

UNA METODOLOGÍA PARA PREDECIR LA DISTRIBUCIÓN ESPACIAL DE LA DIVERSIDAD BIOLÓGICA

JOAQUÍN HORTAL¹ Y JORGE M. LOBO²

RESUMEN

Actualmente es necesario diseñar estrategias de conservación basadas en el conocimiento de la distribución de la diversidad biológica. Desgraciadamente, esta información es escasa y sesgada debido a que desconocemos, parcial o totalmente, la composición faunística de muchas regiones y grupos de organismos, siendo necesario realizar estimaciones fiables de los principales atributos que caracterizan la biodiversidad (riqueza de especies, rareza, etc). Recientemente se han propuesto diversas metodologías para elaborar modelos predictivos capaces de estimar la distribución de estos atributos. Sin embargo, la utilidad de estas aproximaciones se ve cuestionada cuando se desconoce la calidad de los datos de partida y no se realiza una elección previa de las unidades espaciales con inventarios bien establecidos.

En este trabajo se analizan algunos de los problemas que aparecen cuando se pretende modelizar la distribución de los atributos de la biodiversidad, proponiéndose una metodología heurística iterativa que trata de solventarlos utilizando Modelos Lineales Generalizados y una serie de variables ambientales y espaciales. La utilización de esta metodología se ejemplifica mediante la elaboración de un modelo capaz de predecir la riqueza de especies en la Península Ibérica de un grupo de insectos (Coleoptera, Scarabaeinae).

Palabras clave: Pronóstico de atributos de la biodiversidad, Modelos Lineales Generalizados, modelización de la distribución espacial, búsqueda heurística, curvas de acumulación, Scarabaeinae, Península Ibérica

SUMMARY

A methodology to predict the spatial distribution of biological diversity.

Nowadays it is necessary to design conservation strategies based in the knowledge of distribution of biological diversity. Unfortunately, this information is scarce and biased because we ignore, even a part or the total, the faunistic composition of many regions and groups of organisms. Hence, it is necessary to carry out reliable estimates of the main attributes that characterize biodiversity (species richness, rarity, etc). Several methodologies have been proposed recently to elaborate predictive models able to estimate the distribution of these attributes. However, the utility of this approaches must be questioned if the quality of the data used is unknown and a previous selection of the space units with well-established inventories is not carried out.

Museo Nacional de Ciencias Naturales (C.S.I.C.). Departamento de Biodiversidad y Biología Evolutiva. C/José Gutiérrez Abascal, 2 - 28006 Madrid. e-mail: ¹mcnjh521@mncn.csic.es, ²mcnj117@mncn.csic.es

Recibido: 23.04.2002

Aceptado: 19.06.2002

In this paper, some of the problems that appear when we try to modelize the attributes of biodiversity are analysed, and also a heuristic iterative methodology that tries to overcome them using Generalized Linear Models and some environmental and space variables is proposed. The use of this methodology is exemplified by means of the elaboration of a model able to predict the richness of species of a group of insects (Coleoptera, Scarabaeinae) in the Iberian Peninsula.

Key words: Biodiversity attributes forecast, Generalized Linear Models, spatial distribution modelling, heuristic search, species accumulation curves, Scarabaeinae, Iberian Peninsula.

INTRODUCCIÓN

Después de más de 250 años de acumular datos faunísticos y taxonómicos por parte de los científicos dedicados al estudio de la diversidad de la vida y su distribución, aún no existe ninguna localidad en el mundo en la que contemos con el inventario completo de los organismos que habitan en ella. No sólo desconocemos la distribución espacial de la mayoría de las especies, sino que ni siquiera sabemos su número total aproximado (PURVIS & HECTOR 2000). Desdichadamente, este conocimiento es absolutamente necesario si queremos diseñar políticas efectivas de conservación de la biodiversidad (MILLER 1994). Para incorporar la biodiversidad a las estrategias territoriales de conservación, se ha sugerido modelizar la distribución espacial de sus atributos como la alternativa más racional y rápida (SCOTT 1998; LOBO 2000).

