

## MODELOS PREDICTIVOS DE LA DISTRIBUCIÓN DE ESPECIES: UNA REVISIÓN DE SUS LIMITACIONES

JAVIER SEOANE\* Y JAVIER BUSTAMANTE\*

### RESUMEN

En las últimas dos décadas se ha despertado un enorme interés en el modelado de la relación entre las especies y sus hábitat, que responde tanto a la demanda de información aplicable a la gestión del territorio y a la conservación como al fundamento básico de la ecología en estudiar la distribución y abundancia de los organismos. Sin embargo, los modelos predictivos de distribución de especies descansan en ciertas presunciones y tienen unas limitaciones que conviene conocer antes de desarrollarlos. En este trabajo se ofrece primero un breve resumen de los tipos de modelado que pueden encontrarse en estudios de ecología, centrándose en los modelos monoespecíficos de distribución, es decir, en aquellos que relacionan las características del hábitat con la presencia de una especie en particular. Posteriormente, se presenta una síntesis comentada de las limitaciones de carácter biológico y estadístico de los modelos predictivos, analizando en detalle las presunciones en que se sostienen y los problemas metodológicos que dificultan su aplicación. Se concluye que los modelos de distribución de especies están sujetos a numerosos defectos, pero su desarrollo puede ofrecer una interesante herramienta complementaria en la gestión del territorio.

**Palabras clave:** modelos predictivos, relaciones especies-hábitat, selección de hábitat.

### SUMMARY

In the last two decades there has been a growing interest in modelling wildlife-habitat relationships. This is due both to the necessity of basic information for land management and conservation, and to the fundamental interest of Ecology in studying the distribution and abundance of organisms. However, wildlife-habitat models relay on several assumptions, and have some limitations that must be known. This work offer first a brief summary of the type of models that can be found in ecological studies. The focus is on monoespecific models of species distribution, that is, in those that relate habitat characteristics with the presence/absence of a single species, but the discussion can be extended to other model types, in particular those which deal with several species at a time. Second, a commented synthesis on both statistical and biological limitations of the distribution models is given in detail, with an analysis of the underlying assumptions and methodological problems. In conclusion, distribution models have numerous shortcomings but their development may provide a worthy tool for land management.

**Key words:** predictive models, wildlife-habitat relationships, habitat selection.

---

\* Departamento de Biología Aplicada. Estación Biológica de Doñana. Avda. María Luisa, s/n. 41013 Sevilla. e-mail: seoane@ebd.csic.es

Recibido: 15/11/00.

Aceptado: 06/02/01.

## INTRODUCCIÓN

En las últimas dos décadas se ha despertado un enorme interés en el análisis de la relación entre las especies y sus hábitat, extendiéndose los estudios de selección de hábitat a la realización de modelos que predicen la distribución y abundancia de especies. Esta tendencia es un producto del doble interés de estos modelos. Por una parte, existe una fuerte demanda de información en numerosos problemas de conservación en los que las relaciones de las especies con sus hábitat son primordiales y, por otra, la ecología tiene un interés primario en estudiar la distribución y abundancia de los organismos, lo que algunos autores han identificado como su objetivo principal (p.e. BEGON *et al.* 1995, p.124). En consonancia con esta situación, recientemente han aparecido diversos trabajos realizados en la Península Ibérica que desarrollan modelos de distribución de especies (GONZÁLEZ *et al.* 1990; GONZÁLEZ *et al.* 1992; DONÁZAR *et al.* 1993; BUSTAMANTE 1996, 1997; BRITO *et al.* 1999; FRANCO *et al.* 2000; MARTÍNEZ PALAO *et al.* 2000; SÁNCHEZ-ZAPATA & CALVO 1999; SUÁREZ *et al.* 2000), y es de esperar que su número siga aumentando en un futuro próximo, dadas las perspectivas optimistas de su posible uso en la gestión del medio natural.

La utilidad general de los modelos de distribución de especies radica en que permiten trabajar con muestras incompletas acerca de la distribución o abundancia de especies, lo que es especialmente importante en los estudios en áreas remotas o de difícil acceso, donde no resulta práctico llegar a la totalidad del territorio, o bien en trabajos en que los recursos sean insuficientes para ello (OSBORNE & TIGAR 1992; SKOV & BORCHSENIUS 1997; MANEL *et al.* 1999). Los datos recogidos en un muestreo se extienden al conjunto del área de interés mediante la generación de mapas de carácter predictivo (ver p. ej. MLADENOFF *et al.* 1995) entre cuyos valores principales se encuentra, en nuestra opinión, el que pueden ser una herramienta útil para los gestores del territorio. Finalmente, si las variables predictoras pueden derivarse de sensores remotos (fotografía aérea, imágenes de satélite), la información proporcionada por estos sensores podría servir para crear

mapas predictivos fácilmente actualizables (PALMEIRIM 1988; AVERY & HAINES-YOUNG 1990; MILLER & CONROY 1990; ANDRIES *et al.* 1994; PARUELO & GOLLUSCIO 1994). Estos modelos han sido utilizados para evaluar las necesidades de protección en un territorio (SCOTT *et al.* 1993; BOJÓRQUEZ-TAPIA *et al.* 1995). La modelización por separado de un gran número de especies (o la modelización de la riqueza) permite identificar áreas de distinto interés conservacionista, como pueden ser áreas ricas en especies o en táxones amenazados, para tenerlas en cuenta en la declaración de espacios protegidos. El ejemplo paradigmático de esta aproximación es el análisis GAP (SCOTT *et al.* 1993), actualmente muy desarrollado en Estados Unidos, aunque no está exento de críticas (SHORT & HESTBECK 1995; CONROY & NOON 1996). En particular, no hay un acuerdo sobre si existe una coincidencia geográfica de la riqueza, rareza o grado de amenaza entre diferentes táxones (ver p. ej. WILLIAMS & GASTON 1994; CASTRO *et al.* 1996; PRENDERGAST 1997; WILLIAMS *et al.* 1996). Por último, otra extendida aplicación de los modelos de las relaciones entre las especies y sus hábitat es la predicción de impactos, ya sean naturales, como en los estudios que sugieren posibles modificaciones en la distribución de especies relacionados con el cambio climático o incendios (BOX *et al.* 1993; HE & MLADENOFF 1999); ya sean impactos artificiales, como los provocados por infraestructuras, actividades extractivas o cambios en el uso del territorio (AVERY & HAINES-YOUNG 1990; LAVERS & HAINES-YOUNG 1996). En estos casos los modelos son herramientas que permiten decidir entre alternativas de gestión del territorio (TURNER *et al.* 1995).

