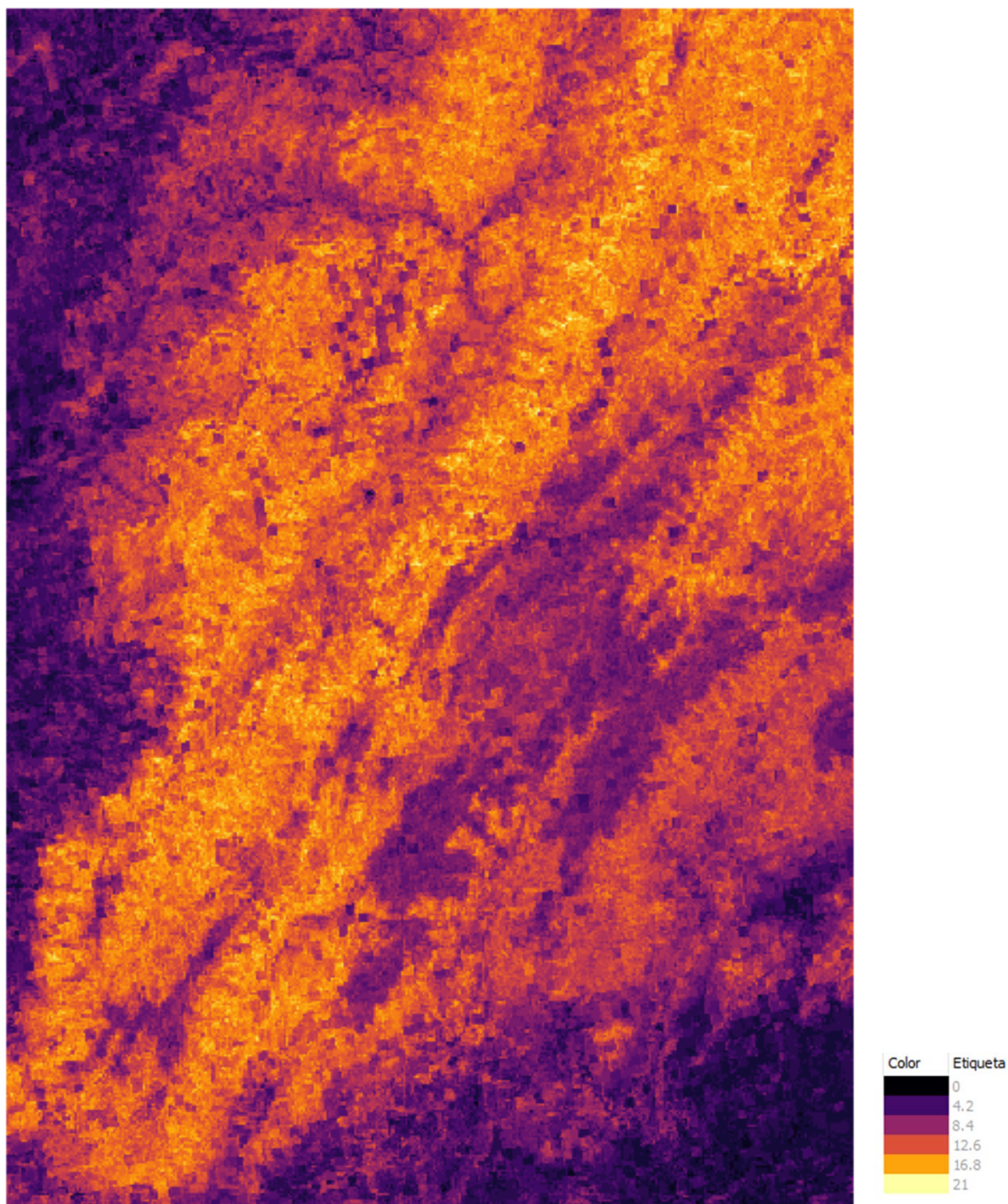


Figura 12. Mapa de riqueza de especies. Los colores más cálidos indican la idoneidad de ese lugar para un mayor número de especies.

La suma de los mapas de presencia/ausencia predicha para cada especie permite obtener un mapa de la riqueza forestal potencial del territorio (Figura 12). El número de especies idóneas por pixel varió entre 0 y 21, con un promedio de 10.30 ± 4.06 .