Desde comienzos de los años 90, la rápida aparición de nuevas técnicas y herramientas, como los ordenadores personales, los sistemas de Bases de Datos, los Sistemas de Información Geográfica (SIG), la Geoestadística o Geografía Cuantitativa y los paquetes de análisis estadístico, han dotado a los científicos del medio ambiente de gran poder de computación y potentes herramientas de análisis espacial, facilitando el tratamiento de información ambiental georeferenciada de alta calidad (JOHNSTON 1998) y la elaboración de modelos predictivos. La utilización de estas nuevas técnicas aumenta las posibilidades de estimar el valor de diversos atributos relacionados con la biodiversidad en las

áreas insuficientemente conocidas. Aunque en la actualidad existen diversas aproximaciones metodológicas que permiten elaborar modelos capaces de predecir la distribución de los organismos (ver por ejemplo GUIBAN & ZIMMERMANN 2000; ALLEN *et al.* 2001), estimar la distribución geográfica de los atributos generales relacionados con la diversidad biológica (riqueza de especies, rareza, diversidad filogenética, etc.) requiere solventar una serie de inconvenientes.

En este artículo revisamos brevemente las metodologías de modelización que han sido utilizadas hasta la fecha, y proponemos una estrategia capaz de predecir la distribución espacial de los atributos que sintetizan nuestro conocimiento sobre la variación espacial de la biodiversidad. Analizamos, además, los principales inconvenientes que deben evitarse al modelizar estos atributos, presentando un procedimiento iterativo diseñado para solucionar algunos de ellos. Dicho procedimiento utiliza una técnica heurística basada en el uso de Modelos Lineales Generalizados por pasos (Stepwise Generalized Linear Models; CRAWLEY 1993; DOBSON 1999), incluyendo la estructura espacial de la variable a modelizar dentro de las variables explicativas. Como ejemplo práctico, desarrollamos un modelo para predecir la riqueza de especies de una familia de coleópteros coprófagos (Coleoptera, Scarabaeinae; MARTÍN-PIERA 2000) en la Península Ibérica (LOBO & MARTÍN-PIERA 2002). Finalmente, discutimos las perspectivas y potencialidades actuales de estas técnicas para el conocimiento de los patrones de la distribución actual de la biodiversidad, así como para su conservación.

MODELOS PREDICTIVOS Y BIODIVERSIDAD

Modelizar consiste en seleccionar una serie de descriptores capaces de reproducir adecuadamente el fenómeno o la variable que se quiere estudiar (PICKETT *et al.* 1994). En unos casos los modelos tratan de reconocer los factores causales que explican el comportamiento de la variable de interés (por ejemplo, McCARTHY *et al.* 2001) y en otros, el propósito es encontrar una serie de predictores capaces de representar los cambios en esa misma variable, sin asumir ninguna relación causal y sin poder extrapolar las predicciones realizadas más allá del rango de las variables explicativas utilizadas (LEGENDRE & LEGENDRE 1998). Aunque sea delicado y se requiera información adicional de tipo experimental para certificar que la relación entre dos variables es, probablemente, de tipo causal (LAWTON 1999; MAC NALLY 2000), ambos tipos de modelizaciones se confunden frecuentemente (ver GUISAN & ZIMMERMANN 2000).

