

APLICACIÓN DE MODELOS DE DISTRIBUCIÓN DE ESPECIES A LA CONSERVACIÓN DE LA BIODIVERSIDAD EN EL SURESTE DE LA PENÍNSULA IBÉRICA

BLAS BENITO DE PANDO¹ y JULIO PEÑAS DE GILES²

Unidad de Conservación Vegetal, Departamento de Botánica, Facultad de Ciencias,
Universidad de Granada, Avda. Severo Ochoa, C.P. 18071, Granada, España

¹blasbp@ugr.es, ²jgiles@ugr.es

RESUMEN

Aplicando modelos de distribución de especies basados sólo en presencias, desarrollamos una estrategia de modelización con múltiples algoritmos y selección de los modelos óptimos mediante la aplicación de la curva ROC y varias muestras aleatorias. Hemos generado modelos de hábitat de una especie vegetal amenazada, *Linaria nigricans*, y su principal motivo de amenaza, los invernaderos. Combinando la información de los modelos resultantes confeccionamos un modelo de riesgo relativo de extinción de poblaciones de la planta causado por la construcción de invernaderos. Este modelo permite estimar la viabilidad de las poblaciones de la planta y es directamente aplicable al diseño de planes de gestión ambiental y conservación de la biodiversidad.

Palabras clave: SIG, invernaderos, curva ROC, *Linaria nigricans*, modelo de extinción, mapas de idoneidad

APPLYING SPECIES DISTRIBUTION MODELS TO THE BIODIVERSITY CONSERVATION IN THE SOUTHEAST OF THE IBERIAN PENINSULA

ABSTRACT

Applying only-presence based species distribution models, we develop a modelling framework using multiple algorithms, evaluating and selecting the most accurate models using ROC curves and several random samples. We generate models of suitable habitat for the threatened plant species *Linaria nigricans* and potential distribution models for greenhouses. Combining the resulting models we built up a relative extinction risk model, useful to estimate the viability of *L. nigricans*' populations and directly applicable to the design of environmental management plans and the conservation of the biodiversity.

Keywords: GIS, greenhouses, ROC curve, *Linaria nigricans*, extinction mode, suitability maps

1. Introducción

En los últimos años se está produciendo un desarrollo muy importante en las técnicas aplicadas a la generación de modelos de distribución de especies, gracias a la amplia oferta de lenguajes de programación, la amplia difusión de información geoespacial y el desarrollo de las técnicas SIG. Los modelos resultantes de estas técnicas se están mostrando eficaces para establecer planes de conservación (Ferrier, 2002), buscar nuevas localizaciones de especies raras y amenazadas (Guisan *et al.*, 2006), avanzar en la comprensión de los patrones espaciales de la biodiversidad (Graham *et al.*, 2006), ó evaluar el impacto del cambio global sobre la distribución de los organismos (Lawler *et al.*, 2006).

Los modelos de distribución de especies indican la idoneidad del hábitat para el desarrollo de poblaciones de una especie concreta o de una comunidad (Ferrier y Guisan, 2006), calculada a partir de observaciones de campo y una serie de variables ambientales que actúan como predictores. La expresión de esta idoneidad del hábitat puede ser binaria, otorgando valor 1 a áreas idóneas y 0 a áreas no idóneas, o continua, según un rango de valores que clasifican el territorio de menor a mayor grado de idoneidad. En la construcción de estos modelos resultan críticos una serie de factores que afectan en gran medida a la precisión del resultado final:

A) La calidad de los datos de localización de la especie: dos tipos de datos son los utilizados habitualmente para generar modelos de idoneidad del hábitat: registros de presencia y registros de presencia-ausencia. Ambos tipos de registros pueden tener un carácter oportunista, como ocurre con las observaciones de herbario, o proceder de un muestreo especialmente diseñado para generar modelos de distribución (Hirzel y Guisan, 2002). Si bien un registro de presencia es incuestionable (especialmente si trabajamos con especies sésiles, y salvo error en la identificación de la especie), no ocurre lo mismo con los registros de ausencia. La dinámica de una población, la fragmentación del hábitat ó la capacidad dispersiva son factores que pueden llevar a designar como ausencia una localización con características óptimas para la especie, alterando el resultado final y restando significación biológica en la interpretación del modelo (Hirzel *et al.*, 2002). Esta circunstancia es habitual cuando se trata con especies que no están en equilibrio con su medio (invasoras, por ejemplo) u organismos con grandes fluctuaciones interanuales en el tamaño poblacional, que pueden generar ausencias temporales, ó también en organismos con baja detectabilidad, que originan ausencias aparentes.

B) Las variables predictoras seleccionadas: idealmente deberían ser aquellas que se consideren las causantes directas de la distribución de la especie, aunque es habitual que se utilicen únicamente las disponibles, generalmente derivadas de modelos digitales de elevaciones ó interpolación espacial a partir de datos de estaciones climáticas. Las variables predictoras pueden clasificarse como: *gradientes de recursos* directamente consumidos (nutrientes, agua, luz, etc.); *gradientes directos*, que son parámetros con importancia fisiológica (temperatura, pH, etc.) y *gradientes indirectos*, relativos a las características físicas del territorio (orientación, elevación, pendiente, geología, etc.), pero que muestran una buena correlación con los patrones de distribución de las especies porque habitualmente reemplazan distintas combinaciones de gradientes de recursos y gradientes directos (Guisan *et al.*, 1999; Guisan y Zimmermann, 2000). La colinealidad entre

variables también es un factor a tener en cuenta, pues añade información redundante al algoritmo de modelización y puede generar resultados de difícil interpretación por sobreajuste del modelo.

C) El algoritmo ó método estadístico seleccionado: pueden utilizarse métodos basados en registros de presencia-ausencia, como GLM (generalized linear models), GAM (generalized additive models) (Guisan *et al.*, 2002), y redes neuronales (Manel *et al.*, 1999), ó algoritmos basados únicamente en presencias como Bioclim, Domain, ENFA, GARP y MaxEnt, que se comentarán más adelante. Los modelos basados únicamente en presencias representan generalmente la distribución espacial del nicho ecológico fundamental de la especie, mientras que los basados en presencia-ausencia indican de modo más aproximado la distribución del nicho ecológico efectivo (Zaniewski *et al.*, 2002).

Otra cuestión crítica en la confección de modelos de distribución es su evaluación estadística. Actualmente, la medida de la precisión de los modelos de distribución es un campo en desarrollo, en el que la discusión está abierta (Fielding y Bell, 1997; Anderson *et al.*, 2003; Segurado y Araújo, 2004; Ottaviani *et al.*, 2004; Elith *et al.*, 2006; Hirzel *et al.*, 2006). Aún no existe un consenso sobre cual es la mejor técnica para cada tipo de modelo, aunque el área bajo la curva ROC (descrita detalladamente en el apartado 2.7), generalmente aplicada a modelos de presencia-ausencia, se perfila como una de las candidatas con mayor aceptación. En este trabajo se propone su aplicación a modelos basados en presencias aplicando modificaciones propuestas por otros autores (Phillips *et al.*, 2006) y añadiendo otras que permiten una comparación estadística robusta entre modelos.

En este trabajo proponemos una técnica de modelización múltiple, con selección y combinación de modelos basados en presencias, para generar mapas de riesgo de extinción de especies amenazadas, aplicándola a la conservación de la biodiversidad en el sureste de la Península Ibérica, concretamente en la provincia de Almería.

En esta provincia, la expansión de los invernaderos presenta diversas consecuencias ambientales negativas, entre las que destaca la destrucción directa de pastizales mediterráneos de alto valor ecológico, incluidos en la Directiva Hábitat 92/43/CEE. Como especie focal (Lambeck, 1997) marcadora de estos pastizales hemos seleccionado la planta *Linaria nigricans* Lange, endémica de la provincia de Almería. Está catalogada como En Peligro según la "Lista Roja de la Flora Vasculare de Andalucía" (Cabezudo *et al.*, 2005), y tanto la planta como su hábitat se encuentran protegidos por la ley de 8/2003, de 28 de Octubre de Flora y Fauna Andaluza. Cierta laxitud en el cumplimiento de esta ley está situando a la especie y su hábitat en una situación comprometida de cara al futuro.