5 DISCUSIÓN

Rendimiento de los distintos métodos de ajuste de SDMs

Las causas que determinan la presencia-ausencia de una especie son de carácter multifactorial, acorde con las múltiples n-dimensiones de su nicho ecológico. Los modelos de distribución de especies intentan representar una información que sintetiza relaciones entre la distribución de una especie y las variables ambientales que determinan su nicho; relaciones que serían difíciles de interpretar por otros medios (Mateo et al., 2012). En el caso de las especies vegetales, suelen ser los factores climáticos y geográficos los que controlan en mayor medida la presencia-ausencia de una determinada especie en una escala geográfica amplia, aunque será un conjunto de factores bióticos (ej., herbivoría, presencia de banco de semillas, dispersión de semillas, etc.) y abióticos (tipo de suelo, pendiente, altitud, etc.) y su interacción los que determinarán su presencia a una escala geográfica más pequeña (Tilman, 1994.). Esta dependencia de múltiples factores y de la escala geográfica, dificulta la identificación de qué métodos de modelación son mejores en cada situación, haciendo que no sea posible recomendar el uso de uno sobre otro (Duan et al. 2014). Sin embargo mediante el examen de nuestros resultados se puede aportar información acerca de en qué circunstancias un modelo manifiesta tener un mejor rendimiento.

Nuestros resultados en cuanto al rendimiento de distintos métodos de elaboración de SDMs concuerdan ampliamente con los de Duan et al. (2014), ya que los modelos DOMAIN y BIOCLIM han dado un rendimiento claramente inferior al del resto de modelos, mientras que Mahalanobis y *Random Forests* alcanzaron los mayores valores promedio de AUC. En cualquier caso, si lo analizamos de manera individualizada, los valores alcanzados con *Random Forests* tampoco resultan altamente satisfactorios en términos generales, ya que solo en 4 de 23 especies han llegado a sobrepasar el valor AUC de 0.90 (valor que se considera como excelente en SDMs). Sin embargo, si consideramos el número de especies para las que los modelos alcanzan $AUC > 0.9$, el método con mejor rendimiento fue Mahalanobis, que alcanzó ese valor para 8 de las 23 especies estudiadas, seguido por DOMAIN y el modelo combinado (5 especies cada uno), *Random Forests* y Logístico (4 especies cada uno) y finalmente SVC y BIOCLIM que solamente alcanzaron este valor de AUC para dos especies. Por tanto podemos decir que *Random Forest* y Mahalanobis son métodos que rinden bastante bien para un diverso conjunto de especies. Aunque no se les puede considerar como idóneos en cualquier circunstancia, nuestros resultados sugieren su inclusión entre los métodos de SDM más recomendables.

La baja capacidad predictiva de los modelos BIOCLIM y DOMAIN se suele asociar a su excesiva simplicidad (ver Walker & Cocks, 1991; Carpenter et al., 1993; Baker et al., 2000; Hughes, 2003). A pesar de estas críticas, estos métodos siguen siendo útiles. Nuestros resultados sugieren que estos métodos son particularmente útiles para modelar especies que presentan una distribución geográfica muy restringida o una menor prevalencia y que, por tanto, proporcionan pocos datos de presencia (p. ej.: *Ilex aquifolium* y *Sorbus torminalis*). Dicho de otro modo, la distribución potencial de especies endémicas o situadas en su límite de distribución, parece explicarse peor con modelos complejos basados en presencia/ausencia que con modelos simples que utilizan solo datos de presencia. Muestra de ello tenemos el caso del acebo (*Ilex aquifolium*), que tiene su límite sur de distribución en las Sierras Béticas de andaluzas y norte de Marruecos. Esta especie requiere de suelos preferiblemente ácidos o neutros, húmedos y ricos en nutrientes, en lugares con buena exposición a la luz (Bañuelos et al. 2004). Es por tanto una especie ecológicamente muy restringida, que crece en nuestra zona de estudio enmarcada por unas condiciones climáticas y edáficas muy particulares. El acebo aparece en nuestra área de estudio tan solo en 19 de los 8693 puntos muestreados en el IFE3. En las Sierras de Cazorla, Segura y las Villas, que aunque predominantemente calcáreas, existen zonas muy localizadas de afloramientos silíceos en el norte y numerosos enclaves dispersos con suelos neutros debido a las elevadas precipitaciones. Luego, debido a que se reúnen puntualmente las singularidades necesarias para la presencia la especie, y a que la variabilidad de las condiciones en estos puntos es pequeña, un modelo simple como DOMAIN es capaz de alcanzar altos niveles de fiabilidad, con un AUC de 0.98. Por otro lado, los modelos que utilizan datos de presencia/ausencia son más exigentes en cuanto al balance entre el número de presencias y de ausencias, por lo que el bajo número de presencias inherente a las especies endémicas o situadas en sus límites de distribución hace que estos modelos tengan un peor rendimiento. En definitiva, nuestros resultados sugieren que los modelos de envoltorio climático simples no deben descartarse por defecto en la elaboración de SDMs.