Existe una numerosísima literatura que ha estudiado las variables relacionadas con la variación de algunos de los atributos de la biodiversidad como la riqueza de especies o la rareza. Este ha sido, probablemente, uno de los temas centrales en Ecología (HUSTON 1994), deduciéndose de esas relaciones posibles mecanismos capaces de controlar la distribución de la diversidad biológica. Desgraciadamente, la relación de los diversos atributos de la biodiversidad con el ambiente dista mucho de ser simple y directa. La distribución actual de la biodiversidad está condicionada por la especificidad de respuesta al medio de cada organismo (HENGEVELD 1997), así como por diversos factores contingentes, únicos e irrepetibles. Es por ello que los modelos que tienen por objeto discernir las causas de la distribución de cualquier organismo, han de poseer, forzosamente, un carácter exploratorio y provisional, debiéndose acudir a técnicas capaces de estimar el porcentaje de variación de cada variable explicativa sin el concurso de las restantes (ver BIRKS 1996; MAC NALLY 2000 o

LOBO *et al.* 2001). Sin embargo, sí es posible crear modelos estáticos (*sensu* GUISAN & ZIMMERMANN 2000), en los que no se trata de estimar la relación funcional de las variables descriptivas que utilizamos (no nos importa su relación causal), sino tan sólo de encontrar una ecuación que nos permita extrapolar la distribución espacial actual de esos atributos a partir de las zonas en los que estos son bien conocidos, creando de este modo un mapa que pronostica su distribución en todo el territorio estudiado.

Según el tipo de datos de partida (solamente presencias, presencias/ausencias o variables continuas) se han utilizado diversas aproximaciones para modelizar la distribución de los organismos en ausencia de datos exhaustivos. Es posible modelizar el nicho potencial de una especie utilizando únicamente los datos sobre las presencias (como en el caso de los atlas de distribución clásicos). La aproximación más básica en este caso es la desarrollada por el proyecto "Gap Analysis". En ella se trabaja a partir de mapas de hábitat, normalmente basados en la vegetación y de la adecuación de la especie cuya distribución se modeliza a cada uno de esos hábitats o asociaciones vegetales (SCOTT *et al.* 1993; SCOTT & JENNINGS 1997; ver <http://www.gap.uidaho.edu/gap>; Lámina 2). La metodología básica del "Gap Analysis" ha sido refinada mediante la comparación de la distribución en el espectro multivariante ambiental de las presencias constatadas frente a la de un conjunto asignando al azar, identificando de este modo el nicho potencial de cada especie (BUSBY 1991; WALKER & COCKS 1991; CARPENTER *et al.* 1993; MITCHELL 1991; STOCKWELL & PETERS 1999; PETERSON *et al.* 1999; HIRZEL 2001). A partir de datos binomiales (presencia/ausencia; 1/0) sobre la distribución de especies concretas también es posible buscar una ecuación capaz de predecir su distribución. Las modelizaciones en este último caso han utilizado muy diversos métodos (GUISAN & ZIMMERMANN 2000) como el análisis discriminante (DENNIS & EALES 1999), los modelos lineares generalizados y los modelos lineares generalizados aditivos, princi-

palmente bajo la forma de regresiones logísticas (PEREIRA & ITAMI 1991; OSBORNE & TIGAR 1992; BUCKLAND & ELSTON 1993; PARKER 1999; AUGUSTIN *et al.* 2001; LOYN *et al.* 2001), los árboles de clasificación y regresión (IVERSON & PRASAD 1998a) o las redes neuronales (MANEL *et al.* 1999a y 1999b) y los algoritmos genéticos (PETERSON *et al.*, 1999). Los pocos estudios existentes sobre la efectividad comparada de estas técnicas no permiten deducir grandes diferencias en la fiabilidad de los modelos obtenidos (MANEL *et al.* 1999a y 1999b; PEARCE & FERRIER 2000; VAYSSIÈRES *et al.* 2000; HIRZEL *et al.* 2001).

