

## INFLUENCIA DEL USO DEL SUELO EN LA DISTRIBUCIÓN DE LA SABINA ALBAR (*Juniperus thurifera* L.) EN LA PENÍNSULA IBÉRICA.

**Gastón González, A.**

*Departamento de Producción Vegetal: Botánica y Protección Vegetal, Universidad Politécnica de Madrid, E.U.I.T. Forestal. Av. Ramiro de Maeztu s/n. 28040 Madrid. [aitor.gaston@upm.es](mailto:aitor.gaston@upm.es)*

**Resumen** Con el objetivo de estudiar la influencia del uso del suelo en la distribución de la sabina albar (*Juniperus thurifera* L.) en la Península Ibérica, se han construido modelos de regresión logística que predicen la probabilidad de presencia de la especie en función del clima y el uso del suelo. Se han ajustado los modelos utilizando datos de presencia/ausencia de la especie a dos resoluciones espaciales: cuadros UTM de 1x1 y de 10x10 km. La precisión de los modelos se ha evaluado utilizando el área bajo la curva ROC. Los resultados en ambas resoluciones espaciales son similares en cuanto a las variables independientes seleccionadas. La precisión de las estimaciones del modelo aumenta cuando se añade el uso del suelo, como variable independiente, a un modelo puramente climático.

**Palabras clave:** Corología, autoecología, regresión logística, curva ROC.

### LAND USE EFFECT ON THE DISTRIBUTION OF THE THURIFEROUS JUNIPER (*JUNIPERUS THURIFERA* L.) IN THE IBERIAN PENINSULA.

**Abstract** In order to assess land use effect on the distribution of the thuriferous juniper (*Juniperus thurifera* L.) in the Iberian Peninsula, logistic regression models have been used to predict presence probability in response to climate and land use. Presence/absence data from two spatial resolutions (1x1 and 10x10 km squares) have been used to fit the models. Area under the ROC curve has been used to assess model performance. Results in both spatial resolutions are similar regarding to selected predictors. Adding land use as predictor to a purely climatic model improves its performance.

**Keywords:** Chorology, autoecology, logistic regression, ROC curve.

### INFLUENCE DE L'USAGE DU SOL SUR LA DISTRIBUTION DES PLANTS DE GENEVRIERS (*JUNIPERUS THURIFERA* L.) DANS LA PENINSULE IBERIQUE.

**Résumé** Afin d'étudier l'influence de l'utilisation du sol sur la distribution des plants de genévriers (*Juniperus thurifera* L.) dans la Péninsule Ibérique, ont été élaborés des modèles de régression logistique qui prévoient la probabilité de présence de l'espèce en fonction du climat et de l'usage du sol. En utilisant les données présence/absence de l'espèce, ces modèles ont été mis au point pour deux résolutions spatiales : des tableaux de 1x1km et de 10x10km. On évalue la précision de ces modèles avec l'aire qui se trouve sous la courbe ROC. Avec les deux résolutions spatiales, les résultats obtenus sont similaires pour les variables indépendantes sélectionnées. La précision des estimations augmente lorsqu'on ajoute à ce modèle purement climatique l'usage du sol comme variable indépendante.

**Mots-clés:** Chorologie, autoécologie, régression logistique, courbe ROC.

## INTRODUCCIÓN

Los intentos de explicar la distribución geográfica de las especies a partir de modelos basados en los factores ecológicos comenzaron en la primera mitad del Siglo XX, pero se han hecho verdaderamente abundantes en los últimos treinta años (GUISAN & THUILLER, 2005). Los modelos de distribución de especies se han construido a diferentes escalas espaciales, desde el detalle de la parcela (véanse trabajos citados en SÁNCHEZ PALOMARES, 2001) a la baja resolución de los atlas corológicos continentales (THUILLER *et al.*, 2003).

En las últimas dos décadas se han publicado numerosos mapas corológicos de plantas vasculares ibéricas, véanse, por ejemplo, las series de artículos a este respecto iniciadas por

FERNÁNDEZ CASAS (1985) o MOLINA (1989), en las que se recopilan datos de herbarios y publicaciones florísticas que se representan en forma de presencia de las especies en cuadros UTM de 10 km. Los modelos de distribución de especies permiten obtener valiosa información autoecológica a partir de los mapas corológicos (THUILLER *et al.*, 2003). Dicha información puede servir para contrastar y formular hipótesis sobre los hábitats de las especies y relacionarla con los datos disponibles a escalas más detalladas. Además para muchas especies no se dispone de información a nivel de parcela para toda la península, para estas especies los modelos basados en mapas corológicos constituyen posiblemente la mejor fuente de información autoecológica disponible.