Para desarrollar un modelo de riesgo de destrucción del hábitat de la planta aplicable a planes de conservación, hemos construido, editado, evaluado y seleccionado modelos de distribución de *L. nigricans* y de invernaderos utilizando distintos algoritmos de modelización de hábitat basados en registros de presencia. Los modelos resultantes se han combinado para estimar el riesgo de destrucción del hábitat de la planta debido a la construcción de invernaderos.

2. Material y métodos

2.1. Área de trabajo

El área de trabajo comprende gran parte de la provincia de Almería ([figura 1](#)), enmarcando el área costera desde la provincia de Granada hasta la Región de Murcia. El límite norte se sitúa en la coordenada UTM 4157730 N y el oeste en UTM 474810, ambas referidas al uso 30 y utilizando el datum ED50. La resolución de trabajo es 20 x 20 metros, con un total de 29 millones de celdas analizadas (11600 Km²).

2.2. Registros de presencia

Con el objetivo de disponer de información adecuada para generar modelos de distribución del hábitat de la planta y de la superficie idónea para invernaderos, se han dispuesto dos conjuntos de datos ([figura 2](#)): 500 registros de presencia de *L. nigricans* recogidos en las poblaciones de la especie durante los años 2004 y 2005, y 500 registros de presencia de invernaderos tomados sobre una imagen satelital Landsat (año 2000). Cada conjunto de presencias se ha dividido aleatoriamente en dos grupos de 300 y 200 registros, utilizados respectivamente para entrenar los algoritmos y evaluar los modelos resultantes (Fielding y Bell, 1997).

2.3 Base de datos medioambiental

Como base de trabajo cartográfica se ha utilizado un modelo digital de elevaciones de Andalucía de 20 metros de resolución horizontal. Se han confeccionado capas topográficas (altitud, orientación, pendiente), climáticas (temperatura, humedad, precipitación y radiación solar), de distancias (a poblaciones, a vías de comunicación, al mar, a embalses, a desaladoras), y se han recortado y editado capas temáticas (espacios protegidos, vías de comunicación y usos de suelo). La cartografía de ocupación de invernaderos para el año 2000 se ha generado mediante un análisis de coberturas con redes neuronales mediante el módulo MLP (Multi-Layer Perceptron) sobre una imagen Landsat (Eastman, 2003).

La radiación solar potencial se ha calculado utilizando el programa SolarAnalyst (www.fs.fed.us/informs/download.php), que utiliza parámetros topográficos, geográficos y atmosféricos, para generar mapas de radiación solar potencial máxima, mínima y media para distintos periodos del año.

Las capas de variables climáticas se han calculado a partir de los registros de la Red de Información Agroclimática Andaluza (RIAC). Se han utilizado técnicas de regresión múltiple (Ninyerola *et al.*, 2000) para obtener las ecuaciones que describen el comportamiento de las variables climáticas en función de la altura, la distancia al mar, la radiación solar, la latitud y la longitud. A partir de estas ecuaciones se han construido los mapas de temperatura, humedad y

precipitación (valores medios, mínimos, máximos y rango) que representan las características climáticas medias del periodo 1999-2005.

2.4. Modelos de distribución de especies

En varios estudios se ha constatado 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). En general los resultados no son concluyentes, funcionando de modo diferente los algoritmos según la especie, la cantidad de registros de presencia, o el número de variables predictoras. Por este motivo proponemos una metodología basada en la generación de múltiples modelos calculados con distintos algoritmos a partir de los mismos datos para posteriormente seleccionar el modelo más óptimo. Teniendo en cuenta la naturaleza de los datos recopilados, hemos seleccionado una serie de algoritmos basados únicamente en presencias para generar los modelos de distribución del hábitat de *L. nigricans* y de territorio idóneo para invernaderos.

Bioclim (Busby 1991): Este algoritmo, implementado en el programa de gestión de recursos florísticos DIVA-GIS (www.diva-gis.org) (Hijmans *et al.*, 2004), genera para la especie un rango ecológico de n dimensiones, siendo n el número de variables predictoras, mediante un análisis de la distribución de los registros de presencia sobre cada variable ambiental. El hiper-rectángulo resultante está caracterizado por una tolerancia mínima y máxima que determinan los límites de idoneidad. El modelo de hábitat se genera otorgando a cada celda un valor de distancia ecológica relativo a la posición que ocupa dentro del rango multidimensional de la especie. Este algoritmo no admite para el análisis capas de información cualitativa como usos del suelo y similares.

Domain (Carpenter *et al.*, 1993): Es un método de distancias ecológicas implementado en DIVA-GIS. Utiliza la métrica de Gower, una medida de distancia que estandariza las variables según su rango en todos los puntos de presencia para igualar su contribución al modelo. La distancia entre una presencia A y una celda B para una variable k se calcula como la diferencia absoluta entre A y B dividida por el rango de k para todos los puntos de presencia. La distancia de Gower (G) es la media de las distancias para todas las variables climáticas, y el estadístico de similitud Domain se calcula como $100(1-G)$. El valor máximo es 100, que se da en todas las celdas con presencia del organismo.

ENFA (Ecological Niche Factor Analysis): El algoritmo está disponible en el programa Biomapper (www.unil.ch/biomapper) (Hirzel, H. *et al.* 2000). Realiza un análisis de componentes principales sobre las variables ambientales que ofrece como resultado una serie de factores no correlacionados entre sí que tienen significado biológico: el factor de *marginalidad*, que describe la distancia existente entre el óptimo ambiental para la especie y el valor medio de las variables ambientales para todo el área de estudio, y el factor de *tolerancia*, que indica el grado de especialización de la especie con respecto al rango ambiental disponible. Para generar el mapa de idoneidad dispone, entre otros (no utilizados en este trabajo), de los algoritmos: *median* (ENFA M) y *geometric mean* (ENFA GM). El valor de idoneidad de

cada celda se obtiene comparando sus valores de marginalidad y especialización con la media (según el algoritmo *geometric mean*) ó la mediana (según el algoritmo *median*) de cada factor calculadas para todo el área de estudio. El resultado se normaliza, expresándose en valores de 0 a 100. ENFA no admite la inclusión de variables cualitativas en el análisis.

GARP (Genetic Algorithm for Rule-set Prediction) (Stockwell y Peters, 1999): El programa DesktopGarp (www.lifemapper.org/desktopgarp/) contiene las herramientas necesarias para generar modelos de distribución utilizando este algoritmo. Es un sistema heurístico basado en inteligencia artificial que busca correlaciones no aleatorias entre los puntos de presencia del organismo y las variables ambientales. Funciona iterativamente, aplicando distintas reglas de modelado (*atomic*, *logistic regresión*, *range rules* y *negated range*), generando y evaluando en cada bucle de cálculo un modelo de distribución binario (1 = presencia, 0 = ausencia) distinto. Los modelos binarios resultantes son seleccionados (ver Anderson *et al.*, 2003) y sumados algebraicamente, proporcionando un mapa final de idoneidad del hábitat con un rango de valores continuo. Los resultados no son determinísticos, por lo que un mismo conjunto de datos origina mapas de idoneidad distintos. Puede utilizar información cualitativa, pero se desconoce el uso que hace de ella.

MaxEnt (Phillips *et al.*, 2006) (www.cs.princeton.edu/~schapire/maxent/): Es un método de inteligencia artificial que aplica el principio de máxima entropía para calcular la distribución geográfica más probable para una especie. MaxEnt estima la probabilidad de ocurrencia de la especie buscando la distribución de máxima entropía (lo más uniforme posible) sujeta a la condición de que el valor esperado de cada variable ambiental según esta distribución coincide con su media empírica. El resultado del modelo expresa el valor de idoneidad del hábitat para la especie como una función de las variables ambientales. Un valor alto de la función de distribución en una celda determinada indica que ésta presenta condiciones muy favorables para la presencia de la especie. MaxEnt puede utilizar variables cualitativas, otorgando a cada valor de la variable un peso relativo al número total de puntos de presencia que contiene. El programa proporciona las curvas de respuesta de la especie ante las distintas variables ambientales y estima la importancia de cada variable en la distribución de la especie.