Utilizar tanto los datos de presencia como los de presencia-ausencia para predecir la distribución de los organismos requiere que exista información de calidad a lo largo de todo el espectro ambiental en el que habita la especie cuya distribución quiere conocerse. La falta de datos fiables sobre las áreas donde la especie está presente y la atribución incorrecta de ausencias, cuando lo que de verdad existe es carencia de información, produce inevitablemente estimas imprecisas. Ese es un inconveniente principal cuya solución pasa por analizar los datos de base con los que se pretende trabajar y mejorar su calidad cuando sea necesario (ver BEARD *et al.* 1999 o GUISAN & ZIMMERMANN 2000). Otro inconveniente es que los modelos producidos asignan una probabilidad de aparición, que es una variable continua entre 0 y 1, mientras que los datos con los que se evalúa el modelo (generalmente algunas localidades no utilizadas en la elaboración del mismo) indican únicamente ausencia (0) o presencia (1), lo que significa elegir un umbral a partir del cual la probabilidad de aparición pueda clasificarse como presencia cierta (MANEL *et al.* 2001). Sin embargo, el mayor inconveniente a la hora de estimar la distribución de la biodiversidad a partir de estas modelizaciones, es que éstas tan sólo nos permiten predecir la distribución de las especies una a una. De este modo, para obtener predicciones de atributos como la riqueza de especies, la rareza o la endemidad, es necesario sumar las distribuciones individuales predichas. Aunque

pueden obtenerse mapas predictivos no muy imprecisos disponiendo únicamente de diez observaciones (STOCKWELL & PETERSON 2002), utilizar estas aproximaciones significa excluir todas aquellas especies con escasa o nula información que, en algunas localidades, pueden ser responsables de un porcentaje importante de la diversidad total. Aunque una parte de las especies raras que pueden colectarse en una localidad pueden considerarse "vagabundas" y son incapaces de reproducirse en el lugar donde han sido colectadas (DENNIS 2001), son muchas las causas de la rareza (RABINOWITZ *et al.* 1986) y poseer poblaciones con pocos individuos o circunscritas a unas pocas localidades es, a menudo, una característica intrínseca de algunas especies que se comportan como metapoblaciones (HANSKI 1999). Como la rareza o la endemidad no se distribuyen aleatoria u homogéneamente en un territorio y como la riqueza de especies de una localidad suele ser altamente dependiente de la cantidad de especies raras (GASTON 1994), no considerar las especies raras significa homogeneizar las diferencias reales entre localidades en la diversidad de especies, disminuyendo la utilidad en conservación de los mapas predictivos de la diversidad biológica de cualquier región.

Otra característica de los mapas de distribución potencial generados con las modelizaciones individuales de cada especie es que, frecuentemente, los modelos sugieren áreas aptas para la supervivencia de la especie en las que esta no aparece, bien por causas históricas, bien porque el tamaño de esas áreas es insuficiente para permitir la supervivencia de las poblaciones (LEATHWICK 1998; LOBO 2000; TEIXERA *et al.* 2001; ALLEN *et al.* 2001). En estos casos, la simple superposición de los mapas predictivos individuales puede tender a incrementar los valores de la diversidad biológica en determinadas partes del territorio. Es por ello que, si queremos conocer la distribución geográfica de diversos atributos que sinteticen la diversidad biológica de cualquier tipo de taxón en ausencia de información exhaustiva, la opción más conveniente consiste en modelizar esos mis-

mos atributos, del mismo modo que se modeliza la distribución individual de cada especie o las variables paisajísticas reconocibles a una gran escala espacial (MOISEN & EDWARDS 1999; FRESCINO *et al.* 2001.). A pesar de la frecuente utilización de las diversas técnicas mencionadas al objeto de estimar la distribución de especies concretas, la elaboración de modelos para conocer la distribución de la riqueza o la rareza de especies con fines conservacionistas ha sido poco aprovechada (MARGULES *et al.* 1987; BOJORQUEZ-TAPIA 1996; IVERSON & PRASAD 1998a y b; ZIMMERMANN & KIENAST 1999; HORTAL *et al.* 2001; LOBO & MARTÍN-PIERA 2002, LOBO *et al.* 2002).

