

MODELOS PREDICTIVOS DE DISTRIBUCIÓN DE ESPECIES DE VEGETACIÓN POTENCIAL EN EL INCENDIO DEL “PUERTO DE LAS PALOMAS”, SIERRA DE CAZORLA”, JAÉN

Gemma Siles, Jesús Bastida, Pedro J. Rey y Julio Alcántara

Departamento de Biología Animal, Biología Vegetal y Ecología, Universidad de Jaén. Paraje Las Lagunillas, s/n; Edificio B-3. 23071-JAÉN (España). Correo electrónico: gsiles@ujaen.es.

Resumen

El objetivo de este trabajo es elaborar modelos de la distribución espacial de especies de la vegetación potencial en el incendio del “Puerto de Las Palomas” (Cazorla, Jaén). Para ello se utilizan tres aproximaciones analíticas: Modelos Lineales Generalizados (GLIM), Análisis Discriminante general (GDA) y Diagramas de regresión (CART). Al identificar así las zonas de plantado más adecuadas para cada especie, se trata de conseguir una cubierta vegetal semejante en composición de especies y distribución espacial a las características de la vegetación potencial de la zona. Además, este procedimiento debe permitir maximizar el éxito de plantación y minimizar costes de cara a la restauración del área afectada. Las variables ambientales usadas para los modelos de distribución fueron: altitud, orientación, existencia de barrancos y ríos, profundidad del suelo, porcentaje de sustrato rocoso, pendiente y coordenadas UTM. Los modelos se desarrollaron a partir de 212 observaciones en zonas próximas al incendio con una vegetación natural poco alterada. De éstas, 162 se utilizaron exclusivamente para ajustar los modelos y las 50 restantes se utilizaron exclusivamente para comprobar la fiabilidad de cada modelo. A partir del modelo más fiable conseguido para cada especie se ha desarrollado un mapa predictivo de su distribución potencial en el área incendiada. Se ilustran los resultados con *Pinus nigra* y *Buxus sempervirens*. En ambos casos, los modelos predictivos reflejan la importancia de aquellas variables ambientales que, de forma cualitativa, han indicado diversos estudios. Sin embargo, nuestros modelos predictivos, al asignar valores cuantitativos a los efectos de esas variables, permiten elaborar predicciones sobre la presencia/ausencia de las especies de forma más precisa a una escala espacial útil en labores de regeneración de la vegetación.

Palabras clave: Restauración, GLIM, CART, GDA, AUC

INTRODUCCIÓN

La restauración de la vegetación potencial implica la re-introducción de numerosas especies, cada una con distintos requerimientos ambientales. Por tanto, para maximizar el éxito de la restauración es necesario analizar la idoneidad del territorio para cada especie. Es más, dado

que tras un incendio gran parte de la vegetación ha desaparecido, un paso previo a la restauración debería ser la elaboración de un modelo que, a partir de información física (clima, topografía y suelo), asigne la presencia/ausencia potencial de cada especie en distintas parcelas de la zona a restaurar. Este tipo de modelos están siendo utilizados, por ejemplo, para evaluar la sensibilidad de

la vegetación a influencias medioambientales (FRANKLIN, 1995), desarrollar predicciones sobre la distribución geográfica potencial de especies y comunidades vegetales, y evaluar patrones anómalos o alteraciones de las distribuciones (REICHAR & HAMILTON, 1997), así como para la predicción de cambios en las comunidades debidos al cambio climático (THUILLER, 2003). En este trabajo se analiza la distribución de *Pinus nigra* y *Buxus sempervirens*, componentes de la vegetación potencial del área de estudio (VALLE et al., 2003), presentándose un modelo predictivo de su distribución en las distintas unidades en que se ha dividido el área del incendio.

Objetivos

Analizar la idoneidad de cada zona para cada especie objetivo, limitando el plantado de cada especie a zonas potencialmente aptas para ellas, de forma que se genere una distribución natural de la vegetación, lo que contribuirá a aumentar la diversidad natural, el valor ecológico y la calidad paisajística de la zona.