A pesar de las numerosas aplicaciones de los modelos de las relaciones entre las especies y sus hábitat, éstos son una representación incompleta de la realidad y, por tanto, tienen limitaciones de las que deben ser conscientes quienes los desarrollan y quienes los utilizan. En este trabajo se ofrece primero un breve resumen de los tipos de modelado que pueden encontrarse en estudios de ecología, centrándose en los modelos mono-específicos de distribución, es decir, en aquellos que relacionan las características del hábitat con la presencia de una sola especie (aunque el razonamiento puede exten-

derse fácilmente a otro tipo de modelos, en particular los multiespecíficos). Posteriormente, se presenta una síntesis comentada de las limitaciones de carácter biológico y estadístico de los modelos predictivos, analizando en detalle las presunciones en que se sostienen y los problemas metodológicos que dificultan su aplicación.

Una versión resumida de este trabajo fue presentada en el I Congreso Ibérico de Ecología, celebrado en Santiago de Compostela (La Coruña) entre el 25 y 28 de septiembre de 2000 y organizado por la Asociación Española de Ecología Terrestre y la Sociedade Portuguesa de Ecología.

## Tipos de modelos

En lo que sigue se considerarán sólo los modelos empíricos que relacionan la distribución de una sola especie (es decir, su existencia y/o abundancia en un área, lo que denominaremos *variable respuesta*) con un conjunto de variables del medio que describen aspectos bióticos, físicos o humanos a través de una formulación matemática o lógica (lo que denominaremos *variables predictoras*). Como modelo empírico se entenderán en este trabajo aquellos que se basan en datos reales, como es común en la mayor parte de los trabajos de campo, en oposición a los modelos teóricos (*sensu* MORRISON *et al.* 1998) cuya formulación parte de supuestos de funcionamiento de un hipotético sistema u organismo. Por tanto, no se tendrán en cuenta los modelos multiespecíficos (p. ej. análisis GAP, SCOTT *et al.* 1993), los que usan como variables predictoras rasgos vitales de los organismos (como el tamaño corporal y parámetros demográficos, ver LAWTON 1993), ni los que se basan exclusivamente en técnicas de interpolación espacial (en estos modelos los valores de la variable respuesta en puntos no muestreados dentro del rango espacial del estudio se estiman en función de su ubicación geográfica con respecto a los puntos vecinos muestreados, ver CRESSIE 1993; MAURER 1994). La discusión se centrará en modelos de tipo correlativo, definidos aquí como los que se basan en correlaciones, no necesariamente causales, entre variables, aunque gran parte de lo que se expone puede aplicarse a la mayoría de los modelos de las relaciones entre las especies y sus hábitat (MORRISON *et al.* 1998).

Los modelos que más se han usado en ecología para predecir la distribución de especies podrían dividirse en tres grandes grupos atendiendo a su funcionamiento: los que estiman el rango de tolerancia ecológica, los modelos de tipo correlativo y ordenación multivariante y redes neuronales artificiales.

Dentro del primer tipo de modelos se encuentran los llamados de análisis de superposición (BRITO *et al.* 1999) y los de envuelta climática (AUSTIN *et al.* 1990; BOX *et al.* 1993), que podrían considerarse extensiones de análisis típicos de los Sistemas de Información Geográfica. El funcionamiento de estos modelos es como sigue. Primero se identifican los lugares en las que una especie está presente y se calculan los valores mínimos y máximos de las variables ambientales que se considera a priori que afectan a su distribución (generalmente variables descriptoras del clima, la altitud, etc.). La extensión de los resultados al conjunto del área de estudio se hace suponiendo que los lugares adecuados para la especie son aquellos cuyos valores de todas las variables predictoras estén dentro de los rangos en los que se ha observado a tal especie. Estos modelos dependen de una selección adecuada de las variables ambientales y normalmente tienden a sobreestimar la extensión areal ocupada (pueden subestimarla si se seleccionan demasiadas variables de escasa relevancia para la especie). La utilidad que se les suele reconocer es la de aportar un primer análisis orientativo, que es particularmente valioso en áreas extensas o escasamente prospectadas (SKOV & BORCHSENIUS 1997).

Al segundo grupo pertenece una gran variedad de modelos cuyo patrón común es que tratan de relacionar la presencia o la abundancia de una especie con distintas variables predictoras a través de una función matemática. Esta función permite establecer el tipo de relación que existe entre la variable respuesta y las predictoras. En general, el uso de técnicas como análisis discriminante (GONZÁLEZ *et al.* 1990; GONZÁLEZ *et al.* 1992) y regresiones lineales múltiples (CARRASCAL *et al.* 1993; DONÁZAR *et al.* 1989) han ido dejando paso a otras enmarcadas dentro de los Modelos Lineales Generalizados («Generalized Linear Models» o GLM, de las que las anteriores pueden consi-

derarse casos particulares), pues permiten una mayor flexibilidad al tratar los datos (NICHOLLS 1989; AUSTIN *et al.* 1990; DONÁZAR *et al.* 1993; BUSTAMANTE *et al.* 1997). Una mención especial merece la regresión logística, que es la técnica más usada, pues utiliza variables binomiales (p. ej. aquellas cuyas respuestas son 1 y 0) fácilmente entendibles en el contexto de análisis de presencia/ausencia. En cuanto a las técnicas de ordenación multivariante, éstas son utilizadas generalmente como paso previo a la modelización (CARRASCAL *et al.* 1993) para resumir un conjunto numeroso de variables en unas pocas variables sintéticas, pero pueden usarse por sí mismas para crear mapas predictivos de la distribución (análisis factorial del nicho ecológico, HAUSSER 1995; HIRZEL *et al.* 2000). Por último, existen técnicas de ajuste no paramétrico como regresiones locales o modelos aditivos generalizados (Generalized Additive Models o GAM, HASTIE & TIBSHIRANI 1990) que se han utilizado para aspectos similares a los que aquí se tratan y son el horizonte hacia el que probablemente irán tendiendo los próximos trabajos (THOMAS & NEIL 1991; FEWSTER *et al.* 2000).