2.5. Calibración de algoritmos

Puede llamarse calibración al ajuste de los parámetros y constantes del modelo para incrementar la coherencia entre los datos de entrada y los resultados (Rykiel, 1996). En este sentido, es posible influir en la calibración de los distintos modelos desde los datos de entrada, preparando convenientemente la muestra de localizaciones de la especie y seleccionando cuidadosamente las variables ambientales, y a través de los parámetros internos de cada algoritmo.

Con las 26 variables ambientales disponibles se realizó un proceso de selección analizando la matriz de correlación entre variables obtenida con Biomapper. Del árbol de correlaciones resultante se eliminaron sucesivamente variables hasta conseguir una correlación máxima de 0,7. Con este proceso se disminuye la multicolinealidad entre variables, reduciéndose la información

redundante proporcionada a los algoritmos. Para todos los algoritmos se utilizaron las capas de información topográfica y climática para calibrar los modelos de hábitat de *L. nigricans*, añadiendo las de distancias a infraestructuras para los modelos de invernaderos. Los parámetros de calibración introducidos en los distintos algoritmos fueron idénticos en ambos casos tratando de generar modelos comparables entre sí.

Bioclim y Domain carecen de parámetros internos que se puedan manipular, por lo que los resultados dependen enteramente de los datos de entrada.

ENFA permite seleccionar el número de factores con los que se va a computar el modelo y modificar el peso específico que tendrá cada uno de estos factores en el resultado final. Además permite la elección de distintos algoritmos como ya se ha comentado anteriormente. Para modelizar la distribución de *L. nigricans* e invernaderos utilizamos en ambos casos 1 factor de marginalidad y 6 de especialización, manteniendo el peso específico calculado por el programa.

Para configurar MaxEnt se introdujo 1000 como número máximo de iteraciones, estableciendo el límite de convergencia en 0.00001 y el valor de regularización en 0.0001. Estos valores, según el autor (Phillips *et al.*, 2006), son adecuados para garantizar la convergencia del algoritmo.

En GARP, el máximo número de iteraciones se estableció en 2000, con un límite de convergencia de 0.001, permitiendo el uso de todas las reglas para el cálculo de los modelos y utilizando todas las variables a la vez. Se utilizó el 95% de los puntos para generar los modelos y el 5% para testarlos. Se calcularon 150 modelos binarios, seleccionando y sumando entre sí los 100 mejores.

2.6. Edición de modelos

Para incorporar la información relativa al uso del suelo, los modelos de hábitat de *L. nigricans* fueron filtrados, reclasificando a cero todas las áreas correspondientes a usos del suelo incompatibles con el desarrollo de poblaciones de la planta, determinándose como incompatibles aquellos usos sin presencia confirmada de la planta. Los modelos de invernaderos se filtraron igualmente reclasificando a cero las áreas correspondientes a edificaciones, canteras y otros usos que impiden la construcción de estas instalaciones agrícolas.

2.7. Evaluación de modelos

Los modelos de distribución basados en presencia-ausencia habitualmente se evalúan utilizando una muestra de evaluación independiente de la utilizada para entrenar los modelos compuesta por registros de presencia-ausencia independientes de los utilizados para el entrenamiento de los algoritmos. A partir de esta muestra de evaluación pueden calcularse: a) los casos correctamente clasificados, que son verdaderos positivos cuando el modelo predice presencia y ésta es confirmada por las presencias de la muestra de evaluación y verdaderos negativos cuando

el modelo predice ausencia y ésta es confirmada por las ausencias de la muestra de evaluación. b) los casos erróneamente clasificados, que son falsos positivos cuando el modelo predice presencia y la muestra indica ausencia (error de comisión), y falsos negativos cuando el modelo predice ausencia y la muestra indica presencia (error de omisión). Estos datos se tabulan en una matriz de confusión a partir de la cual se pueden calcular distintos índices de evaluación (para una revisión exhaustiva sobre técnicas de evaluación de modelos de distribución, ver Fielding y Bell, 1997).

El error de comisión y los falsos positivos no puede calcularse para modelos basados en presencia, excepto si realizamos un cambio de concepto y sustituimos los puntos de ausencia por puntos generados aleatoriamente, en lo que algunos autores denominan pseudo-ausencias (Zaniewski *et al.*, 2002; Engler *et al.*, 2004). De este modo obtenemos una matriz de confusión modificada ([tabla 1](#)), a partir de la cual puede calcularse la capacidad del modelo para discriminar entre registros de presencia y puntos generados aleatoriamente.

Para comparar la capacidad de discriminación de los distintos modelos hemos seleccionado la técnica del área bajo la curva ROC (Receiver Operating Characteristic) (Hanley y McNeil, 1982), procedente del procesamiento de señales de radar, que ha sido aplicada con éxito a la evaluación de modelos de distribución basados en algoritmos de presencia-ausencia (Fielding y Bell 1997, Segurado y Araújo 2004; Muñoz y Felicísimo 2004) y en algoritmos de solo presencia (Phillips *et al.*, 2006).

La construcción de la curva ROC es un proceso iterativo aplicado a cada umbral de idoneidad del modelo: el modelo de distribución se reclasifica en dos categorías a partir del umbral, para obtener un modelo binario, en el que los valores del modelo original inferiores al umbral indican ausencia y los valores iguales o superiores al umbral indican presencia. Sobre este modelo binario se superponen los puntos de presencia y los puntos aleatorios, y se calculan los valores de la matriz de confusión. A partir de estos valores se calcula la fracción de verdaderos positivos (sensibilidad = $a / a + c$) y la fracción de aleatorios positivos (especificidad = $d / b + d$). En el gráfico de la curva ROC se dibuja el valor de sensibilidad en el eje x y el valor $1 -$ especificidad en el eje y . Repitiendo el proceso para todos los umbrales de idoneidad del modelo y conectando los puntos mediante líneas rectas se obtiene la función ROC. El área bajo la curva ROC (AUC) indica, para un punto de presencia y uno aleatorio seleccionados al azar, la probabilidad de que el valor de idoneidad previsto por el modelo para el punto de presencia sea mayor que el previsto para el punto aleatorio. 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, y cercanos a 0.5 cuando el ajuste no es mejor que el obtenido por azar.

El procedimiento se encuentra automatizado en el programa Simtest (cortesía de Nikaus E. Zimmermann, disponible en www.wsl.ch/staff/niklaus.zimmermann/progs.html), una rutina de Fortran que calcula la mayoría de los índices de evaluación habitualmente utilizados.

Los modelos fueron evaluados mediante Simtest comparando la muestra de presencias con 30 muestras distintas de 200 puntos aleatorios (ejemplo de muestra aleatoria en la [figura 2](#)), obteniéndose para cada modelo un valor de AUC medio (AUC_m) y una desviación estándar característicos, que determinan su capacidad de discriminación media ante distintos casos de puntos

aleatorios. Los resultados de la evaluación de los distintos modelos fueron comparados mediante la prueba de rangos de Wilcoxon, seleccionándose como mejores modelos para el hábitat de la planta e invernaderos aquellos con mayor AUC_m y menor desviación estándar.

2.8. Modelo de riesgo de extinción

Para analizar la preferencia de *L. nigricans* e invernaderos por áreas con valores de idoneidad altos según sus modelos de distribución, generamos para ambos las curvas de saturación del hábitat, que resultan de calcular el porcentaje de puntos de presencia de la muestra de evaluación situados en cada valor del rango de idoneidad del hábitat. Las curvas de saturación indican el valor de idoneidad mínimo para cada modelo, a partir del cual resulta improbable la presencia de *L. nigricans* o de invernaderos. Los valores inferiores a estos umbrales fueron reclasificados a cero, como áreas no idóneas para el desarrollo de la planta y para la construcción de invernaderos respectivamente.