METODOLOGÍA

Área de Estudio

La zona de estudio se localiza en el Parque Natural de las Sierras de Cazorla, Segura y Las Villas (provincia de Jaén), en la zona afectada por el incendio acaecido en Agosto de 2001 en la zona del “Puerto de Las Palomas” (ver Figura 1). Presenta un relieve muy accidentado, con fuertes pendientes, suelos poco profundos (en su mayoría litosoles, aunque aparecen también regosoles calcáreos y cambisoles cálcicos) y gran presencia de afloramientos rocosos. Las precipitaciones medias rondan los 1000 mm (ombrotipo subhúmedo). La vegetación previa al incendio eran pinares de repoblación de *Pinus pinaster* y puntualmente de *P. halepensis* y *P. nigra*, aunque existían pequeños rodales de vegetación natural y, en algunas zonas, abundante sotobosque de matorral nativo (LUQUE, 1995). Por tanto, la vegetación se encontraba muy alterada con respecto a las series de vegetación potencial propuestas por VALLE et al.

(2003) para la zona: serie edafoxerófila meso-supramediterránea subhúmeda bética del pino salgareño (*Pinus nigra*); serie meso-supramediterránea basófila del quejigo (*Quercus faginea*) en su faciación típica y umbrófila con *Pistacia terebinthus*; serie supramediterránea seco-subhúmeda de la encina (*Quercus rotundifolia*); y la serie mesomediterránea seco-subhúmeda de la encina (*Quercus rotundifolia*) en su faciación típica y termófila con *Pistacia lentiscus*.

Muestreo de vegetación y medio físico

El área del incendio se dividió en 746 cuadrículas georreferenciadas de 100 x 100 m (unidades espacialmente explícitas). En cada nodo de la malla se muestreó una parcela de 10 x 10 m., realizándose estimaciones sobre la presencia y cobertura de especies leñosas así como de las siguientes variables ambientales: *altitud* (relacionada con aspectos climáticos generales); *orientación* (determina el grado de insolación y el riesgo e intensidad de heladas), presencia de *barrancos* y *ríos* (que crean microclimas particulares); *pendiente* (relacionada con las características generales del suelo); *profundidad del suelo* (relacionada con la capacidad de retención de agua y disponibilidad de materia orgánica); *cobertura de roca* (relacionada con la estructura del suelo y la disponibilidad de espacio para el asentamiento de plantas); y *coordenadas geográficas* (relacionada con la variación climática debida a la ubicación de cada punto dentro del macizo montañoso). Para cada cuadrícula se promediaron los datos de sus cuatro nodos. Estas variables son las utilizadas para predecir la presencia/ausencia de cada especie.

Para elaborar el modelo de predicción se recabó información en zonas próximas al incendio con vegetación dominada por especies nativas y una estructura vertical poco alterada. Los lugares escogidos para los muestreos cumplían, además, los criterios de: abarcar el rango de altitud, pendiente y orientación existente en la zona del incendio, y poseer alguno de los tipos de vegetación potencial descritos para la zona por VALLE et al. (2003). Se realizaron recorridos por carriles y pistas forestales seleccionando puntos de muestreo espaciados 200-500 metros. Cada punto fue georreferenciado y se anotó la presencia de especies leñosas, así como las variables

ambientales descritas anteriormente. Se realizaron en total 14 recorridos con un total 212 puntos de muestreo.

Modelos de predicción

Con el fin de obtener el mejor modelo predictivo de la presencia potencial de cada especie en cada cuadrícula del incendio, hemos empleado tres métodos de análisis: diagramas de regresión (CART), modelos lineales generalizados (GLIM) y análisis discriminantes generales (GDA), todos ellos implementados en el programa informático Estadística (STATSOFT INC. 1998) y ampliamente utilizados en estudios similares (REICHAR & HAMILTON, 1997; MILLER & FRANKLIN, 2002).

La información utilizada para elaborar los modelos procede de los muestreos de vegetación potencial en zonas próximas al incendio. La estructura elegida para ajustar el modelo predictivo ha sido la misma en los tres métodos de análisis empleados. La variable dependiente fue la presencia/ausencia de la especie y como variables predictoras se incluyeron: presencia/ausencia de barrancos y ríos, coordenadas N y E, altitud, pendiente, profundidad del suelo, porcentaje de roca, orientación, así como las interacciones entre pares de ellas. Por su naturaleza, el modelo CART no requiere la inclusión de estas interacciones.