Un último tipo de modelos, que se separa en esta revisión por sus peculiaridades, es el de redes neuronales artificiales, denominadas así porque están basados en un modelo conceptual del funcionamiento del cerebro. En este caso los efectos de las distintas variables predictoras sobre la respuesta se transforman y se combinan en grupos denominados *neuronas* cuyo número se determina subjetivamente. Estas combinaciones y el peso relativo de cada neurona en la respuesta final se modifican iterativamente (a través de los denominados «algoritmos de entrenamiento») hasta dar con un ajuste a los datos que se considere apropiado (STATSOFT 1999). Se argumenta en favor de su utilización que pueden modelar relaciones no lineales muy complejas (LEK *et al.* 1996) y en su contra que no permiten reconocer fácilmente posibles relaciones causales entre los predictores y la respuesta (originan modelos de tipo «caja negra») y que requieren mayor tiempo de computación (MANEL *et al.* 1999). El modelado de la distribución de especies a través de redes neuronales

se ha emprendido recientemente y sus ejemplos son todavía escasos (MASTROIRILLO *et al.* 1997; MANEL *et al.* 1999).

## LIMITACIONES DE LOS MODELOS

### Presunciones

El modelado de la abundancia o presencia de especies en función de variables del hábitat reposa sobre dos presunciones básicas: 1) que la variable respuesta es independiente entre localidades y 2) que todas las variables predictoras importantes se incluyen en el modelo (LENNON 1999). Puesto que habitualmente no se sabe a priori cuáles son las importantes, es necesario hacer una selección a través de tests estadísticos, de manera que el modelo final retiene sólo las variables que se consideran significativas, de acuerdo con el principio de parsimonia por el que se prefieren modelos sencillos con pocas variables a otros más complejos que expliquen lo mismo. Sin embargo, la primera de las presunciones es probablemente falsa en la mayoría de los trabajos, ya que las condiciones ambientales en un punto de estudio tenderán a ser similares en un área próxima y, por tanto, las especies ligadas a tales condiciones tenderán a presentarse también en los puntos vecinos. Además, no es raro que las especies aparezcan distribuidas de forma agregada, puesto que los individuos establecidos en un área pueden ejercer un efecto de atracción hacia nuevos colonizadores o condicionar la dispersión de los descendientes, de forma que la probabilidad de encontrar a una especie en un lugar podría no ser independiente de la probabilidad de encontrarla en lugares vecinos (LEGENDRE & TROUSSELLIER 1988; AUGUSTIN *et al.* 1996). Estos dos aspectos originan lo que se conoce como autocorrelación espacial de la variable respuesta. Cuando ésta existe, los tests estadísticos que seleccionan las variables predictoras tienden a incorporar en los modelos aquellas variables que cambien espacialmente de una forma gradual, lo cual impide hacer una interpretación biológica del modelo y perjudica su capacidad de ser aplicado a otros lugares, aunque puede resultar conveniente si el objetivo es explicar una distribución en un área determinada (AUGUSTIN

*et al.* 1996; LENNON 1999). El método más comúnmente utilizado para tener en cuenta la autocorrelación espacial se basa en incorporar a los modelos una o distintas variables predictoras que informen del estado de la variable respuesta en un área vecina cuya extensión se decide empíricamente. Por tanto, en estos modelos, denominados autologísticos por ser un refinamiento de la regresión logística, las relaciones de vecindad se analizan como una variable predictora más (PREISLER 1993; SMITH 1994; AUGUSTIN *et al.* 1996; CHOUT & SORET 1996; WU & HUFFER 1997).

Por otro lado, los modelos de la distribución de especies asumen implícitamente que los hábitat están saturados, es decir, que todo hábitat adecuado para una especie estará ocupado por ella. Sin embargo, si un organismo cuya selección de hábitat se pretende modelar muestra una dinámica poblacional en la que hay efectos de fuente-sumidero, un área adecuada podría estar vacía si aún no hubiera sido colonizada o si la población existente se hubiera extinguido por causas naturales o provocadas por el hombre (DÍAS 1996). Además, puede haber interacciones entre especies (p.ej., predación o competencia) que hagan que un hábitat en otro caso adecuado no esté ocupado (LAWTON & WOODROFFE 1991). Esta situación origina los tipos de error por comisión, en los que se predice erróneamente la presencia de una especie en un lugar (FIELDING & BELL 1997).

Otra presunción fundamental es que la probabilidad de detección de una especie será mayor en sus hábitat óptimos. En los modelos de distribución se mide generalmente la abundancia de una especie en distintas áreas caracterizadas por un conjunto de variables y se equipara la abundancia con la calidad del hábitat para esa especie. No obstante, existen ciertos procesos naturales que hacen que esta presunción pueda ser falsa en algunos casos (VAN HORNE 1983). Por un lado, la distribución actual de una especie podría reflejar situaciones pasadas, si existen cambios en la densidad de los individuos de frecuencia plurianual que sigan variaciones a escala local en factores que influyan en la demografía, como la intensidad de depredación o la cantidad de ali-

mento. Por otro lado, en poblaciones animales que desarrollen jerarquía social, los individuos desfavorecidos (subadultos inexpertos, ejemplares enfermos, etc.) pueden ser desplazados a ambientes subóptimos en los que podrían adquirir gran abundancia (DÍAS 1996). Finalmente, los índices de selección de hábitat pueden estar afectados por el tamaño poblacional, de manera que, por ejemplo, un hábitat de elevados recursos podría usarse mucho hasta que la población creciera tanto que la presión de competencia intraespecífica condujera a ocupar hábitat subóptimos con menor competencia (HOBBS & HANLEY 1990). En esta línea de razonamiento se ha destacado que una población más numerosa no significa que esté en mejores condiciones; así se ha mostrado (HOBBS & SWIFT 1985) que un área con abundantes recursos de baja calidad puede mantener a una gran población infraalimentada, mientras que otra área de escasos recursos de alta calidad soporta a pocos individuos de, probablemente, mayor eficacia biológica (*fitness*). Por último, existen simulaciones en las que se recrea un hábitat fragmentado y revelan que el tamaño medio poblacional de los fragmentos está influido principalmente por la dispersión de individuos entre ellos y no por su capacidad de carga (FAHRIG & PALOHEIMO 1988). Por estos motivos se ha propuesto que la adecuación de un hábitat se mida baremando la abundancia de las especies con la eficacia biológica de los individuos que lo ocupan (VAN HORNE 1983).