Dejando al margen qué métodos puedan dar mejor o peor resultado, lo cierto es que en términos generales los modelos que hemos generado alcanzan un rendimiento bastante alto, con valores de AUC superiores a 0.9 para 11 de las 23 especies. Podemos comparar estos resultados con los de Siles et al. (2010) en los que se modelaron 13 de las especies que hemos incluido en el presente trabajo (Tabla 5). En su caso los modelos estadísticos utilizados fueron CART, GLM y GDA (*General Discriminant Analysis*, utilizan GLM para analizar funciones discriminantes, ajustando el modelo lineal con variables dependientes categóricas y continuas). De igual manera utilizaron la curva de ROC para comprobar la capacidad predictiva de los modelos ajustados con cada procedimiento estadístico. En su

caso ninguno de los modelos alcanzó valores de AUC superiores a 0.9, y el mejor modelo para 5 de las 13 especies alcanzó un poder predictivo inferior al 0.7.

Especies	Mejor modelo y su valor de AUC			
	Nuestros resultados		Siles, G. et al., 2010	
<i>Acer monspessulanum</i>	Combinado	0.93	GLM	0.59*
<i>Amelanchier ovalis</i>	Mahalanobis	0.93	GLM	0.80
<i>Arbutus unedo</i>	Combinado	0.91	GLM	0.87
<i>Celtis australis</i>	Random F.	0.84	CART	0.82
<i>Juniperus communis</i>	Random F.	0.97	GDA	0.77
<i>Juniperus oxycedrus</i>	Random F.	0.86	GLM	0.62
<i>Juniperus phoenicea</i>	Random F.	0.88	GLM	0.72
<i>Phillyrea latifolia</i>	Combinado	0.94	GLM	0.78
<i>Pinus nigra</i>	GLM	0.96	CART	0.89
<i>Pistacia terebinthus</i>	Random F.	0.83	GLM	0.78
<i>Quercus faginea</i>	BioClim	0.82	GLM	0.59*
<i>Quercus ilex</i>	Combinado	0.86	CART	0.50*
<i>Sorbus torminalis</i>	DOMAIN	0.99	GLM	0.67

Tabla 5. Comparación de los resultados de AUC obtenidos por el mejor modelo para cada especie extraídos en el presente trabajo, frente a los conseguidos en el trabajo de Siles et al. (2010). En asterisco se destacan aquellos valores más bajos conseguidos en ese anterior trabajo, manifestando su bajo rendimiento frente a este.

Las razones para esta gran diferencia en rendimiento entre ambos estudios, realizados para las mismas especies y en el mismo marco geográfico pueden ser múltiples. Por un lado, en el trabajo de Siles et al., no se emplearon algunas propiedades topográficas (p. ej. rugosidad, seno y coseno de la orientación) y edáficas que claramente pueden afectar a la distribución de especies de plantas, debido a la ausencia de capas con la extensión y resolución necesarias. Por otro lado, el tamaño de muestra disponible para Siles et al. era de solo 212 puntos de muestreo, frente a los 8693 recogidos del INF3 para este trabajo. En cualquier caso, el mayor rendimiento de modelado que hemos conseguido en este trabajo demuestra claramente que la elaboración de SDMs debe considerarse siempre una tarea inacabada, susceptible de mejora.

Correspondencia entre los mapas de distribución potencial y la distribución conocida de las especies.