Validación y Elección del Modelo

Un modelo puede predecir muy bien la presencia de una especie en las muestras empleadas en su ajuste y, sin embargo, generar una predicción de la distribución que refleje pobremente la realidad. Esto podría suponer, por ejemplo, la plantación de una especie en una gran superficie cuando en realidad solo debería aparecer en unos pocos lugares. Para minimizar este problema es recomendable analizar la fiabilidad predictiva del modelo a través de un procedimiento de validación basado en su capacidad de predicción frente a muestras independientes (validación externa). De los 212 puntos de muestreo de vegetación natural, 162 fueron utilizados para ajustar los modelos, mientras que el resto (50 casos a los que llamaremos "muestra de validación") se utilizó para evaluar su poder predictivo (FIELDING & BELL, 1997). La muestra de validación se seleccionó de forma aleatoria de entre los puntos de muestreo que presentaban condi-

ciones ambientales dentro del rango presente en zona del incendio, sobre la cual se va a realizar la predicción. Este tipo de procedimientos son empleados de forma habitual en aplicaciones predictivas (OLDEN & JACKSON; 2002).

Este procedimiento se aplicó para cada uno de los tres modelos de una especie. Cada modelo ofrece como resultado la probabilidad de aparición de la especie en cada uno de los puntos de la muestra de validación, en los cuales se conoce, además la presencia/ausencia real de la especie. Evidentemente, la probabilidad de aparición predicha para cada punto (y que toma un valor continuo entre 0 y 1) puede ser más o menos aproximada a la realidad. Esto plantea la pregunta: ¿a partir de que valor de probabilidad predicha podemos asumir que la especie debería estar, en realidad, presente? Según se tome un valor u otro para este umbral, se cometerán más o menos errores en la predicción. Una forma de estimar la capacidad predictiva de un modelo, independientemente del umbral elegido, es el uso de las llamadas curvas "Receiver-Operating Characteristic", en adelante ROC. (PEARCE & FERRIER, 2000; THUILLER *et al.*, 2003b). La curva ROC es un método que analiza la relación entre la fracción de falsos positivos (observaciones que siendo negativas en la realidad son predichas como positivas) y la sensibilidad (proporción de observaciones que siendo positivas en la realidad son predichas como tales) para un rango de umbrales. Si las predicciones se realizasen de forma aleatoria, la relación entre sensibilidad y falsos positivos sería una línea de 45°. La curva ROC para un modelo correcto se caracteriza porque maximiza la sensibilidad para valores bajos de falsos positivos. El área entre la línea de 45° y la curva (área bajo la curva, AUC) mide la capacidad del modelo para clasificar correctamente la presencia/ausencia de una especie en un punto dado, independientemente del umbral elegido. Según SWETS (1988) los rangos de precisión del modelo en función de los valores de AUC serían: 0.5-0.6 = malo; 0.6-0.7 = pobre; 0.7-0.8 = satisfactorio; 0.8-0.9 = bueno; y 0.9-1.00 = excelente. Este procedimiento permite comparar la capacidad predictiva de distintos modelos. En nuestro caso, se ha seleccionado para cada especie el modelo que presentó mayor valor de AUC. Cuando el valor de AUC fue muy similar entre modelos, se

ha recurrido como método complementario de elección a la Robustez del modelo dentro del área del incendio: porcentaje de unidades del incendio que presentaban regeneración de una especie y que fueron predichas para la presencia de la misma, ya que la existencia de regeneración de una especie indica la idoneidad de esa unidad.

Una vez elegido el mejor modelo, el siguiente paso es determinar el umbral óptimo para la transformación de los datos de probabilidad a valores de presencia/ausencia. El umbral elegido fue el que maximiza los valores del estadístico Kappa (k): $k = (O - E)/(I - E)$; donde O es la proporción de muestras clasificadas correctamente por el modelo utilizando un umbral dado y E es la proporción de muestras que se clasificarían correctamente si la asignación de presencia/ausencia se realizase al azar. K toma valores entre 0 y 1, teniendo un valor de cero cuando O no es distinto de E , es decir, cuando el modelo no hace una predicción mejor que el puro azar. El valor máximo se alcanza cuando la predicción del modelo es perfecta.