Si los modelos trabajan sobre un muestreo de áreas disponibles (o unidades de recursos disponibles en la terminología de MANLY *et al.* 1993), se asume que éstas han sido escogidas independientemente y al azar, y que todos los individuos tienen las mismas probabilidades de acceso a ellas (BOYCE & McDONALD 1999). Además, el significado que en cada caso se dé al concepto de «disponibilidad» se ha identificado también como un problema importante en estudios de selección de hábitat (McCLEAN *et al.* 1998; WILSON *et al.* 1998) que cabe extender al contexto de modelos de distribución. Así, en un modelo de regresión logística típico se comparan las variables predictoras de un conjunto de localidades en el que se ha observado a una especie con otro en

la que se la supone ausente, de manera que, a efectos analíticos, el área disponible es la suma de las áreas de los dos conjuntos de localidades: a mayor área en que se midieron los predictores en torno a cada localidad, mayor área disponible y, según los trabajos anteriores, mayor probabilidad de cometer error de tipo I al incorporar variables espurias a los modelos.

**Limitaciones**

*Unas de carácter biológico...*

Existen varias razones para esperar que los modelos no funcionen correctamente. Las dos primeras que se subrayan aquí son producidas por el propio fenómeno que se quiere modelar.

Por un lado, cabría esperar que, hasta cierto punto, la distribución actual de una especie estuviera afectada por acontecimientos pasados (DÍAS 1996; y referencias en FIELDING & BELL 1997), lo cual podría ser especialmente relevante en organismos sésiles de larga vida, como muchos táxones vegetales. Así, por ejemplo, una especie podría habitar un área que colonizó hace tiempo y que hoy en día carece de las condiciones que le son más favorables; tal especie podría estar sufriendo un lento declive en esa zona, pero los modelos (que generalmente se desarrollan en un intervalo de tiempo breve) no lo detectarían.

Por otro lado, clases distintas de individuos de una población podrían mostrar una selección de hábitat diferente, dependiendo, por ejemplo, de sus estatus social (es decir, no por una simple expulsión de individuos subordinados a hábitat subóptimos, sino por una selección activa de hábitat diferentes con distinta oferta de recursos, ver ARDIA & BILDSTEIN 1997). Esta situación se ha encontrado en aves rapaces (CADE 1955; KOPLIN 1973; SMALLWOOD 1987; BUSTAMANTE *et al.* 1997) y podría ser típica de organismos animales que exhiban una jerarquía social cuyos distintos grupos fueran dimórficos (p. ej., en las rapaces diurnas las hembras suelen ser mayores que los machos y los tamaños de las presas que les suponen un beneficio óptimo son diferentes).

Por último, los modelos de distribución se limitan implícitamente a poblaciones en equilibrio cuya relación con el hábitat no cambia (BOYCE & McDONALD 1999), de otra manera sería necesario hacer un modelo para cada situación (ARTHUR *et al.* 1996).

*... y otras de corte metodológico*

Existen además varios problemas metodológicos en el tipo de modelización que aquí se trata que impiden que los modelos sean perfectos. Éstos se refieren a la comparación de modelos mediante medidas de error de la predicción, a la conversión de probabilidades dadas por los modelos a valores de presencia o ausencia y a la naturaleza correlativa de las relaciones que se establecen entre la variable respuesta y las predictoras.

Las predicciones de los modelos de la distribución que tratan con datos de presencia/ausencia se analizan con una matriz de confusión (figura 1) y pueden estar erradas de dos formas: en las presencias (falsos positivos) y en las ausencias (falsos negativos). Las distintas medidas de error (tabla 1) tienen características diferentes y, en particular, algunas están influenciadas por la prevalencia. Por ejemplo, suponiendo que se realizó un muestreo de 100 lugares y en sólo 10 apareció la especie objeto del estudio (N=100, a+c=10, b+d=90), un modelo trivial sería suponer que ninguno de los lugares es apto para ella (a+b=0, c+d=100), lo que daría una tasa de clasificación correcta del 90% [(a+d)/N=0+90/100, ver tabla 1].

Además, los modelos dan generalmente valores continuos para las predicciones dentro del intervalo (0,1), pero los valores de la matriz de confusión que se utiliza para compararlos son valores

		Observado	
		Presencia	Ausencia
Predicho	Presencia	a	b
	Ausencia	c	d

Fig. 1. Matriz de confusión. a: presencias predichas correctamente; b: falsos positivos; c: falsos negativos; d: ausencias predichas correctamente. [Confusion matrix. a: presences correctly predicted; b: false positives; c: false negatives; d: absences correctly predicted.]

TABLA 1

MEDIDAS DE ERROR (MODIFICADO DE FIELDING & BELL 1997) Y PREGUNTAS A LAS QUE RESPONDEN.  
 [ERROR MEASURES (MODIFIED FROM FIELDING & BELL 1997) AND QUESTIONS THAT THEY ADDRESS.]