Existiendo en los modelos una relación positiva entre porcentaje de presencias e idoneidad, cabe pensar que cuanto mayores sean los valores de idoneidad para invernaderos y *L. nigricans* en un mismo punto del espacio, mayores serán las probabilidades de que una porción de hábitat adecuado para la planta sea destruida por la construcción de un invernadero. Siguiendo este razonamiento, confeccionamos el modelo de riesgo multiplicando los modelos de distribución de invernaderos y *L. nigricans* entre sí, escalando el resultado a valores de 0 a 100 para obtener una escala relativa de riesgo de ocupación. Eliminando de este modelo las áreas en las que ya existen invernaderos, obtenemos la estima del riesgo de desaparición del hábitat de la planta en las áreas en las que aún no se han construido invernaderos.

Para analizar el riesgo relativo de extinción de las poblaciones confirmadas de *L. nigricans*, los 500 registros de presencia disponibles fueron puestos sobre el modelo de riesgo calculando el porcentaje de puntos situados en cada nivel de riesgo.

3. Resultados

Del proceso de selección de variables resultan 12 variables ambientales y 4 de distancias a infraestructuras, mostradas en la [tabla 2](#).

Los resultados gráficos de los distintos algoritmos quedan representados en la [figura 3](#). Los resultados de la evaluación de modelos ([figura 4](#)) indican que MaxEnt ($AUC_m = 0.9466 \pm 0,0064$) es el modelo que mejor se ajusta a la distribución de invernaderos ([figura 5](#)). Respecto a la distribución de *L. nigricans*, es el modelo Domain ([figura 6](#)) ($AUC_m = 0.9961 \pm 0,0030$) el que presenta un mejor ajuste con los datos de evaluación.

Las curvas de saturación ([figura 7](#)) muestran las preferencias de invernaderos y *L. nigricans* por áreas con niveles altos de idoneidad en sus respectivos modelos. Los umbrales

mínimos de idoneidad señalados por las curvas fueron 91 para *L. nigricans* y 81 para invernaderos. El modelo de riesgo de extinción de poblaciones de *L. nigricans* se representa en la [figura 8](#).

4. Discusión

4.1. Modelos de distribución de especies

Son muchos los trabajos en los que se comparan múltiples algoritmos de modelización con el objetivo de determinar cuales son las técnicas más precisas (Zaniewski *et al.*, 2002; Segurado y Araújo, 2004; Elith *et al.*, 2006; Phillips *et al.*, 2006). En general, aunque los resultados apuntan a los métodos de presencia-ausencia como los de mayor precisión, es habitual que en los muestreos de especies raras ó amenazadas no se registren las ausencias, por lo que es importante trabajar en el desarrollo de entornos de modelización basados en algoritmos de solo presencia. En este sentido, proponemos una metodología de modelización múltiple y selección del resultado más óptimo aplicable a este tipo de algoritmos.

Existen grandes diferencias entre los distintos algoritmos en cuanto al coste computacional: Bioclim, Domain y MaxEnt son los algoritmos más rápidos, pudiendo completar los modelos en menos de una hora. Para ENFA, el proceso de análisis de los datos y la generación del modelo supone entre 2 y 3 horas de trabajo, mientras que GARP exige días ó semanas dependiendo de la cantidad de modelos binarios que se desee generar.

Los resultados de la evaluación sugieren que el modelo de distribución de *L. nigricans* es muy preciso. La curva de saturación indica que la planta tiene requerimientos estrictos en cuanto a idoneidad para desarrollar poblaciones viables, concentrándose los registros en las áreas con idoneidad comprendida entre 98 y 100.

No hemos podido encontrar en la literatura una aplicación de los modelos de distribución a una entidad no biológica como son los invernaderos, pero creemos que las particulares características de estas instalaciones (requerimientos de insolación, temperatura, humedad, recursos hídricos, etc.) y el amplio grado de expansión que han alcanzado en la provincia de Almería los convierten en sujetos adecuados para aplicarles las técnicas de modelización comentadas. Han sido tratados como una especie invasora, por lo que su modelo de distribución representa el área potencial que podrían ocupar si continúa su expansión. El análisis de la curva de saturación del modelo ([figura 7](#)) muestra cómo los registros de presencia están repartidos a lo largo del rango de idoneidad a partir del valor 81, y están especialmente concentrados en áreas con idoneidad situada entre 95 y 100, con el máximo de puntos de presencia sobre el valor 98. Estos datos indican que los invernaderos se construyen preferentemente en localizaciones con una determinada combinación de condicionantes ambientales. Los resultados de la evaluación indican que hemos obtenido un modelo de ocupación potencial de invernaderos muy preciso, demostrando la utilidad de los modelos de distribución de especies en su aplicación a entidades no biológicas, cuando las variables que determinan su distribución son conocidas y se pueden representar espacialmente.

Aunque la evaluación de modelos de solo presencia utilizando puntos aleatorios en lugar de ausencias no es nueva (Zaniewski *et al.*, 2002; Phillips *et al.*, 2006), no hemos encontrado una aplicación similar a la que proponemos de la curva ROC, utilizando múltiples muestras de puntos aleatorios. Creemos que este método es más preciso cuanto mayor sea el número de muestras aleatorias utilizado y podría estar influenciado por el tamaño de la muestra de evaluación, por lo que resulta necesario investigar a fondo las posibilidades de esta técnica. Sin embargo, la evaluación de múltiples modelos con varias muestras aleatorias utilizando Simtest es un proceso largo y repetitivo, y se impone la necesidad de automatizar el procedimiento de evaluación y selección de modelos para optimizar los tiempos de trabajo en las distintas fases de la metodología.

4.2. Conservación de *Linaria nigricans*

El modelo de riesgo señala dos áreas sensibles por sus valores de riesgo: Los campos de Níjar y las áreas cercanas a las Salinas de Cabo de Gata. La primera es un área en la que existe un proceso activo de fragmentación del hábitat de *L. nigricans*, y según el modelo, de continuar la tendencia actual, toda la población de esta localidad podría desaparecer. Las poblaciones de las Salinas de Cabo de Gata se encuentran dentro del Parque Natural de Cabo de Gata-Níjar, por lo que no está previsto que se cumplan las predicciones del modelo en este área concreta, salvo cambios en el plan de ordenación del parque o la construcción de invernaderos ilegales.

El análisis de riesgo de los 500 registros de presencia de *L. nigricans* situados sobre el modelo de riesgo revela que un 40 por ciento de las poblaciones conocidas se encuentran el valor 80 de riesgo relativo de extinción, situándose otro 30 por ciento en niveles de riesgo bajos (entre 23 y 35), mientras que el 30 por ciento restante no se encuentra bajo riesgo directo de extinción por la construcción de invernaderos ([figura 9](#)). La validez del modelo de riesgo está condicionada a la continuación de la expansión de los invernaderos. A nuestro entender, las infraestructuras hídricas proyectadas para la región (plantas desaladoras y canales de trasvase) están impulsando la construcción de nuevos invernaderos. Por este motivo pensamos que el modelo propuesto puede ayudar a enfocar la atención en las áreas amenazadas de extinción, sirviendo de guía para el diseño de planes de conservación de la biodiversidad y gestión medioambiental.

5. Conclusiones

Las técnicas de modelización múltiple mediante algoritmos de solo presencia, selección y evaluación de los modelos más óptimos permiten obtener la distribución de la entidad modelada con un alto grado de precisión.

Generar conjuntos de muestras aleatorias aplicados a la evaluación de modelos mediante la curva ROC permite testar la capacidad discriminadora de modelos basados en presencia y establecer diferencias entre los resultados de distintos algoritmos.

En la evaluación de nuestros modelos, los algoritmos MaxEnt y Domain se mostraron superiores en cuanto a capacidad de discriminación al resto de los algoritmos en competencia,

Benito de Pando, B. y Peñas de Giles, J. (2007): "Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica", *GeoFocus (Artículos)*, nº 7, p. 100-119, ISSN: 1578-5157

demostrando por una parte la robustez de un sistema de modelización clásico (Domain) y la capacidad de predicción de las técnicas emergentes basadas en inteligencia artificial (MaxEnt).