La veracidad de un SDM no solo debe contrastarse mediante el uso de estadísticos tipo AUC, también es fundamental contrastar el resultado de los modelos, plasmado en el mapa de distribución potencial, frente al conocimiento científico acerca de la distribución geográfica e incidencia de la especie. Por ejemplo, es conocido que las dos especies de pinos y enebros que hemos modelado se reemplazan altitudinalmente, algo que se aprecia claramente en sus respectivos mapas de distribución potencial (Figura 11).

Una visión general de los 23 mapas que hemos generado (Figura 11) sugiere la existencia de dos patrones en cuanto al grado de incidencia de las especies en el área de estudio. Los SDMs de las especies típicamente dominantes en la vegetación de la zona (*Pinus nigra*, *P. halepensis*, *Juniperus communis*, *J. oxycedrus*, *Quercus ilex* y *Quercus faginea*) muestran una gran homogeneidad de la distribución potencial, mientras que muchas otras especies presentan distribuciones potenciales mucho más dispersas. A éste último grupo pertenecen especies de amplia distribución geográfica pero cuya presencia es siempre puntual, como ocurre con *Acer monspessulanum*, *Crataegus monogyna*, *Celtis australis*, *Quercus pyrenaica* y *Sorbus aria*. Estas especies tienen posiblemente unos requerimientos ambientales relativamente estrechos, que hacen que la idoneidad de un lugar pueda variar en escalas espaciales pequeñas. La realización de SDMs de alta resolución es capaz de captar esta “micro-heterogeneidad” y generar mapas de distribución mucho más detallados. De haber empleado píxeles de menor resolución (200 x 200 m o más), nuestros modelos no habrían podido captar este tipo de diferencias en la incidencia de las especies.

No obstante, el SDM de algunas especies que cuya distribución es típicamente dispersa se asemeja más al de las especies dominantes en cuanto a la continuidad de la presencia potencial de la especie. Los ejemplos más claros de esto son *Acer opalus*, *Amelanchier ovalis* y *Sorbus torminalis*. Esa discrepancia no parece deberse a la fiabilidad de sus respectivos modelos, ya que todas alcanzaron valores de AUC claramente superior a 0.9. Cabe la posibilidad, por tanto, de que los modelos sean correctos pero claramente incompletos para estas especies. Es decir, probablemente estas especies son capaces de tolerar un amplio rango de las condiciones ambientales existentes en la zona de estudio, pero algún factor ambiental (biótico o abiótico) que no hemos tenido en cuenta y que varía a pequeña escala espacial, podría ser muy limitante para la presencia o no de la especie.

Distribución espacial de la riqueza de especies

La riqueza de especies, así como la ubicación de centros de diversificación y endemismo (hot-spots), son parámetros útiles en la toma de decisiones para tareas de conservación (Cruz-Cárdenas, 2013). Como muestra la figura 12, las áreas con mayor riqueza potencial en nuestra zona de estudio corresponden a la media montaña, sobre todo en orientaciones NO a lo largo del Valle del Guadalquivir. Tanto las zonas de alta montaña (alineadas en dirección SO-NE) como, especialmente, las zonas periféricas de baja altitud (situadas la NO y SE) presentan una riqueza potencial muy baja. Las especies con mejor potencial para ocupar la zona SE de baja altitud son *Pinus halepensis* y *Juniperus oxycedrus*, mientras que la zona de baja altitud del NO es especialmente apta para *Olea europea*, *Rhamnus alaternus* y *Pistacia terebinthus*. En cuanto a las zonas de alta montaña, las especies más adecuadas son *Pinus nigra* y *Juniperus communis*.

6 CONCLUSIONES

Elaborar un modelado de distribución de especies mediante el manejo de RStudio proporciona un recurso fácil de usar. Este software permite realizar todo el proceso de modelado dentro del mismo marco de trabajo. El progreso de los SDMs se ha beneficiado durante más de dos décadas del desarrollo de herramientas como esta, por lo que ahora los SDMs son lo suficientemente maduros como para asumir un papel más importante en el apoyo a la conservación.

Un aspecto fundamental de la elaboración de SDMs es que no es fácil identificar qué métodos son mejores y, por lo tanto, no es posible recomendar el uso de un método sobre otro *a priori*. Nuestros resultados refrendan los de otros estudios en el sentido de que ningún modelo, por simple o complejo que sea, debe considerarse de antemano como el más adecuado en cualquier circunstancia. Debido a los incesantes avances en los SIG, el proceso de mejora continua de las bases de datos y al surgimiento de nuevos métodos de distribución potencial de especies, se potencia su aplicación y se abre paso a una disciplina sometida a una constante renovación.