Medida	Cálculo
Prevalencia <i>¿en qué fracción de puntos ha aparecido una especie?</i>	$(a+c)/N$
Tasa de clasificación correcta total <i>¿qué fracción de puntos se predijo correctamente?</i>	$(a+d)/N$
Tasa de clasificación incorrecta <i>¿qué fracción de puntos se predijo incorrectamente?</i>	$(b+c)/N$
Sensibilidad <i>¿qué fracción de las presencias se predijeron correctamente?</i>	$a/(a+c)$
Especificidad <i>¿qué fracción de ausencias se predijeron correctamente?</i>	$d/(b+d)$
Poder predictivo positivo <i>de las presencias predichas ¿qué fracción es correcta?</i>	$a/(a+b)$
Poder predictivo negativo <i>de las ausencias predichas ¿qué fracción es correcta?</i>	$d/(c+d)$
Kappa <i>¿qué fracción de puntos se predijo correctamente teniendo en cuenta lo esperable al azar?</i>	$[(a+d)-\{(a+c)(a+b)+(b+d)(c+d)\}/N] / [N-\{(a+c)(a+b)+(b+d)(c+d)\}/N]$

discretos 0 ó 1. Esto hace que los valores de probabilidad de aparición hayan de ser convertidos, de manera que se adjudique la presencia de la especie a todas las áreas cuya probabilidad de aparición supere un umbral. El problema reside en la elección de este punto umbral, al que son sensibles las medidas de error (ver figura 2 y BRITO *et al.* 1999; FRANCO *et al.* 2000, para ejemplos reales). La elección de 0,5 como umbral en el ejemplo de la figura 2 conduciría a un bajo poder predictivo positivo (aproximadamente la mitad de las presencias predichas serían reales). La adopción de un umbral más bajo, por ejemplo 0,3, aumentaría el poder predictivo para las presencias (hasta 0,85), mientras que un umbral mayor, como 0,8, aseguraría un elevado poder predictivo para las ausencias (cerca del 0,9). Cada estrategia se adecuaría a distintos escenarios, por ejemplo, la primera en la selección de espacios que albergaran a una especie a proteger y la segunda en la selección de áreas alternativas para la ubicación de actividades humanas de gran impacto para cierta especie.

Se ha descrito un método de comparación de modelos que evita el problema de la influencia del punto de corte sobre las medidas de error. Se trata de los diagramas ROC (de Receiver Operating Characteristic, ZWEIG & CAMPBELL 1993)

en los que se representa la sensibilidad de un modelo, en ordenadas, contra su especificidad, en abscisas, para todos los puntos umbral, de forma que el modelo que esté por encima en el diagrama tendrá una mayor exactitud relativa. Los diagramas ROC no informan de cuál es el punto umbral óptimo, pero existen métodos por los que pueden ser utilizados para conseguir esta

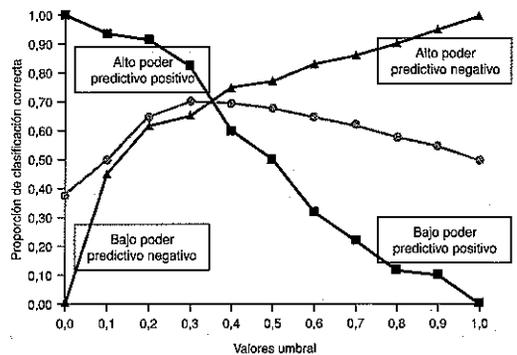


Fig. 2. Ejemplo de diagrama de la relación entre la tasa de clasificación correcta (cuadrados para las presencias, triángulos para las ausencias y círculos para el total) y el punto umbral para convertir los valores de probabilidad en valores de presencia/ausencia. [Example of plot of the relation between correct classification rate (squares for presences, triangles for absences and circles for the overall) and the threshold used to convert probability values on presence/absence.]

información (referencias en FIELDING & BELL 1997). Sin embargo, existen pocos ejemplos prácticos en ecología del uso de los diagramas ROC (MANEL *et al.* 1999).

Como corolario cabe decir que tanto la elección de las medidas de error con las que se comparen los modelos como, si procede, la elección del punto de corte, han de ser escogidos con especial atención a las preguntas más relevantes en el contexto de la investigación que se esté realizando (FIELDING & BELL 1997; MORRISON *et al.* 1998).

Las relaciones entre las variables respuesta y explicativas que se modelan suelen tener una naturaleza correlativa, por lo que no revelan necesariamente pautas de causa y efecto. Esto hace que los modelos de distribución puedan fracasar en su aplicación a otras áreas (o tiempos). La solución sería utilizar variables predictoras causales en los modelos, pero esto excede nuestro conocimiento actual sobre la mayor parte de las especies. Además, el desarrollo de modelos causales probablemente exigiría más tiempo del que permiten las necesidades de conservación que conducen a la modelización (en particular, ¿cómo se identifican a qué variables del medio realmente responde una especie?; y una vez hecho esto, ¿cómo conocer la distribución espacial de esas variables predictoras, pues la cartografía disponible se ha creado con otro propósito?). Sin embargo, algunos autores han sugerido que la única solución posible reside en el uso de variables causales, pues la concatenación de presunciones podría conducir necesariamente a un bajo poder predictivo (BEUTEL *et al.* 1999).

Por otro lado, los modelos suelen desarrollarse en un contexto multivariante, donde las correlaciones entre las variables (colinearidad) son muy probables. La multicolinealidad hace que puedan incorporarse a los modelos variables espurias, y que queden fuera otras más próximas a la causales (FLACK & CHANG 1987). De nuevo, este problema reduce la capacidad de extrapolación de los modelos (es decir, la fiabilidad con que pueden ser aplicados en otras áreas distintas a aquellas en que se generaron) y hace que las distintas técnicas de modelado pro-

duzcan resultados diferentes (MAC NALLY 2000). Una solución interesante a este problema es calcular todos los modelos posibles con las variables de que se dispone y escoger entre ellos según criterios que promedian la información recogida por el modelo y su complejidad (criterios de información de Akaike, AIC o bayesiano de Schwarz, BIC; AKAIKE 1978; SCHWARZ 1978). Se ha propuesto además que la incertidumbre asociada a la selección de los modelos, es decir, a la elección tanto de las variables predictoras como de la forma en que estas varían con la variable respuesta, se tenga en cuenta en las predicciones, ponderando los resultados procedentes de distintos modelos (BUCKLAND *et al.* 1997). Sin embargo, estas aproximaciones no son muy frecuentes en ecología (ver un ejemplo en TONER & KEDDY 1997). Finalmente, un método diseñado para identificar variables causales —pero no para generar modelos predictivos— reduciendo los problemas de multicolinealidad es la partición jerárquica (CHEVAN & SUTHERLAND 1991; CHRISTENSEN 1992), mediante la cual se calcula la influencia relativa de una variable en todos los modelos en los que aparece. Sus resultados pueden compararse con los modelos seleccionados mediante un criterio de información para ilustrar su grado de causalidad (MAC NALLY 2000).