En este trabajo proponemos una metodología diseñada para obtener modelos estáticos (sin una hipótesis funcional asociada), que nos permitan obtener estimas fiables de atributos de la biodiversidad en forma de mapas. Para ello, vamos a utilizar un conjunto de variables explicativas, sin importarnos si su relación con el atributo estudiado es causal o no. Este último punto es especialmente importante, pues facilita utilizar como predictores diversas variables que, aunque no estén funcionalmente relacionadas con la distribución de la variable a conocer, pueden mejorar la capacidad predictiva del modelo. La construcción de estos modelos estáticos implica tres fases secuenciales:

- i) *Recopilación y análisis de la información*, en la que se compila y almacena toda la información disponible tanto de la variable a modelizar, como de las variables predictoras, se definen las unidades territoriales de trabajo (escala de trabajo) y se determina para cuáles de ellas la información es fiable
- ii) *Establecimiento del modelo*, en la que se seleccionan las variables explicativas, se estudia la relación de estas con el atributo de biodiversidad escogido y se genera una función o modelo capaz de explicar la variable dependiente.
- iii) *Evaluación del modelo*, en la que se determina la fiabilidad de los resultados obtenidos.

CONSTRUCCIÓN DE MODELOS PREDICTIVOS

Recopilación y análisis de la información

Una vez establecido el grupo taxonómico sobre el que vamos a trabajar y definida el área geográfica de interés (el extent sensu WHITTAKER *et al.* 2001), hemos de proceder a recopilar toda la información existente que sea relevante sobre ese taxón en la zona elegida. Las fuentes de información, tanto faunística como ambiental, son muy heterogéneas en su origen y calidad, encontrándose dispersas en multitud de publicaciones, instituciones y, desde hace varios años, páginas web. Debido a ello, es necesario procesar la información recopilada para determinar la escala espacial de trabajo y encontrar la información faunística y ambiental fiable y relevante a dicha escala.

Obtención y compilación de datos biológicos

Tras las experiencias pioneras de países como el Reino Unido (LAWTON *et al.* 1994; GRIFFITHS *et al.* 1999), y ante la inmediatez de la denominada "Crisis de la Biodiversidad", la creación de grandes bancos de datos sobre la distribución de la biota está recibiendo un importante impulso internacional, mediante la creación del Banco de Datos Global sobre Biodiversidad, denominado GBIF (Global Biodiversity Information Facility; ver <http://www.gbif.org> y EDWARDS *et al.* 2000a). Sin embargo, este proceso es lento y complejo. En la gran mayoría de los países, incluida España, esta información aún ha de ser recopilada debiéndose, en primer lugar, definir un diseño de base de datos cuyos campos incluyan información estandarizada y consensuada (BISBY 2000a y b; EDWARDS *et al.* 2000a y b; SMITH *et al.* 2000; RAMOS *et al.* 2001).

Cuando se trata de recopilar la información taxonómica y faunística disponible sobre cualquier grupo, esta resulta ser casi siempre escasa, heterogénea, y dispersa en multitud de fuentes como publicaciones científicas, museos de historia natural, colecciones privadas, etc. Para utilizar esta información, es necesario desarrollar una base de datos capaz de compilarla de

manera exhaustiva, de modo que incluya los datos procedentes de todas las fuentes disponibles para el grupo estudiado en el territorio de interés. La estructura de las bases de datos diseñadas para compilar la información sobre la distribución de las especies es bastante compleja (ver por ejemplo ATLANTIS, en <http://www.gobcan.es/medioambiente/biodiversidad/ceplam/bancodatos/atlantid.html>, BIOTICA, en http://www.conabio.gob.mx/-biotica/acerca_biotica.html o BIODI, en <http://biodi.sdsc.edu/>). Cuando no sea posible elaborar una base de datos en un formato tan acabado, puede resultar igual de útil confeccionar una más sencilla que incluya, al menos, los siguientes campos (Figura 1): fecha de captura u observación, lugar (y también las coordena-