A partir de modelos predictivos es posible confeccionar mapas de riesgos que pueden favorecer las labores de identificación de amenazas y conservación de especies de interés.

La propagación de invernaderos hacia el interior de la provincia de Almería está poniendo en peligro poblaciones de *L. nigricans* en los Campos de Níjar, afectando previsiblemente a proporciones significativas del hábitat de esta planta. Es necesario implantar planes de protección adecuados para garantizar la conservación de este valor natural.

Agradecimientos

Queremos agradecer la financiación de esta investigación a la Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía (proyecto de excelencia "Conservación de flora endémica y amenazada de ecosistemas frágiles: Zonas áridas y altas montañas de Andalucía" RNM-1067), y al Ministerio de Ciencia y Tecnología del Gobierno de España (proyecto I+D "Desarrollo de herramientas aplicables a la definición de criterios y diseño de estrategias de conservación de endemoflora del SE árido Ibérico" REN2003-09427-CO2). Agradecemos igualmente a N. Donadio, J. Cabello, C. Oyonarte, M. Piquer, D. Alcaraz, H. Schwarzeger, V. Vargas y J.L. Caparrós su ayuda en los trabajos de campo.

Referencias bibliográficas

- Anderson, R. P., Lew, D. y Peterson, A. T. (2003): "Evaluating predictive models of species distributions: criteria for selecting optimal models", *Ecological Modelling*, 162, pp. 211-232.
- Busby, J.R. (1991): "BIOCLIM – a bioclimate analysis and prediction system", en Margules, C.R., y Austin, M.P., (Eds.), *Nature conservation: cost effective biological surveys and data analysis*. CSIRO, pp- 64-68.
- Cabezudo, B. et al. (2005): *Lista Roja de la Flora Vasculare de Andalucía*. Sevilla, Consejería de Medio Ambiente de la Junta de Andalucía.
- Carpenter, G., Gillson, A. N. y Winter, J. (1993): "DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals", *Biodiversity and Conservation*, 2, pp. 667-680.
- Eastman, J. R. (2003): *Idrisi Kilimanjaro. Guía para SIG y Procesamiento de Imágenes*, Worcester, Clark University.
- Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T., Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. y Zimmermann, N. E. . (2006): "Novel methods improve prediction of species' distributions from occurrence data", *Ecography*, 29, pp. 129-151.

- Engler, R., Guisan, A. y Rechsteiner, L. (2004): "An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data", *Journal of Applied Ecology*, 41, pp. 263-274.
- Ferrier, S. (2002): "Mapping spatial pattern in biodiversity for regional conservation planning: where to from here?", *Systematic Biology*, 51, pp. 331-363.
- Ferrier, S., Guisan, A. (2006): "Spatial modelling of biodiversity at the community level", *Journal of Applied Ecology*, 43, pp. 393-404.
- Fielding, A. H. y Bell, J. F. (1997): "A review of methods for the assessment of prediction errors in conservation presence/absence models", *Environmental Conservation*, 24, pp. 38-49.
- Graham, C.H., Moritz, C., y Williams, S.E. (2006): "Habitat history improves prediction of biodiversity in a rainforest fauna", *Proceedings of the National Academy of Science of USA*. 103, 632-636.
- Guisan, A. y Zimmermann, N.E. (2000): "Predictive habitat distribution models in ecology", *Ecological Modelling*, 135, 147-186.
- Guisan, A., Weiss, S.B., Weiss, A.D. (1999): "GLM versus CCA: Spatial modelling of plant species distributions", *Plant Ecology*, 143, 107-122.
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., y Zimmermann, N.E. (2006): "Using niche-based models to improve the sampling of rare species", *Conservation Biology*, 20, 2, pp- 501-511.
- Guisan, A., Edwards, T.C. y Hastie, T. (2002): "Generalized linear and generalized additive models in studies of species distributions: setting the scene", *Ecological Modelling*, 157, 89-100.
- Hanley, J. A., McNeil, B. J. (1982): "The meaning and use of the area under a receiver operating characteristic (ROC) curve", *Radiology*, 143, pp. 29-36.
- Hijmans, R.J., et al., (2004): "DIVA-GIS. "A geographic information system for the analysis of biodiversity data", www.diva-gis.org.
- Hirzel, A.H., Lay, G.L., Helfer, V., Randin, C. y Guisan, A. (2006): "Evaluating the ability of habitat suitability models to predict species presences", *Ecological Modelling*, 199, 142-152.
- 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, pp. 2027-2036.
- Hirzel, A. H., Guisan, A. (2002): "Which is the optimal sampling strategy for habitat suitability modelling", *Ecological Modelling*, 157, pp. 331-341.
- Hirzel, A. H., Hausser, J., y Perrin, N. (2000). *Biomapper 2.0*, Laboratory for Conservation Biology, University of Lausanne.
- Lambeck, R. J. (1997): "Focal species: a multi-species umbrella for nature conservation", *Conservation Biology*, 11, 849-856.
- Lawler, J.J., White, D., Neilson, R.P. y Blaustein, A.R. (2006): "Predicting climate-induced range shifts: model differences and model reliability", *Global Change Biology*, 12, pp. 1568-1584.
- Manel, S., Dias, J.M. y Ormerod, S.J., (1999): "Comparing discriminant analysis, neural networks and logistic regression for predicting species' distributions: a case study with Himalayan river bird", *Ecological Modelling*, 120, pp. 337-347.
- Muñoz, J. y Felicísimo, A.M. (2004): "Comparison of statistical methods commonly used in predictive modelling", *Journal of Vegetation Science*, 15, pp. 285-292.
- Ninyerola, M., Pons, X. y Roure, J. M. (2000): "A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques", *International Journal of Climatology*, 20, pp. 1823-1841.

Benito de Pando, B. y Peñas de Giles, J. (2007): "Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica", *GeoFocus (Artículos)*, nº 7, p. 100-119, ISSN: 1578-5157

- Ottaviani, D., Lasinio, G.J. y Boitani, L. (2004): "Two statistical methods to validate habitat suitability models using presence-only data", *Ecological Modelling*, 179, pp. 417-443.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006): "A maximum entropy modelling of species geographic distributions", *Ecological Modelling*, 190, pp. 231-259.
- Rykiel, E.J., Jr., (1996): "Testing ecological models: the meaning of validation", *Ecological Modelling*, 90, pp. 229-244.
- Segurado, P. y Araújo, M.B. (2004): "An evaluation of methods for modelling species distributions", *Journal of Biogeography*, 31, pp. 1555-1568.
- Stockwell, D. y Peters, D. (1999): "The GARP modelling system: Problems and solutions to automated spatial prediction", *International Journal of Geographic Information Systems*, 13, pp. 143-158.
- Thuiller, W. (2003): "BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change", *Global Change Biology*, 9, pp. 1353-1362.
- Zaniewski, A.E., Lehmann, A., Overton, J.McC. (2002): "Predicting species spatial distributions using presence only data: a case study of native New Zealand ferns", *Ecological Modelling*, 157, pp. 261-280.

TABLAS

Tabla 1. Matriz de confusión modificada

		MUESTRA DE EVALUACIÓN	
		PRESENCIA	ALEATORIO
MODELO	PRESENCIA	a (verdadero positivo)	b (aleatorio negativo)
	AUSENCIA	c (falso negativo)	d (aleatorio positivo)

Fuente: adaptada y modificada de Fielding y Bell (1998).

Tabla 2. Listado de variables predictoras

Tipo	Variable
Topográfico	Elevación
	Pendiente
	Radiación solar directa máxima
	Radiación solar directa media
	Rango de radiación solar directa
Climático	Precipitación media anual
	Temperatura media máxima
	Temperatura mínima absoluta
	Rango medio de temperatura
	Humedad media máxima
	Humedad media mínima
	Rango medio de humedad
Distancias	Distancia a desaladoras
	Distancia a embalses
	Distancia a poblaciones
	Distancia a vías de comunicación

Fuente: Elaboración propia.