El factor ecológico más relevante en la distribución de las plantas es el clima (WOODWARD, citado en THUILLER *et al.*, 2004). Lo mismo se puede afirmar para la sabina albar (*Juniperus thurifera* L.) a la vista de la buena capacidad predictiva de un modelo puramente climático para los sabinares peninsulares (GARCÍA & ALLUÉ, 2005). Este hecho no debe hacernos olvidar que las características del suelo, la competencia con otras especies o la actividad humana pueden influir decisivamente en la distribución de las especies, sobre todo cuando se estudia a escala local.

Este trabajo es parte de una investigación más amplia que tiene como objetivo principal diseñar métodos eficaces para la caracterización autoecológica de plantas vasculares a partir de mapas corológicos. En esta ocasión se aborda el problema de utilizar los usos del suelo como variable explicativa, con el objeto de entender mejor la diferencia entre las áreas potenciales obtenidas a partir de modelos climáticos y las áreas de distribución observadas.

Partiendo de un modelo de distribución de la sabina albar exclusivamente climático, se ha estudiado si los usos del suelo contribuyen de forma estadísticamente significativa al modelo y si aumenta la capacidad predictiva del mismo. Esta comparación se ha realizado a dos escalas espaciales diferentes: cuadros UTM de 10 km y 1 km.

## **MATERIAL Y MÉTODOS**

### **Selección del modelo estadístico**

El modelo estadístico seleccionado, la regresión logística, se usa frecuentemente en modelos de distribución de especies (GUISAN & ZIMMERMANN, 2000). Este tipo de regresión es un caso particular del modelo lineal generalizado adecuado para explicar la presencia y ausencia de especies en función de variables ambientales cuantitativas (LEGENDRE & LEGENDRE, 1998). Se pueden usar funciones lineales o cuadráticas de las variables explicativas que producirán curvas de respuesta sigmoidales o gaussianas. Una vez ajustado, el modelo devuelve probabilidades de presencia de la especie en función de las variables ambientales. Las probabilidades se pueden convertir en estimaciones de presencia/ausencia fijando un umbral de probabilidad, generalmente 0,5.

### **Muestreo**

Los resultados de los modelos de regresión logística son sensibles a la proporción de casos positivos en la muestra (prevalencia) y mejoran cuando dicha proporción es cercana a la mitad (McPHERSON *et al.*, 2004). Para evitar problemas derivados de la prevalencia, las muestras se han construido incluyendo todos los cuadros con presencia conocida de sabina albar y un número igual de cuadros tomados al azar fuera del área de distribución conocida.

Se han elaborado dos muestras a resoluciones espaciales diferentes: cuadros UTM de 1 y 10 km de lado. Los cuadros de 10 km se han tomado de GASTÓN & SORIANO (2006). Los cuadros de 1 km de lado se han obtenido a partir de las teselas del Mapa Forestal de España (RUIZ DE LA TORRE, 1990-99), de dos formas: (1) todos los cuadros UTM de 1 km superpuestos a teselas en las que la agrupación vegetal dominante es el sabinar albar y (2) los cuadros que incluyen el centro de las teselas en las que se cita la presencia de sabina albar sin ser sabinares. En este último caso, se han descartado las teselas que tengan algún punto a más de 2 km del centro para minimizar la incertidumbre respecto a la localización. Para reducir errores se ha forzado a que las presencias en la muestra de 1 km pertenezcan a cuadros de 10 km con citas de sabina albar según GASTÓN & SORIANO (2006) y viceversa para las ausencias. Las muestras se han dividido al azar en dos submuestras de igual tamaño, una de ellas se ha utilizado para calibrar el modelo y otra para validarlo. El tamaño de muestra resultante ha sido de 1216 cuadros de 10 km y 16322 de 1 km.