RESULTADOS Y DISCUSIÓN

Para cada uno de los modelos seleccionados se muestran el valor de AUC, así como medidas

de precisión basadas en los resultados de validación externa con 50 casos de vegetación natural (Tabla 1), y los mapas de distribución resultantes de la aplicación del mejor modelo (Figura 1). Las variables que más poder discriminante han tenido sobre *Buxus sempervirens* han sido tanto la pendiente como la cobertura de roca, influyendo también la existencia de un barranco o río próximo, así como las interacciones entre la altitud y la pendiente o la orientación con la pendiente. Para *Pinus nigra* la que más influencia tiene sobre la presencia o no de la especie es la altitud, coincidiendo con su carácter edafoxerófilo que le aporta mayor resistencia a las condiciones de xericidad estival e inviernos fríos. Según la clasificación de SWETS (1988), la precisión de los modelos ha sido buena para *Pinus nigra* y satisfactoria para *Buxus sempervirens*. La proporción de falsos positivos en nuestro caso es del 20-30%, lo que indica una moderada tasa de error. Sin embargo, debido a que una especie no siempre ocupa toda su área potencial, diversos autores (FIELDING & BELL, 1997), estiman el error ocasionado por los falsos positivos como de menor gravedad que el error producido al estimar una especie como ausente cuando realmente está presente (falsos negativos). Cabe, además, destacar los altos valores de sensibilidad conseguidos con ambas especies, que superan claramente el 90%.

	AUC	Umb	CCR	Ss	Sp	PPP	PFP	PFN
<i>Pinus nigra</i>								
GLIM	0.86	0.69	78	76.92	79.17	80	20.83	23.08
GDA	0.87**	0.49	82	92.31	70.83	77.42	29.17	7.69
CART	0.74	0.5	74	76.92	70.83	74.07	29.17	23.08
<i>Buxus sempervirens</i>								
GLIM	0.56	0.13	69.39	66.67	69.57	12.50	3.43	33.33
GDA	0.79*	0.15	72	100	70.21	17.65	29.79	0
CART	0.5	0.5	94	0	100	6	0	100

Tabla 1. Medidas de precisión y error de los tres modelos ensayados para *Buxus sempervirens* y *Pinus nigra*. En negrita se marca el modelo seleccionado para cada especie. La significación estadística de los valores de AUC se realizó mediante análisis de 1000 combinaciones aleatorias de las muestras, y se indica: * $p < 0.05$, ** $p < 0.0001$. Umb: umbral de probabilidad de presencia utilizado en la predicción. Para cada uno de los modelos seleccionados se han escogido cuatro medidas de precisión basadas en los resultados de una validación externa con 50 casos. Porcentaje de casos correctamente clasificados (CCR): porcentaje de observaciones correctamente predichas de presencia/ausencia. Sensitividad (Ss): porcentaje de observaciones que, siendo positivas en la realidad, son predichas como tales, es decir, porcentaje de verdaderos positivos. Especificidad (Sp) porcentaje de observaciones que, siendo negativas en la realidad, son predichas como tales, es decir, es porcentaje de verdaderos negativos. Poder predictivo positivo (PPP): porcentaje de observaciones predichas como positivas siendo realmente positivas. Como medidas que indican posibles errores en el modelo se encuentran: Porcentaje de falsos Positivos (PFP): porcentaje de observaciones que siendo negativas en la realidad, son predichas como positivas = $100 - \text{Especificidad}$. Porcentaje de Falsos Negativos (PFN): porcentaje de observaciones que siendo positivas en la realidad, son predichas como negativas