das espaciales en un sistema de referencia de uso general, como coordenadas geográficas (lat/long) o Universal Transverse Mercator (UTM)), datos ecológicos de la captura relevantes para el grupo estudiado (tipo de hábitat, alimentación, altitud, especie hospedadora, etc.), número y sexo, si es posible, de los ejemplares capturados u observados, método de captura u observación, identidad del colector u observador, responsable de la identificación del taxón, lugar de almacenamiento (para especímenes procedentes de colecciones de Historia Natural) o referencia bibliográfica, y otros datos de utilidad, como secuencias genéticas, morfotipo, etc. El núcleo central de una base de datos como esta deben ser los especímenes, de modo que cada registro informático contenga toda aquella

Base de datos de Scarabaeoidea de Madrid					
Familia	Aphodidae		Género	Aphodius (Aphodius)	
Especie	loetoides				
Provincia	Madrid	Municipio	Colmenar Viejo	Topónimo	Colmenar Viejo
Altitud localidad	883	altura captura			
UTM completa	30TVL30	UTM 10 Km	30TVL30	Longitud	435300
				Latitud	4501243
				CUTM Propia	
día	30	mes	11	año	80
				ejemplares	4
				sexo	M
Habitat		método captura		alimentación	HEZ. vacuno
colección	Veiga, C.M.			Muestreo	
leg	Veiga, C.M.	del	Veiga, C.M.	serie típica	
autor artículo	Material estudiado		año artículo	Trampa	
título artículo					
revista					

Figura 1: Registro procedente de la base de datos biológica SCAMAD, que recoge la información disponible sobre la distribución y fenología de los coleópteros coprófagos (escarabajos; Col., Scarabaeoidea) de la Comunidad de Madrid.
Figure 1: Record from the biologic data-base SCAMAD, that compiles the available distribution and phenology information for Comunidad de Madrid dung beetles (scarabeids; Col., Scarabaeoidea).

información biológica, ecológica y geográfica de aquellos individuos capturados en idénticas condiciones, fecha y lugar.

Elaboración y almacenamiento de la información ambiental

En la actualidad, existe una gran cantidad de información ambiental geográficamente referenciada. Muchos países e instituciones transnacionales cuentan ya con grandes bancos de datos procesados, en los que se almacena de manera centralizada la información ambiental generada en diferentes proyectos, así como su origen y metodología de obtención (ver GARCÍA HERNÁNDEZ & BOSQUE SENDRA 2001). En otros casos, esta información está igualmente disponible, aunque no centralizada, por lo que es necesario un trabajo previo de recopilación y tratamiento de la información para incluirla en un Sistema de Información Geográfica (SIG; MILLER 1994). En España, este trabajo está aún en proceso de realización (actualmente se lleva a cabo en el Banco de Datos de la Naturaleza, dependiente de la Dirección General de Conservación de la Naturaleza; ver <http://www.mma.es>).

A partir de los datos accesibles, hemos de crear una Base de Datos territorial. La información mínima incluida en ella depende de los objetivos a conseguir. Si el objetivo es crear un banco de datos ambiental polivalente, esta información ha de incluir, al menos:

- Datos medioambientales como: i) datos climáticos, incluyendo variables comunes como temperatura media o precipitación anual y otras variables que puedan ser importantes para el grupo estudiado, tales como evapotranspiración potencial, precipitación estival o número de días con heladas; ii) datos topográficos, como un modelo digital del terreno (MDT; mapa de alturas) y variables derivadas de él (pendientes, orientaciones, cuencas de drenaje, etc.) o datos relativos al sustrato, como geología, tipo y composición del suelo e hidrología.
- Mapas de cobertura y usos del suelo que aporten información sobre la situación actual del medio.

- Otra información relevante, como variables que puedan producir impactos negativos, variables socioeconómicas, distribución de especies importantes para el grupo, etc.