### **Variables explicativas**

Las variables climáticas utilizadas (véase tabla 1) son las propuestas en SÁNCHEZ PALOMARES (2001). Las precipitaciones y temperaturas se han estimado utilizando el programa ESTCLIMA (SÁNCHEZ PALOMARES *et al.*, 1999) aplicado al modelo digital de elevaciones GTOPO30 (U.S. Geological Survey), cuyo tamaño de celda es de aproximadamente 1 km de lado. El resto de variables se han calculado a partir de las anteriores. A cada cuadro de la muestra se le ha asignado el promedio de las variables climáticas de las celdas del GTOPO30 que le corresponden. McPHERSON *et al.* (2005) denominan a esta forma de asignar valores de las variables a los cuadros “direct approach”, sus datos no muestran diferencias significativas en la precisión de las predicciones entre modelos ajustados con datos obtenidos de esta manera y los ajustados con datos de mayor resolución.

Con el objeto de evitar la multicolinealidad, que es un problema para el ajuste de modelos de regresión (LEGENDRE & LEGENDRE, 1998), se estudiaron las relaciones entre las variables climáticas utilizando el Análisis de Componentes Principales (véase figura 1). Las variables cuyos vectores se encuentran separados por ángulos cercanos a cero están muy correlacionadas entre sí. De cada grupo de variables correlacionadas se ha seleccionado una que las represente (véase tabla 1 y figura 1).

Como indicadores de los usos del suelo se han calculado para cada cuadro las proporciones en tanto por uno de superficie arbolada (ARBO), matorrales (MAT), herbazales (HERB) y cultivos (CULT). Los datos para estas variables se han obtenido a partir del Programa Corine Land Cover (©EEA, 1990) que aporta información de usos del suelo para celdas de 200 m de lado. Estas variables son, en la mayoría de los cuadros, una combinación lineal de las otras ( $1 = CULT + ARBO + MAT + HERB$ ) y esto es un problema para el ajuste de modelos de regresión múltiple. Para evitar este problema solo se ha usado la variable “porcentaje de superficie cultivada” (CULT), ya que es la que mejor explica por sí sola la distribución de la sabina albar, según una regresión logística (véase figura 2).

### **Ajuste de los modelos**

El modelo se ha ajustado usando un procedimiento por pasos que utiliza el estadístico Wald, partiendo de un modelo con todas las variables y eliminando aquellas cuya ausencia no reduzca significativamente la verosimilitud.

En total se han ajustado cuatro modelos: (1) resolución 10 km y solo variables climáticas, (2) añadiendo el uso del suelo al anterior, (3) resolución 1 km y solo variables climáticas y (4) añadiendo el uso del suelo al anterior. Para el ajuste solo se ha usado la mitad de cada muestra, separada de forma aleatoria de la submuestra usada para la validación.

### **Validación de los modelos**

Los modelos obtenidos se han evaluado por su capacidad de discriminación de presencias y ausencias en la submuestra no usada para el ajuste. McPHERSON *et al.* (2004) recomiendan el uso del área bajo la curva ROC (AUC), por ser independiente del umbral de discriminación de positivos y de la prevalencia en la muestra. Este estadístico puede alcanzar valores entre 0 y 1. Los valores mayores de 0,5 indican que el modelo predice mejor que el puro azar y los mayores de 0,9 que la precisión del modelo es buena (SWETS, citado en McPHERSON *et al.*, 2005).

El ajuste y la validación de los modelos se ha realizado con el programa estadístico SPSS (versión 11).

## **RESULTADOS**

La precisión de las predicciones de todos los modelos es buena ( $AUC > 0,9$ ), pero es mejor en los modelos de resolución de 1 km que en los de 10 km. Las variables seleccionadas varían poco entre los modelos ajustados a resoluciones espaciales diferentes (véase tabla 2).

En ambas resoluciones la introducción de la variable indicadora de los usos del suelo ha contribuido de forma estadísticamente significativa a aumentar la verosimilitud del modelo, ya que si no fuera así la variable CULT no hubiera sido seleccionada por el procedimiento de ajuste (véase tabla 2).

En ambas resoluciones la incorporación de la proporción de superficie cultivada en el cuadro (CULT) ha resultado en un ligero aumento de la AUC y por lo tanto de la calidad de ajuste del modelo (véase tabla 2).

La incorporación de la variable CULT ha supuesto una mejora en la predicción en regiones potencialmente adecuadas para la sabina albar y dominadas por los cultivos (véase la provincia de Palencia en las figuras 3 y 4). Según el modelo, la sabina albar tiende a desaparecer cuando el porcentaje de superficie cultivada supera el 90 %. La actividad agrícola parece no afectarle hasta que alcanza niveles altos (véase figura 2).