Los sistemas naturales, debido a su inherente complejidad, no permiten una modelación de la distribución geográfica precisa y exacta (independientemente del modelo utilizado). En cualquier caso, las herramientas de modelación actuales y las bases de datos más recientes permiten generar modelos de distribución que se aproximan bastante a la realidad. Luego, utilidad real de los SDMs no debe de subestimarse, pues ofrecen un punto

de partida útil para predecir la distribución potencial tanto de diferentes especies como a distintas escalas.

Asociado a la eficacia de los modelos subyace la ya mencionada heterogeneidad ambiental de la vegetación a pequeña escala espacial. El empleo de datos de alta resolución permite apreciar la importancia de esta heterogeneidad al poder discriminar entre la extensión de los rangos de distribución (que pueden ser más o menos extensos) y la prevalencia de una especie que puede aparecer de forma más o menos dispersa dentro de su rango de distribución particular. Esta discriminación puede no tener especial importancia en el marco de estudios biogeográficos, pero es fundamental en estudios dirigidos a la planificación de la conservación de especies.

Por último, la distribución potencial de las 23 especies seleccionadas, asocia un alto nivel de riqueza en toda la franja diagonal de la zona de proyección, correspondiente a las laderas orientadas al NO en la media montaña del Valle del Guadalquivir. Un dato importante que ayuda a identificar áreas prioritarias para incrementar el nivel de protección de la biodiversidad de cara a un posible plan preventivo.

7 REFERENCIAS

- Alba-Sánchez, F., López-Sáez, J. A., Pando, B. B. D., Linares, J. C., Nieto-Lugilde, D., & López-Merino, L. 2010. Past and present potential distribution of the Iberian *Abies* species: a phytogeographic approach using fossil pollen data and species distribution models. **Diversity and Distributions** 16: 214-228.
- Allouche, O., Tsoar, A., & Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). **Journal of Applied Ecology** 43: 1223–1232.
- Araque Jiménez, E. 2013. Medio siglo de grandes incendios forestales en Andalucía (1961-2011). **Revue Géographique des Pays Méditerranéens** 121: 41-52.
- Araújo, M. B. & Williams, P. H. 2000. Selecting areas for species persistence using occurrence data. **Biological Conservation** 96: 331–345
- Araújo, M. B., & New, M. 2007. Ensemble forecasting of species distributions. **Trends in Ecology & Evolution** 22: 42-47.
- Baker, R.H.A., Sansford, C.E., Jarvis, C.H., Cannon, R.J.C., MacLeod, A. & Walters K.F.A. 2000. The role of climatic mapping in predicting the potential geographical distribution on non-indigenous pests under current and future climates. **Agriculture, Ecosystems & Environment** 8: 57–71
- Ballabio, C., Panagos, P., & Montanarella, L. 2016. Mapping topsoil physical properties at European scale using the LUCAS database . **Geoderma**, 261, pp. 110-123.
- Bañuelos, M. J., Kollmann, J., Hartvig, P., & Quevedo, M. 2003. Modelling the distribution of *Ilex aquifolium* at the north-eastern edge of its geographical range. **Nordic Journal of Botany** 2: 129-142.
- Benito, B., & Peñas, J. 2007. Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica. **GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica** 7: 100-119.
- Benito Garzón, M., Sánchez de Dios, R., & Sáinz Ollero, H. 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. **Ecography** 30: 120-134.