FIGURAS

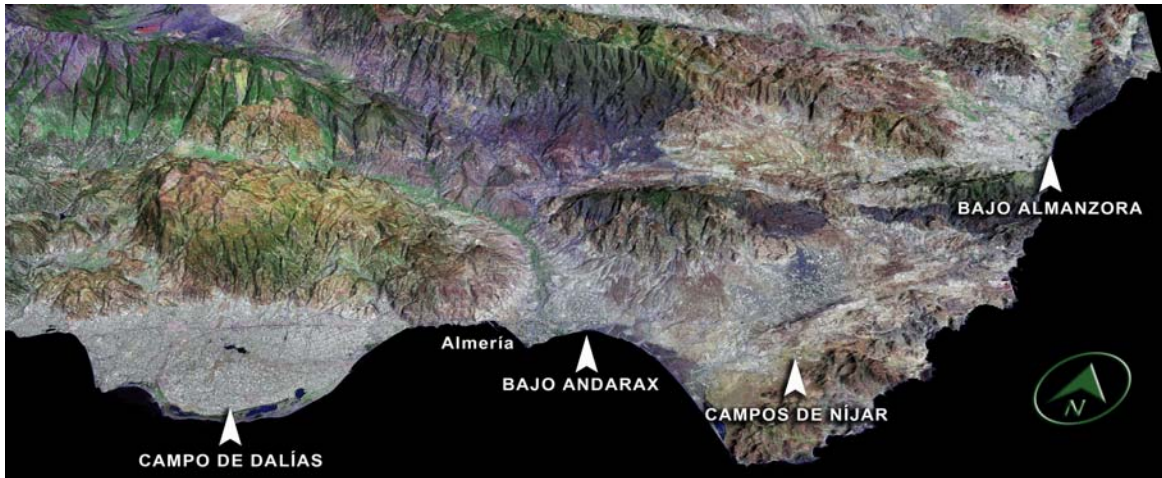


Figura 1. Área de estudio.

Fuente: Elaboración propia, a partir de imagen Landsat y modelo de elevaciones.

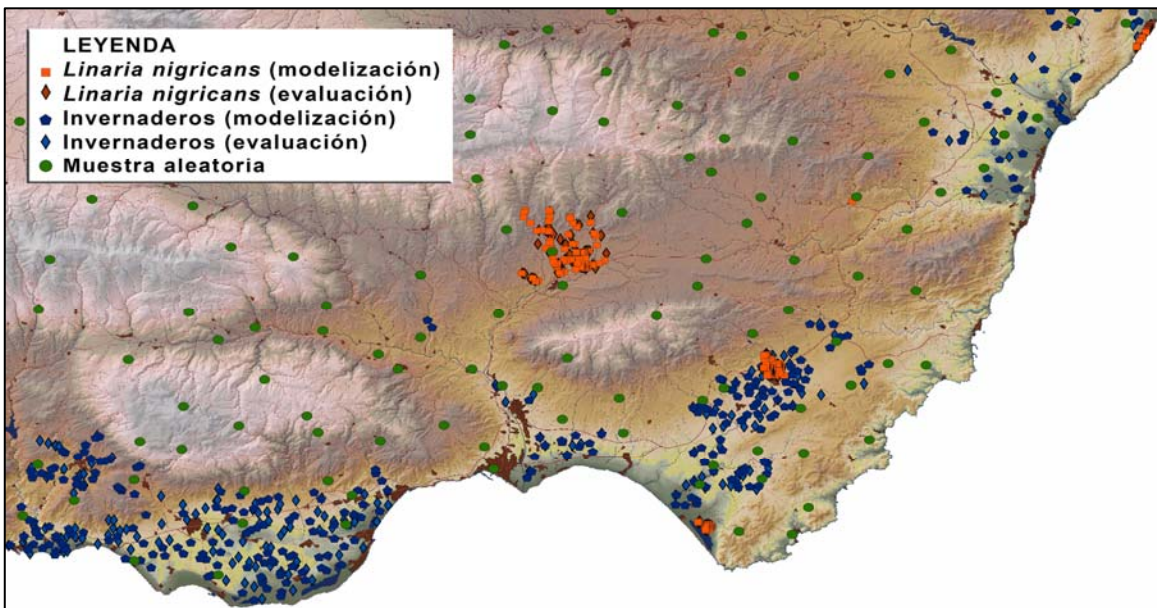


Figura 2. Registros de presencia y muestra aleatoria.

Fuente: Elaboración propia.

Benito de Pando, B. y Peñas de Giles, J. (2007): "Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica", *GeoFocus (Artículos)*, nº 7, p. 100-119, ISSN: 1578-5157

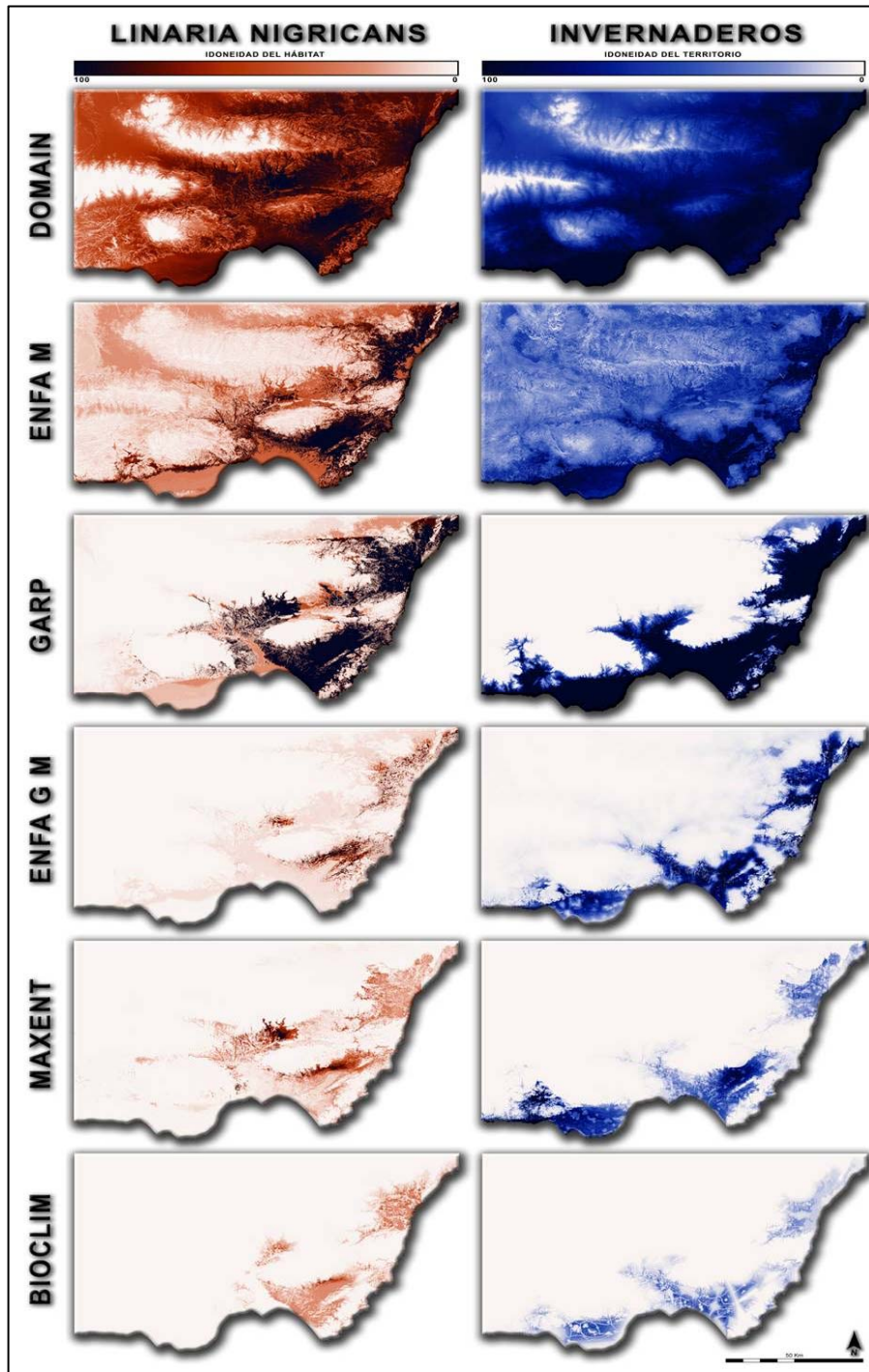


Figura 3. Resultados de los distintos algoritmos.