Cabe hacer un último comentario acerca de la posibilidad de validar los modelos de distribución. La verificación de modelos numéricos de sistemas naturales, si se entiende como tal la demostración de su certeza, es imposible porque tales sistemas no son cerrados y los resultados no son singulares (es decir, varios modelos pueden originar los mismos resultados). Según ORESKES *et al.* (1994) el término validación se emplea con dos significados erróneos; el primero es el de que las predicciones son consistentes con las observaciones, y el segundo el de que el modelo refleja con precisión la realidad. Estos autores afirman que los modelos sólo se pueden confirmar, entendiendo este término como la comprobación de que las observaciones coinciden con las predicciones; y subrayan que la confirmación no demuestra la hipótesis (el modelo), sólo apoya su probabilidad (ORESKEKES *et al.* 1994).

## CONCLUSIÓN

Enfrentados a las limitaciones que se han expuesto, y ante un éxito muy variable al extrapolar modelos entre zonas geográficas distintas, algunos autores han recomendado cautela en la aplicación de los modelos a problemas de conservación, llegando a sugerir que la distribución de las especies podría ser impredecible (FIELDING & HAWORTH 1995).

Sin embargo, los modelos de la relación especies-hábitat han mostrado su utilidad en distintas áreas y proporcionan una herramienta de, al menos, interés heurístico. La incorporación de variables causales y la atención a medidas de eficacia biológica aumentarían probablemente su valor. En cualquier caso merece la pena citar aquí a MORRISON (1999, p. 313) que da una perspectiva prác-

tica a la utilización de modelos: «Models are like politicians: support them, use them, but don't unquestioningly trust them».

## AGRADECIMIENTOS

Este trabajo es una contribución al proyecto «Cartografía predictiva de la distribución de aves terrestres: un estudio piloto en Andalucía Occidental», financiado por la Dirección General de Enseñanza Superior e Investigación Científica (Ministerio de Ciencia y Tecnología) y fondos FEDER de la UE (proyecto # 1DF-97-0648). J.S. disfruta de una beca predoctoral del Ministerio de Educación y Cultura. Los autores desean agradecer los comentarios de dos revisores anónimos que contribuyeron a mejorar el trabajo.

## REFERENCIAS BIBLIOGRÁFICAS

- AKAIKE, H. 1978. A Bayesian analysis of the minimum AIC procedure. *Annals Institute Statistics Mathematics* 30: 9-14.
- ANDRIES, A.M., GULINCK, H. & HERREMANS, M. 1994. Spatial modelling of the barn owl *Tyto alba* habitat using landscape characteristics derived from SPOT data. *Ecography* 17: 278-287.
- ARDIA, D.R. & BILDSTEIN, K.L. 1997. Sex-related differences in habitat selection in wintering American kestrels *Falco sparverius*. *Animal Behaviour* 53: 1305-1311.
- ARTHUR, S.M., MANLY, B.F.J., McDONALD, L. & GARNER, G.W. 1996. Assessing habitat selection when availability changes. *Ecology* 77: 215-227.
- AUGUSTIN, N.H., MUGGLESTONE, M.A. & BUCKLAND, S.T. 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33: 339-347.
- AUSTIN, M.P., NICHOLLS, A.O. & MARGULES, C.R. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Environmental Management* 60: 161-177.
- AVERY, M.I. & HAINES-YOUNG, R.H. 1990. Population estimates for the dunlin *Calidris alpina* derived from remotely sensed satellite imagery of the Flow Country of northern Scotland. *Nature* 344: 860-862.
- BEGON, M., HARPER, J.L. & TOWNSEND, C.R. 1995. *Ecología: individuos, poblaciones y comunidades*, Omega, Barcelona.
- BEUTEL, T.S., BEETON, R.J.S. & BAXTER, G.S. 1999. Building better wildlife-habitat models. *Ecography* 22: 219-223.
- BOJÓRQUEZ-TAPIA, L.A., AZUARA, I., EZCURRA, E. & FLORES-VILLELA, O. 1995. Identifying conservation priorities in Mexico through geographic information systems and modeling. *Ecological Applications* 5: 213-231.