- Blonder, B. 2017. Hypervolume concepts in niche -and trait- based ecology. **Ecography** 41: 1441-1455
- Breiman, L. 2001. Random forests. **Machine learning** 45: 5-32.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. 1984. Classification and Regression Trees (Wadsworth & Brooks Cole, Monterey, CA). A powerful yet simple technique for ecological data analysis. **Ecology** 81: 3178-3192.
- Brown, M.P.S., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Haussler, D., 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. **Proceedings of the National Academy of Sciences of the United States of America** 97: 262–267.
- Busby, J. R. 1986. Bioclimatic prediction system (BIOCLIM) user's manual version 2.0. **Australian Biological Resources Study**. Canberra, Australia.
- Cabezudo, B et al., 2005. Lista Roja de la Flora Vascular de Andalucía. Consejería de Medio Ambiente de la Junta de Andalucía. Sevilla
- Calenge, C., Darmon, G., Basille, M., Loison, A., & Jullien, J. M. 2008. The factorial decomposition of the Mahalanobis distances in habitat selection studies. **Ecology** 89: 555-566.
- Carpenter, G., Gillison, A. N., & Winter, J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. **Biodiversity & Conservation** 2: 667-680.
- Consejería de Medio Ambiente: Dirección General de la Red de Espacios Naturales Protegidos y Servicios Ambientales, Servicio de Fomento de Espacios Naturales, 2003. Plan de Desarrollo Sostenible del Parque Natural de la Sierra de Cazorla, Segura y Las Villas
- Cruz-Cárdenas, G., Villaseñor, J. L., López-Mata, L., & Ortiz, E. 2013. Distribución espacial de la riqueza de especies de plantas vasculares en México. **Revista Mexicana de Biodiversidad** 84: 1189-1199.
- Fernández de Trespalacios, F. J. G. 2007. El agua subterránea en el Parque Natural de Sierra Magina (Jaen). Instituto Geológico y Minero de España. Madrid
- Duan, R. Y., Kong, X. Q., Huang, M. Y., Fan, W. Y., & Wang, Z. G. 2014. The predictive performance and stability of six species distribution models. **PLoS One**, 9: e112764.

- Edwards Jr, T. C., Cutler, D. R., Zimmermann, N. E., Geiser, L., & Moisen, G. G. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. **Ecological Modelling** 199: 132-141.
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A. & Li, J. 2006. Novel methods improve prediction of species' distributions from occurrence data. **Ecography**, 29: 129-151.
- Ferrier, S., & Guisan, A. 2006. Spatial modelling of biodiversity at the community level. **Journal of Applied Ecology** 43: 393-404.
- Ferrier, S. 1984. The Status of the Rufous Scrub Bird *Atrichornis Rufescens*: Habitat, Geographical Variation and Abundance. Doctoral dissertation. Royal Australasian Ornithologists Union.
- Franklin, S. E. & Wulder, M. A. 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. **Progress in Physical Geography** 26: 173–205.
- Friedman, J. H. 1991. Multivariate adaptive regression splines. **The Annals of Statistics** pp.1-67
- Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. 2004. New developments in museum-based informatics and applications in biodiversity analysis. **Trends in Ecology & Evolution** 19: 497-503.
- Grinnell, J. 1917a. The niche-relationships of the California Thrasher. **The Auk** 34: 427-433.
- Grinnell, J. 1917b. Field tests of theories concerning distributional control. **The American Naturalist** 51: 115-128.
- Groom, M. J., Meffe, G. K., & Carroll, C. R. 2006. Principles of conservation biology. **Sinauer Associates**. Massachusetts, U.S.A. 779 pp.
- Guisan, A., & Harrell, F. E. 2000. Ordinal response regression models in ecology. **Journal of Vegetation Science** 11: 617-626.
- Guisan, A., & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. **Ecological Modelling** 135: 147-186.
- Guisan, A., Theurillat, J. P., & Kienast, F. 1998. Predicting the potential distribution of plant species in an alpine environment. **Journal of Vegetation Science** 9: 65-74.

- Guisan, A., Edwards Jr, T. C., & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. **Ecological Modelling** 157: 89-100.
- Guo, Q., Kelly, M., & Graham, C. H. 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. **Ecological Modelling** 18: 75-90.
- 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.
- Hastie, T., & Tibshirani, R. 1987. Generalized additive models: some applications. **Journal of the American Statistical Association** 82: 371-386.
- Hirzel, A., & Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modeling. **Ecological Modelling** 157: 331-341.
- Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data?. **Ecology** 83: 2027-2036.
- Houlder D., Hutchinson M., Nix H. A., & McMahon J. 1999. 'ANUCLIM version 5.0 User Guide. Centre for Resource and Environmental Studies. University of Canberra. Australia.
- Hugall, A. *et al.* 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). **Proceedings of the National Academy of Sciences** 99: 6112–6117.
- Hughes, L. 2003. Climate change and Australia: Trends, projections and impacts. *Austral. Ecology* 28, 423–443.
- Joachims, T., 1998. Text categorization with support vector machines: learn with many relevant features. In: *Proceedings of European Conference on Machine Learning*. Springer-Verlag. Berlin, pp. 137–142.
- Nally, R. M., & Fleishman, E. 2004. A successful predictive model of species richness based on indicator species. **Conservation Biology** 18: 646-654.
- Manel, S., Williams, H. C., & Ormerod, S. J. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. **Journal of Applied Ecology**, 38: 921-931.