Fuente: Elaboración propia.

Benito de Pando, B. y Peñas de Giles, J. (2007): "Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica", *GeoFocus (Artículos)*, nº 7, p. 100-119, ISSN: 1578-5157

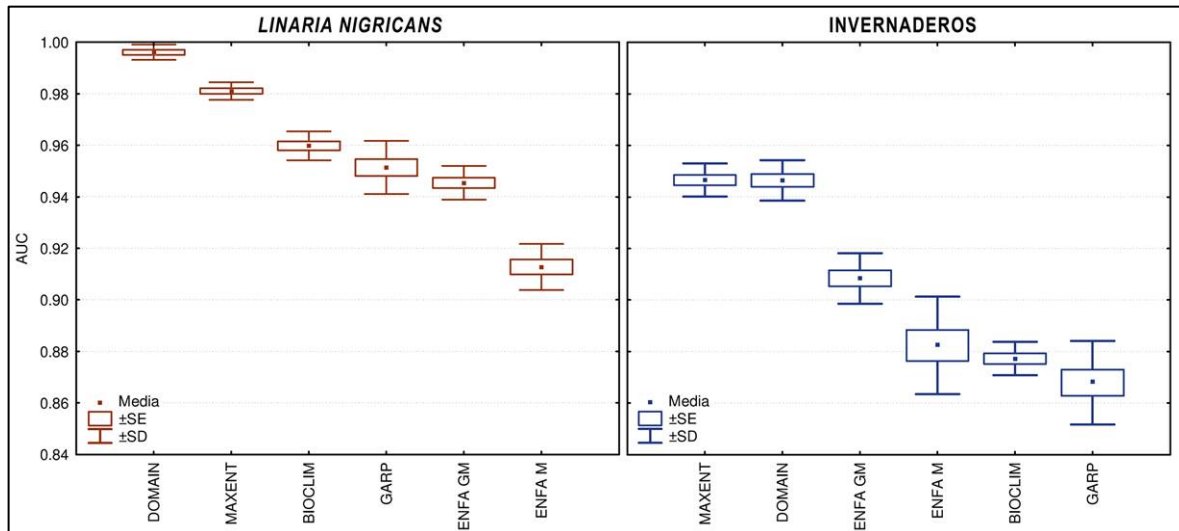


Figura 4. Resultados de la evaluación de modelos.

Fuente: Elaboración propia.

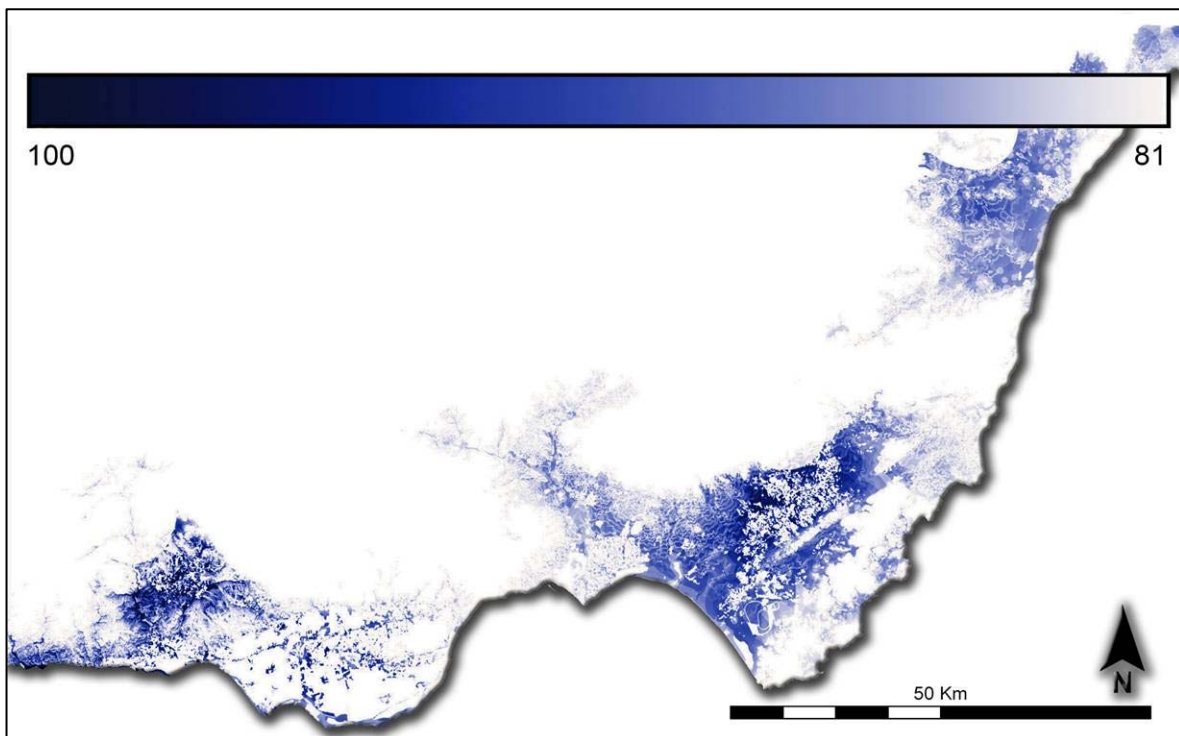


Figura 5. Idoneidad del territorio para invernaderos.

Fuente: Elaboración propia.

Benito de Pando, B. y Peñas de Giles, J. (2007): "Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica", *GeoFocus (Artículos)*, nº 7, p. 100-119, ISSN: 1578-5157

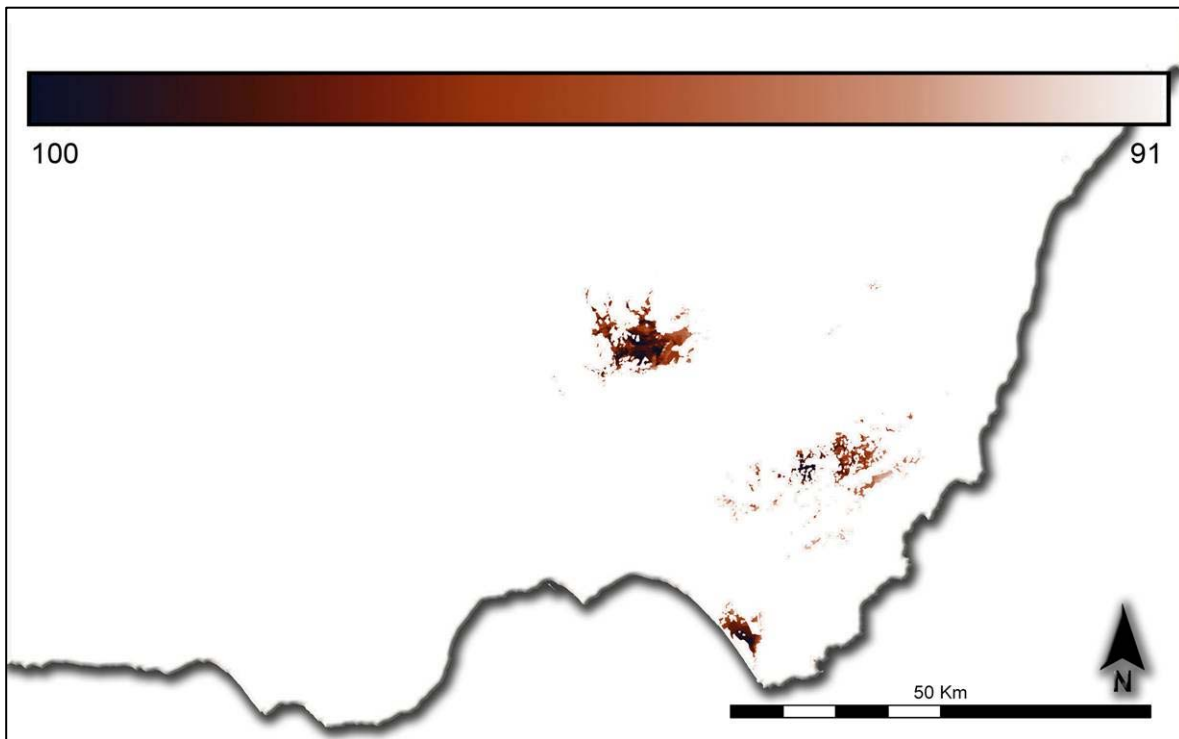


Figura 6. Hábitat potencial de *L. nigricans*.