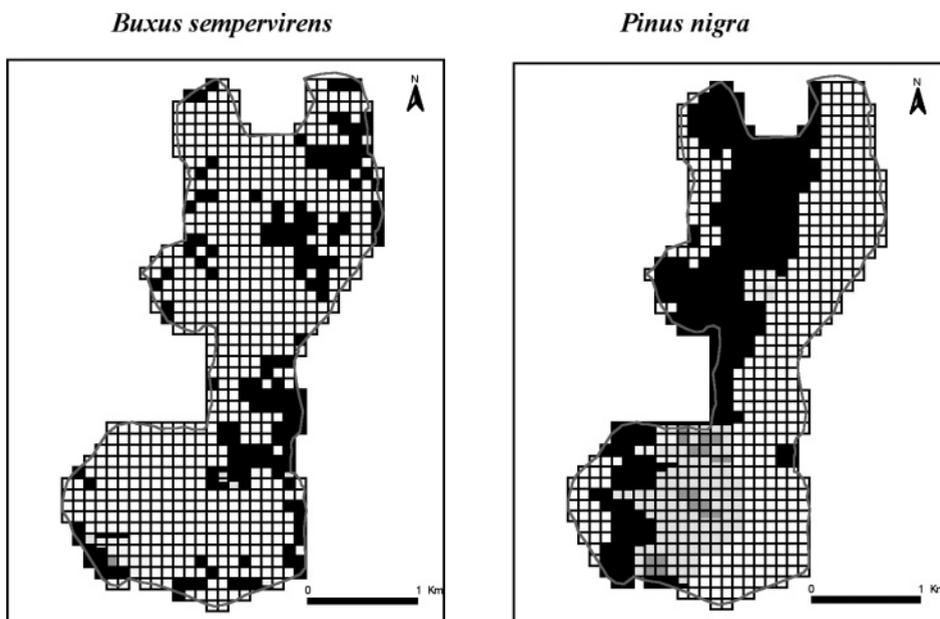


Figura 1. Distribución de *Buxus sempervirens* y *Pinus nigra* en cada unidad espacial explícita de la zona de estudio. Se representa en verde las celdas en las que el modelo ha predicho la existencia de la especie y que en la actualidad tenían rebrote de la misma; en amarillo, las celdas donde existe rebrote sin haber sido predichas por el modelo; y en azul, las celdas predichas como aptas por el modelo para la existencia de esa especie pero donde no existe rebrote en la actualidad

CONCLUSIONES

Los modelos de predicción desarrollados han probado su utilidad como herramienta fundamental para el futuro trabajo de reforestación, ya que sirven para conocer con bastante fiabilidad la distribución de cada especie (en nuestro caso, *Pinus nigra* y *Buxus sempervirens*) en cada unidad explícita del incendio. Su utilización en la asignación de especies a las unidades espaciales explícitas del incendio nos permitirá crear una distribución de especies, que refleje la heterogeneidad natural de la vegetación y que minimice futuros costes tanto económicos como ecológicos.

Agradecimientos

Este trabajo se financió con cargo al Convenio "Asesoramiento y Seguimiento de las actuaciones de restauración del incendio del "Puerto de Las Palomas" entre la Universidad de Jaén y la Empresa de Gestión Medio Ambiental, S.A. (EGMASA).

BIBLIOGRAFÍA

- FIELDING, A.H. & BELL, J.F.; 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38-49.
- FRANKLIN, J.; 1995. Predictive vegetation mapping: Geographic modelling of biospacial pattern in relation to environmental gradients. *Prog. Phys. Geogra.* 19: 474-499.
- LUQUE, P.; 1995. *Mapa de vegetación del Parque Natural de las Sierras de Cazorla, Segura y Las Villas*. Consejería de Medio Ambiente. Junta de Andalucía.
- MILLER, J. & FRANKLIN, J.; 2002. Modelling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecol. Model.* 157: 227-247.
- OLDEN, J.D. & JACKSON, D.A.; 2002. A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biol.* 47: 1976-1995.

- REICHAR, S.H. & HAMILTON, C.W.; 1997. Predicting invasions of woody plants introduced into North America. *Conserv. Biol.* 11: 193-203.
- STATSOFT INC.; 1998. *Statistica for Windows*. StatSoft Inc. Tulsa. OK
- SWETS, K.A.; 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285-1293.
- THULLER, W.; 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biol.* 9: 1353-1362.
- THULLER, W.; ARAUJO, M.B. & LAVOREL, S.; 2003a. Generalised models versus classification tree analysis: a comparative study for predicting spatial distributions of plant species at different scales. *J. Veg. Sci.* 14: 669-680.
- THULLER, W.; VAIDERA, J.; PINO, J.; SABATÉ, S.; LABOREL, S. & GRACIA, C.; 2003b. Large-scale environmental correlatos of forest tree distributions in Catalonia (NE Spain). *Global Ecol. Biog.* 12: 313-325.
- VALLE, F.; ALGARRA, J.A.; ARROJO, E.; ASENSI, A.; CABELLO, J.; CANO, E.; CAÑADAS, E.; CUETO, M.; DANA, E.; DE SIMÓN, E.; DÍEZ, E.; GARCÍA, A.; GIMÉNEZ, E.; GÓMEZ, F.; JIMÉNEZ, M.; LINARES, J.E.; LORITE, J.; MELENDO, M.; MONTOYA, M.; MOTA, J.F.; NAVARRO, F.B.; PÉÑAS, J.; SALAZAR, C. Y TÓRRES, J.A.; 2003. *Mapa de Series de Vegetación de Andalucía. Consejería de Medio Ambiente*. Junta de Andalucía. En: F. Valle (ed.), Editorial Rueda, S.L. Madrid.