## **DISCUSIÓN**

Los resultados de este estudio apoyan la idea de que los usos de suelo pueden servir para afinar los modelos climáticos de distribución de especies. La contribución este tipo de variables a la mejora de los modelos parece disminuir al reducirse la resolución de los datos utilizados. Al añadir variables de uso del suelo a modelos climáticos, THUILLER *et al.* (2004) no encontraron un aumento de la capacidad predictiva de los modelos de distribución

de animales y plantas a escala continental (cuadros de 50 km), pero PEARSON *et al.* (2004) si lo hicieron a resoluciones más altas (cuadros de 10 km). En todo caso, parece que los usos del suelo contribuyen a la capacidad explicativa del modelo incluso a bajas resoluciones (THUILLER *et al.*, 2004).

Las variables de usos del suelo más útiles serán aquellas menos correlacionadas con el clima, THUILLER *et al.* (2004) señalan la superficie cultivada como una de ellas y los resultados de este estudio apoyan esta afirmación.

En cuanto a la distribución ibérica de la sabina albar, los resultados de este trabajo apoyan la hipótesis de que la sabina ha sido desplazada por los cultivos de lugares climáticamente idóneos pero de topografía más favorable para la actividad agrícola (GARCÍA & ALLUÉ, 2005).

En la medida que se integren nuevos factores ambientales a los modelos de distribución de la sabina albar se podrá avanzar en el conocimiento de sus relaciones con el medio y, tal vez, pronosticar la dinámica de su área de distribución. En este sentido es fundamental integrar en los modelos información sobre la capacidad de dispersión para valorar el alcance de la colonización de áreas en las que se abandonen cultivos.

## CONCLUSIONES

El uso de variables que informan sobre los usos del suelo ha mejorado la precisión de los modelos climáticos de distribución de la sabina albar. Esta mejora ha sido ligera y más acusada a mayor resolución espacial de los datos, pero estadísticamente significativa en ambas resoluciones consideradas, cuadros de 1 y 10 km.

Según el modelo, la sabina albar parece no verse afectada por la actividad agrícola hasta que no se alcanza un porcentaje alto de superficie cultivada. El uso del suelo puede explicar la ausencia de la especie en zonas climáticamente idóneas pero ocupadas en su mayoría por cultivos.

### Agradecimientos

No puedo dejar pasar la ocasión de expresar mi más sincero agradecimiento a Carlos Soriano y a las personas que colaboraron en la recopilación de los datos del Mapa Forestal, a Juan Ignacio García Viñas por sus sabios consejos, a Enrique Sadornil, Alejandro Vivar y Pepa Aroca por la ayuda en temas estadísticos y a Marta Ampudia por revisar el borrador.

## BIBLIOGRAFÍA

- GARCÍA LÓPEZ, J.M. & ALLUÉ CAMACHO, C.; 2005. Caracterización y potencialidades fitoclimáticas de la sabina albar (*Juniperus thurifera* L.) en la Península Ibérica. Invest. Agrar.: Sist. Recur. For. 14(1): 98-109.
- GASTÓN, A. & SORIANO, C.; 2006. Contribution of the Forest Map of Spain to the chorology of woody plants. Invest. Agrar.: Sist. Recur. For. (aceptado).
- GUISAN, A. & THUILLER, W.; 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8 (9): 993-1009.
- GUISAN, A. & ZIMMERMANN, N.E.; 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135: 147-186.