Fuente: Elaboración propia.

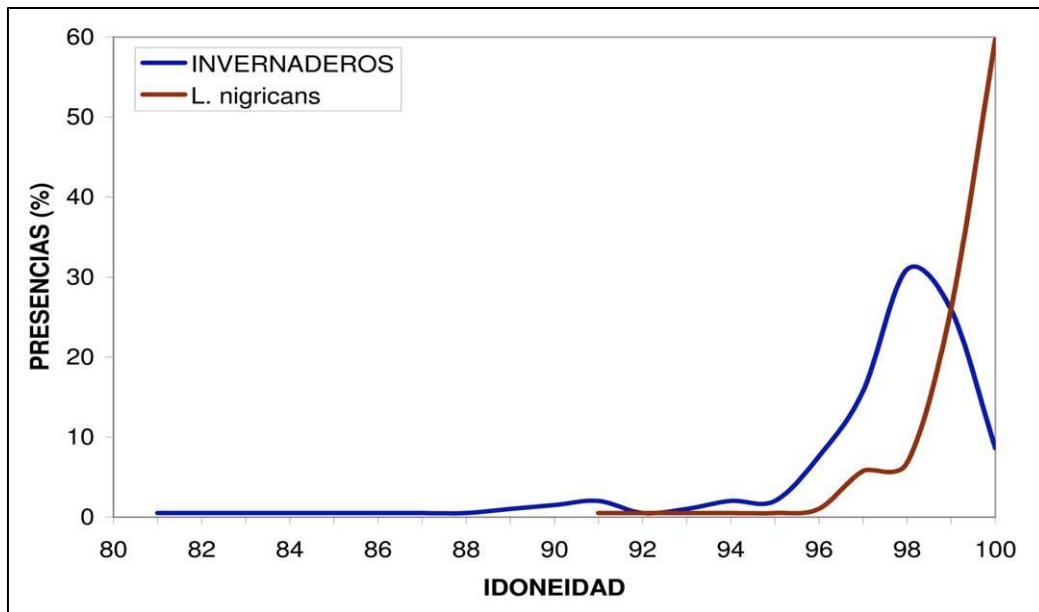


Figura 7. Curvas de saturación de hábitat.

Fuente: Elaboración propia.

Benito de Pando, B. y Peñas de Giles, J. (2007): "Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica", *GeoFocus (Artículos)*, nº 7, p. 100-119, ISSN: 1578-5157

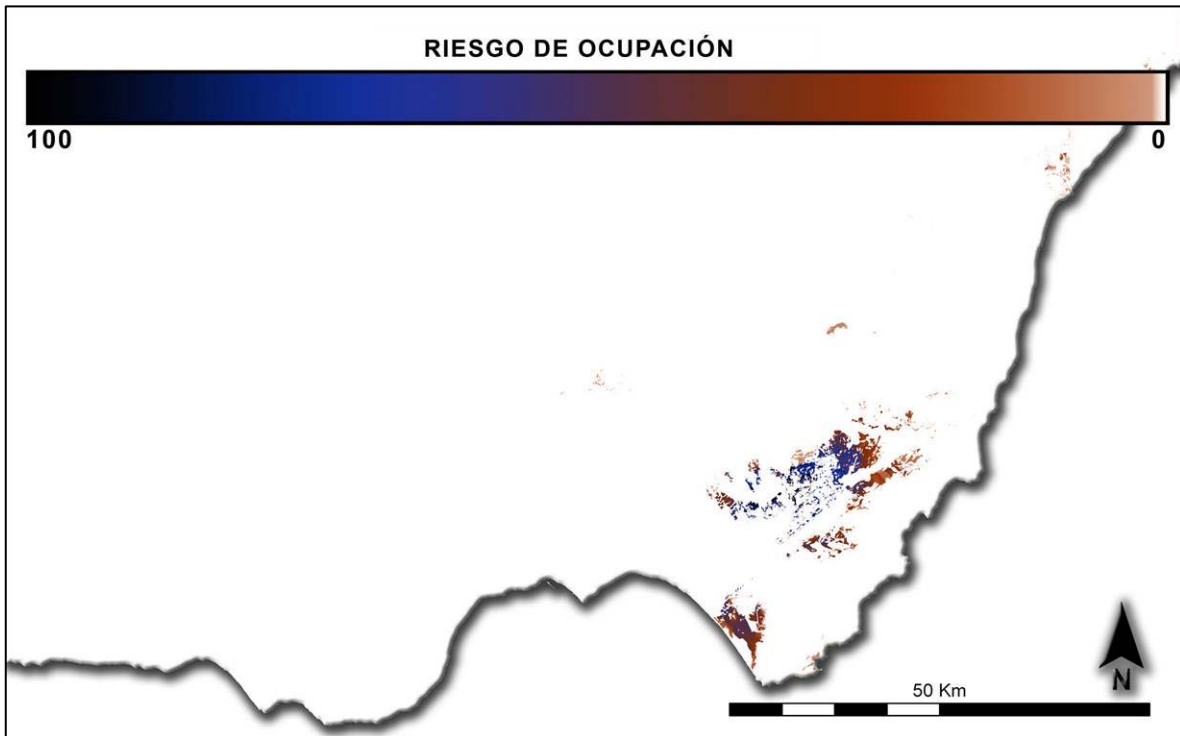


Figura 8. Modelo de riesgo de extinción.

Fuente: Elaboración propia.

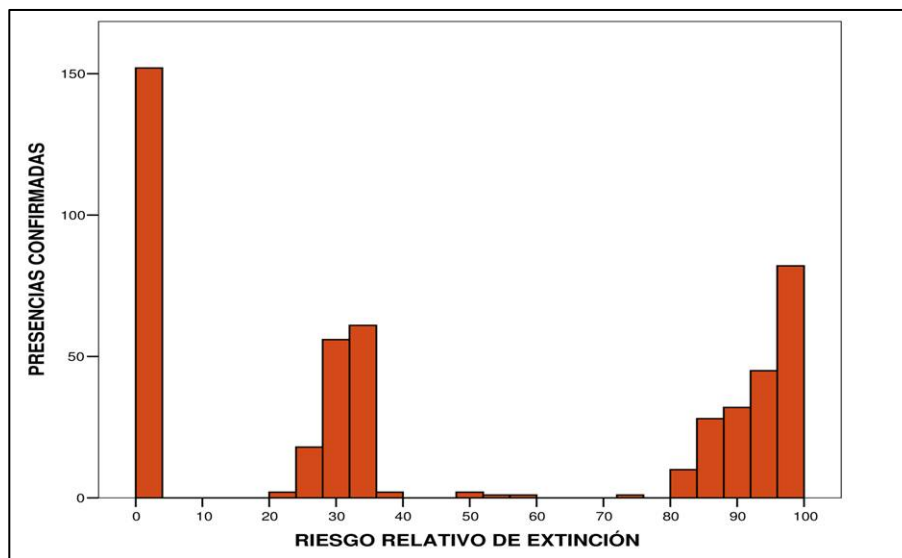


Figura 9. Riesgo de extinción relativo de puntos de presencia de *L. nigricans*.

Fuente: Elaboración propia.