- BOX, E.O., CRUMPACKER, D.W. & HARDIN, E.D. 1993. A climatic model for location of plant species in Florida, USA. *Journal of Biogeography* 20: 629-644.
- BOYCE, M. & McDONALD, L.L. 1999. Relating populations to habitats using resource selection functions. *Trends in Ecology and Evolution* 14: 268-272.
- BRITO, J.C., CRESPO, E.G. & PAULO, O.S. 1999. Modelling wildlife distributions: logistic multiple regression vs overlap analysis. *Ecography* 22: 251-260.
- BUCKLAND, S.T., BURNHAM, K.P. & AUGUSTIN, N.H. 1997. Model selection: an integral part of inference. *Biometrics* 53: 603-618.
- BUSTAMANTE, J. 1996. Statistical model of nest-site selection for the Bearded Vulture (*Gypaetus barbatus*) in the Pyrenees and evaluation of the habitat available with a geographical information system. En: J. Muntaner J. y M. (ed.). *Biología y conservación de las rapaces mediterráneas*, 1994. SEO, Madrid. pp. 393-400.
- BUSTAMANTE, J. 1997. Predictive models for lesser kestrel *Falco naumanni* distribution, abundance and extinction in southern Spain. *Biological Conservation* 80: 153-160.
- BUSTAMANTE, J., DONÁZAR, J.A., HIRALDO, F., CEBALLOS, O. & TRAVAINI, A. 1997. Differential habitat selection by immature and adult Grey Eagle-buzzards *Geranoaetus melanoleucus*. *Ibis* 139: 322-330.
- CADE, T. 1955. Experiments on winter territoriality of the American kestrel, *Falco sparverius*. *The Wilson Bulletin* 67: 5-17.
- CARRASCAL, L.M., BAUTISTA, L. M. & LÁZARO, E. 1993. Geographical variation in the density of the White stork (*Ciconia ciconia*) in Spain: influence of habitat structure and climate. *Biological Conservation* 65: 83-87.
- CASTRO, I., MORENO, J.C., HUMPHRIES, C.J. & WILLIAMS, P.H. 1996. Strengthening the Natural and National Park system of Iberia to conserve vascular plants. *Botanical Journal of the Linnean Society* 121: 189-206.
- CONROY, M.J. & NOON, B.R. 1996. Mapping of species richness for conservation of biological diversity: conceptual and methodological issues. *Ecological Applications* 6: 763-773.
- CRESSIE, N.A.C. 1993. *Statistics for spatial data*, John Wiley & Sons, Nueva York.
- CHEVAN, A. & SUTHERLAND, M. 1991. Hierarchical partitioning. *The American Statistician* 45: 90-96.
- CHOU, Y.H. & SORET, S. 1996. Neighborhood effects in bird distributions, Navarre, Spain. *Environmental Management* 20: 675-687.
- CHRISTENSEN, R. 1992. Comment on Chevan and Sutherland. *The American Statistician* 46: 74.
- DÍAS, P.C. 1996. Sources and sinks in population biology. *Trends in Ecology and Evolution* 11: 326-330.
- DONÁZAR, J.A., CEBALLOS, O. & FERNÁNDEZ, C. 1989. Factors influencing the distribution and abundance of seven cliff-nesting raptors: a multivariate study. En: B.-U. Meyburg & R.D. Chancellor, (eds.). *Raptors in the modern world*, pp. 545-552. WWGBP, Londres.
- DONÁZAR, J.A., HIRALDO, F. & BUSTAMANTE, J. 1993. Factors influencing nest site selection, breeding density and breeding success in the bearded vulture (*Gypaetus barbatus*). *Journal of Applied Ecology* 30: 504-514.

- FAHRIG, L. & PALOHEIMO, J. 1988. Determinants of local population size in patchy habitats. *Theoretical Population Biology* 34: 194-213.
- FEWSTER, R.M., BUCKLAND, S.T., SIRIWARDENA, G.M., BAILLIE, S. R. & WILSON, J.D. 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology* 81: 1970-1984.
- FIELDING, A.H. & BELL, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.
- FIELDING, A.H. & HAWORTH, P.F. 1995. Testing the generality of bird-habitat models. *Conservation Biology* 9: 1466-1481.
- FLACK, V.F., & CHANG, P.C. 1987. Frequency of selecting noise variables in subset regression analysis: a simulation study. *The American Statistician* 41: 84-86.
- FRANCO, A.M.A., BRITO, J.C. & ALMEIDA, J. 2000. Modelling habitat selection of Common Cranes *Grus grus* wintering in Portugal using multiple logistic regression. *Ibis* 142: 351-358.
- GONZÁLEZ, L.M., BUSTAMANTE, J. & HIRALDO, F. 1990. Factors influencing the present distribution of the Spanish imperial eagle *Aquila adalberti*. *Biological Conservation* 51: 311-319.
- GONZÁLEZ, L.M., BUSTAMANTE, J. & HIRALDO, F. 1992. Nesting habitat selection by the Spanish imperial eagle *Aquila adalberti*. *Biological Conservation* 59: 45-50.
- HASTIE, T.J. & TIBSHIRANI, R.J. 1990. *Generalized Additive Models*, Chapman & Hall, Londres.
- HAUSSER, J. 1995. Introduction. En: *Mammifères de la Suisse*, pp.1-14. Birkhäuser, Bâles.
- HE, H.S. & MLADENOFF, D.J. 1999. Spatially explicit and stochastic simulation of forest-landscape fire disturbance and succession. *Ecology* 80: 81-99.
- HIRZEL, A.H., HAUSSER, J. & PERRIN, N. 2000. *Biomapper*. Laboratory for Conservation Biology, University of Lausanne, Lausanne.
- HOBBS, N.T. & HANLEY, T.A. 1990. Habitat evaluation: do use/availability data reflect carrying capacity? *Journal of Wildlife Management* 54: 515-522.
- HOBBS, N.T. & SWIFT, D.M. 1985. Estimates of habitat carrying capacity incorporating explicit nutritional constraints. *Journal of Wildlife Management* 49: 814-822.
- KOPLIN, J.R. 1973. Differential habitat use by sexes of American kestrels wintering in northern California. *Raptor Research* 7: 39-42.
- LAVERS, C.P. & HAINES-YOUNG, R.H. 1996. Using models of bird abundance to predict the impact of current land-use and conservation policies in the Flow Country of Caithness and Sutherland, northern Scotland. *Biological Conservation* 75: 71-77.
- LAWTON, J.H. 1993. Range, population abundance and conservation. *Trends in Ecology and Evolution* 8: 409-413.
- LAWTON, J.H. & WOODROFFE, G.L. 1991. Habitat and distribution of water voles: why are there gaps in a species' range? *Journal of Animal Ecology* 60: 79-91.
- LEGENDRE, P. & TROUSSELLIER, M. 1988. Aquatic heterotrophic bacteria: modelling in the presence of spatial autocorrelation. *Limnology and Oceanography* 33: 1055-1067.
- LEK, S., DELACOSTE, M., BARAN, P., DIMOPOULOS, I., LAUGA, J. & AULANIER, S. 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling* 90: 39-52.