- LEGENDRE, P. & LEGENDRE, L.; 1998. Numerical ecology. Second English edition. Developments in environmental modelling 20. Elsevier Science B.V. Amsterdam. 853 pp.
- McPHERSON, J. M.; JETZ, W. & ROGERS, D. J.; 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?. *J. Appl. Ecol.* 41 (5): 811-823.
- McPHERSON, J. M.; JETZ, W. & ROGERS, D. J.; 2005. Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions - possibilities and limitations. *Ecol. Model.* 192: 499-522.
- PEARSON, R. G.; DAWSON, T. P. & LIU, C.; 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography* 27 (3): 285-298.
- RUIZ DE LA TORRE, J. (ed.); 1990-1999. Mapa Forestal de España. Ministerio de Medio Ambiente. Madrid, 93 vol.
- SANCHEZ PALOMARES, O.; 2001. Los estudios autoecológicos paramétricos de especies forestales. Modelos digitales. III Congreso Forestal Español. Granada.
- SÁNCHEZ PALOMARES, O.; SÁNCHEZ SERRANO, F. & CARRETERO CARRERO, M.P.; 1999. Modelos y cartografía de estimaciones climáticas termopluviométricas para la España peninsular. INIA, col. Fuera de Serie. Madrid. 192 pp.
- THUILLER, W.; ARAÚJO, M. B. & LAVOREL, S.; 2003. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *J. Veg. Sci.* 14: 669-680.
- THUILLER, W.; ARAÚJO, M. B. & LAVOREL, S.; 2004. Do we need land-cover data to model species distributions in Europe?. *J. Biogeogr.* 31: 353-361.

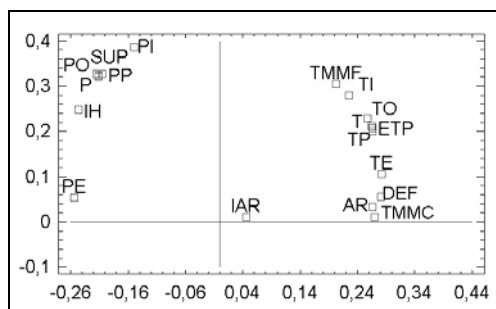
## TABLAS Y FIGURAS

Variable	Descripción	Representada por
P	Precipitación media anual	P
PI	P. media invernal	P
PP	P. media primaveral	P
PE	P. media estival	PE
PO	P. media otoñal	P
T	Temperatura media anual	T
TI	T. media invernal	T
TP	T. media primaveral	T
TE	T. media estival	T
TO	T. media otoñal	T
TMMF	T. media de las mínimas del mes más frío	TMMF
TMMC	T. media de las máximas del mes más cálido	TMMC
ETP	Evapotranspiración potencial (Thornthwaite)	T
SUP	Superávit hídrico anual (Thornthwaite)	P
DEF	Déficit hídrico anual (Thornthwaite)	TMMC
IH	Índice hídrico anual (Thornthwaite)	P
AR	Duración del periodo árido (Walter-Lieth)	TMMC
IAR	Intensidad de la aridez (Walter-Lieth)	IAR

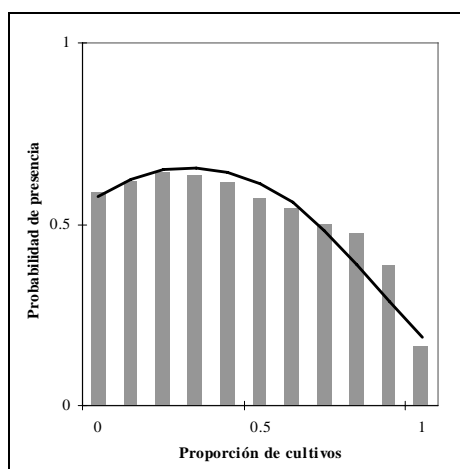
**Tabla 1.** Descripción de las variables climáticas consideradas inicialmente y las finalmente usadas para ajustar el modelo (última columna).

Modelo	Resolución	Variables introducidas	Variables seleccionadas	AUC
1	10 km	Solo climáticas	P, PE2, T, TMMC2, TMMF2	0,906
2	10 km	Clima y usos del suelo	P2, PE2, T, TMMC2, TMMF2, CULT2	0,913
3	1 km	Solo climáticas	P2, PE2, T, TMMC, TMMF2, IAR2	0,968
4	1 km	Clima y usos del suelo	P2, PE2, T, TMMC, TMMF2, IAR2, CULT2	0,971

**Tabla 2.** Modelos ajustados, variables y precisión de las predicciones. El número 2 tras el nombre de la variable indica que el término cuadrático es estadísticamente significativo.



**Figura 1.** Resultado del Análisis de Componentes Principales. Variables climáticas representadas frente a los dos primeros componentes



**Figura 2.** Probabilidad de presencia de sabina albar en función de la proporción de superficie cultivada en cuadros de 1 km (regresión logística,  $R^2$  de Nagelkerke = 0,15). Las barras representan la frecuencia relativa de la especie en clases de proporción de cultivos.