- Mateo, R. G., Felicísimo, Á. M., & Muñoz, J. 2011. Modelos de distribución de especies: Una revisión sintética. **Revista Chilena de Historia Natural** 84: 217-240.
- Mateo, R. G., Felicísimo, Á. M., & Muñoz, J. 2012. Modelos de distribución de especies y su potencialidad como recurso educativo interdisciplinar. **Reduca** 5: 137-153.
- McCullagh P & Nelder, J.A., 1989. Generalized Linear Models. Second Edition. Monographs on Statistics and Applied Probability. Chapman & Hall. Londres. Inglaterra.
- Molina, M. 2018. Pruebas diagnósticas con resultados continuos o politómicos. Curvas ROC. **Evid Pediatr.** 2017: 13-12.
- Nix, H.A. & Switzer, M.A. 1991. Rainforest animals: atlas of vertebrates endemic to Australia's wet tropics. Australian National Parks and Wildlife Service. Canberra. Australia.
- Olivier, F., & Wotherspoon, S. J. 2006. Modelling habitat selection using presence-only data: case study of a colonial hollow nesting bird, the snow petrel. **Ecological Modelling** 195: 187-204.
- Ottaviani, D., Lasinio, G. J., & Boitani, L. 2004. Two statistical methods to validate habitat suitability models using presence-only data. **Ecological Modelling**, 179: 417-443.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. **The Quarterly Review of Biology** 78: 419-433.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. **Ecological Modelling** 190: 231-259.
- Pliscoff, P., & Fuentes-Castillo, T. 2011. Modelación de la distribución de especies y ecosistemas en el tiempo y en el espacio: una revisión de las nuevas herramientas y enfoques disponibles. **Revista de Geografía Norte Grande** 48: 61-79.
- Pulliam, H. R. 1988. Sources, sinks, and population regulation. **The American Naturalist** 132: 652-661.
- Richardson, D. M., & Whittaker, R. J. 2010. Conservation biogeography foundations, concepts and challenges. **Diversity and Distributions** 16: 313-320.
- Siles, G., Alcántara, J. M., Rey, P. J., & Bastida, J. M. 2010. Defining a target map of native species assemblages for restoration. **Restoration Ecology** 18: 439-448.

- Scotts, D. & Drielsma, M., 2003. Developing landscape frameworks for regional conservation planning: an approach integrating fauna spatial distributions and ecological principles. **Pacific Conservation Biology** 8: 235–254
- Stockwell, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. **International Journal of Geographical Information Science** 13: 143-158.
- Thomas, C. D. *et al.* 2004. Extinction risk from climate change. **Nature** 427: 145–148.
- Tilman, D. 1994. Competition and Biodiversity in Spatially Structured Habitats. **Ecology** 75:2-16.
- Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., & Kadmon, R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. **Diversity and Distributions** 13: 397-405.
- Valle Tendero, F., Gomez Mercado, F., Mota Poveda, J. F., & Díaz de la Guardia, C. 1989. Parque natural de Cazorla, Segura y las Villas. Guía botánico-ecológica, Editorial Rueda, Madrid.
- Vapnik, V. 2013. The nature of statistical learning theory. Springer science & business media. New York, E.E.U.U.
- Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. **Journal of Biogeography** 36: 2290-2299.
- Walker, P. A., & Cocks, K. D. 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. **Global Ecology and Biogeography Letters**, pp.108-118.
- Williams, J. W., & Jackson, S. T. 2007. Novel climates, no analog communities, and ecological surprises. **Frontiers in Ecology and the Environment** 5: 475-482.
- Zaniewski, A. E., Lehmann, A., & Overton, J. M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. **Ecological Modelling** 157: 261-280.