- LENNON, J.J. 1999. Resource selection functions: taking space seriously? *Trends in Ecology and Evolution* 14: 399-400.
- MAC NALLY, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between -and reconciliation of- 'predictive' and 'explanatory' models. *Biodiversity and Conservation* 9: 655-671.
- MANEL, S., DIAS, J.M. & ORMEROD, S.J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120: 337-347.
- MANEL, S., DIAS, J.M., BUCKTON, S.T. & ORMEROD, S.J. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology* 36: 734-747.
- MANLY, B., McDONALD, L. & THOMAS, D. 1993. Resource selection by animals. Statistical design and analysis for field studies, Chapman & Hall, Londres.
- MARTÍNEZ PALAO, M., GIMÉNEZ, A., MARTÍNEZ, J., ESTEVE, M.A., ANADÓN, J.D. & PÉREZ, I. 2000. Construcción y validación de un modelo de distribución de tortuga mora (*Testudo graeca graeca*) en la región de Murcia basado en su respuesta a factores climáticos. En: Actas del I Congreso Ibérico de Ecología, Univ. de Santiago de Compostela-AEET-SPECO (25-28/09/2000), pp. 46.
- MASTRORILLO, S., LEK, S., DAUBA, F. & BELAUD, A. 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology* 38: 237-246.
- MAURER, B.A. 1994. Geographical population analysis: tools for the analysis of biodiversity, Blackwell Scientific Publications, Oxford.
- MCCLEAN, S., RUMBLE, M.A., KING, R.D. & BAKER, W.L. 1998. Evaluation of resource selection methods with different definitions of availability. *Journal of Wildlife Management* 62: 793-801.
- MILLER, K.V. & CONROY, M.J. 1990. Spot satellite imagery for mapping Kirtland's warbler wintering habitat in the Bahamas. *Wildlife Society Bulletin* 18: 252-257.
- MLADENOFF, D.J., SICKLEY, T.A., HAIGHT, R. G. & WYDEVEN, A.P. 1995. A regional landscape analysis and prediction of favorable Gray Wolf habitat in the Northern Great Lakes region. *Conservation Biology* 9: 279-294.
- MORRISON, M.L., MARCOT, B.G. & MANNAN, R.W. 1998. Wildlife-habitat relationships. Concepts and applications, The University of Wisconsin Press, Madison.
- MORRISON, M.L. & HALL, L.S. 1999. Standard terminology: toward a common language to advance ecological understanding and applications.
- NICHOLLS, A.O. 1989. How to make biological surveys go further with generalised linear models. *Biological Conservation* 50: 51-75.
- ORESQUES, N., SHRADER-FRECHETTE, K. & BELITZ, K. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263: 641-646.
- OSBORNE, P.E. & TIGAR, B.J. 1992. Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa. *Journal of Applied Ecology* 29: 55-62.
- PALMEIRIM, J.M. 1988. Automatic mapping of avian species habitat using satellite imagery. *Oikos* 52: 59-68.
- PARUELO, J.M. & GOLLUSCIO, R.A. 1994. Range assessment using remote sensing in Northwest Patagonia (Argentina). *Journal of Range Management* 47: 498-502.

- PREISLER, H.K. 1993. Modelling spatial patterns of trees attacked by bark beetles. *Applied Statistics* 42: 501-514.
- PRENDERGAST, J.R. 1997. Species richness covariance in higher taxa: empirical test of the biodiversity indicator concept. *Ecography* 20: 210-216.
- SÁNCHEZ-ZAPATA, J.A. & CALVO, J.F. 1999. Raptor distribution in relation to landscape composition in semi-arid Mediterranean habitats. *Journal of Applied Ecology* 36: 254-262.
- SCOTT, J.M., DAVIS, F., CSUTI, B., NOSS, R., BUTTERFIELD, B., GROVES, C., ANDERSON, H., CAICCO, S., D'ERCHIA, F., EDWARDS, J., T.C., ULLIMAN, J. & WRIGHT, R.G. 1993. GAP analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs* 123: 1-41.
- SCHWARZ, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461-464.
- SHORT, H.L. & HESTBECK, J.B. 1995. National biotic resource inventories and GAP analysis: problems of scale and unproven assumptions limit a national program. *Bioscience*, 45: 535-539.
- SKOV, F. & BORCHSENIUS, F. 1997. Predicting plant species distribution patterns using simple climatic parameters: a case study of Ecuadorian palms. *Ecography*, 20: 347-355.
- SMALLWOOD, J.A. 1987. Sexual segregation by habitat in American kestrels wintering in southcentral Florida-vegetative structure and responses to differential prey availability. *The Condor*, 89: 842-849.
- SMITH, P.A. 1994. Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters*, 4: 47-61.
- STATSOFT. 1999. *Electronic Statistics Textbook*, StatSoft, Tulsa, OK.
- SUÁREZ, S., BALBONTÍN, J. & FERRER, M. 2000. Nesting habitat selection by booted eagles *Hieraaetus pennatus* and implications for management. *Journal of Applied Ecology* 37: 215-223.
- THOMAS, Y. & NEIL, M. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science*, 2: 587-602.
- TONER, M. & KEDDY, P. 1997. River hydrology and riparian wetlands: a predictive model for ecological assembly. *Ecological Applications* 7: 236-246.
- TURNER, M.G., ARTHAUD, G.J., ENGSTROM, R.T., HEJL, S.J., LIU, J., LOEB, S. & MCKELVEY, K. 1995. Usefulness of spatially explicit population models in land management. *Ecological Applications* 5: 12-16.
- VAN HORNE, B. 1983. Density as a misleading indicator of habitat quality. *Journal of Wildlife Management* 47: 893-901.
- WILLIAMS, P.H. & GASTON, K.J. 1994. Measuring more of biodiversity: can higher-taxon richness predict wholesale species richness? *Biological Conservation* 67: 211-217.
- WILLIAMS, P., GIBBONS, D., MARGULES, C., REBELO, A., HUMPHRIES, C. & PRESSEY, R. 1996. A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. *Conservation Biology* 10: 155-174.
- WILSON, S.F., SHACKLETON, D. M. & CAMPBELL, K.L. 1998. Making habitat-availability estimates spatially explicit. *Wildlife Society Bulletin* 26: 626-631.
- WU, H. & HUFFER, F.W. 1997. Modelling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics* 4: 49-64.
- ZWEIG, M.H. & CAMPBELL, G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39: 561-577.