La información no disponible en formato digital ha de ser introducida en el SIG a partir de su versión en papel, digitalizándola mediante un escáner o una tableta digitalizadora y teniendo especial cuidado en el error espacial debido a la resolución espacial del escaneado y/o a los fallos en la digitalización de los elementos vectoriales. Si esta información procede de datos puntuales (por ejemplo, estaciones meteorológicas), es necesario interpolar los valores de esa variable en el conjunto del territorio a partir de ese conjunto fragmentario de datos. Existen diversos procedimientos de interpolación que difieren en su complejidad. La elección de uno u otro dependerá del tipo de datos interpolados (cuantitativos o cualitativos), de la calidad de los datos de partida, y de la resolución de salida que deseemos. Para una revisión de estos métodos, ver BOSQUE SENDRA (1997) y JOHNSTON (1998).

Tras recopilar toda la información necesaria en forma de mapas digitales (llamados capas o coberturas), esta debe ser encajada en la misma base cartográfica, es decir, en una cobertura espacial con la misma extensión, resolución espacial, situación y sistema de referencia (por ejemplo, UTM o Latitud/Longitud). Los métodos necesarios en este proceso final son sencillos de utilizar y se encuentran implementados en todos los programas SIG existentes en la actualidad.

Elección de la unidad territorial de trabajo

Una vez que toda la información corológica y ambiental disponible ha sido recopilada, es necesario definir las unidades territoriales sobre las que vamos a realizar el modelo y, por tanto, la resolución espacial del mismo. Estas unidades han de ser las de menor área posible teniendo en cuenta: i) la disponibilidad y precisión de la información biológica, ii) la resolución espacial de la información ambiental disponible, y iii) la extensión total de la región a

estudiar (sensu WHITTAKER *et al.* 2001). Normalmente, la información ambiental no suele ser un problema ya que es posible obtenerla a una resolución inferior a la de los datos corológicos. Una condición ineludible para la consecución de cualquier modelo predictivo mediante la utilización de variables ambientales explicativas, es que los lugares de partida con datos biológicos fiables que vayamos a utilizar para construir el modelo, recojan la variabilidad ambiental del territorio a la escala considerada. Aumentar la extensión del territorio a estudiar puede ocasionar que no existan suficientes unidades espaciales con información biológica relevante que, a la vez, constituyan una buena representación de su heterogeneidad ambiental. Por otra parte, disminuir la escala de trabajo implica estudiar procesos a escalas menores, aumentando la complejidad ambiental percibida, por lo que exige utilizar un mayor número de lugares con datos biológicos fiables para recoger toda esa variación ambiental. Por el contrario, utilizar unidades espaciales relativamente grandes puede impedir elaborar modelos con utilidad práctica, especialmente en territorios relativamente pequeños. Un modelo generado a una resolución de 100 x 100 km² para toda Europa puede ser de gran utilidad, mientras que si se realiza para Cataluña su utilidad es nula.

Por ello, dependiendo de la calidad de la información biológica de partida, es necesario establecer un compromiso entre la extensión del territorio a estudiar y la escala espacial del análisis. En este contexto, es necesario recordar que los patrones de la distribución espacial de la diversidad biológica y los procesos que los generan difieren entre regiones y escalas (LEVIN 1992; LAWTON 1999), de modo que los modelos elaborados a una resolución y en un territorio no deben aplicarse a otros lugares y escalas.

Es también recomendable escoger unidades espaciales que mantengan superficies y formas similares o comparables dentro de la región estudiada. Las mallas de referencia comunes, como las basadas en coordenadas geográficas, o las derivadas del sistema Universal

Transverse Mercator (por ejemplo, malla UTM de 50 Km; Figura 2a), son una buena opción. No obstante, es posible realizar modelizaciones utilizando unidades territoriales irregulares en forma y tamaño (Figura 2b), como manchas de vegetación o divisiones administrativas (ver por ejemplo WOHLGEMUTH 1998), aunque sus resultados son difícilmente utilizables para realizar mapas interpolados.

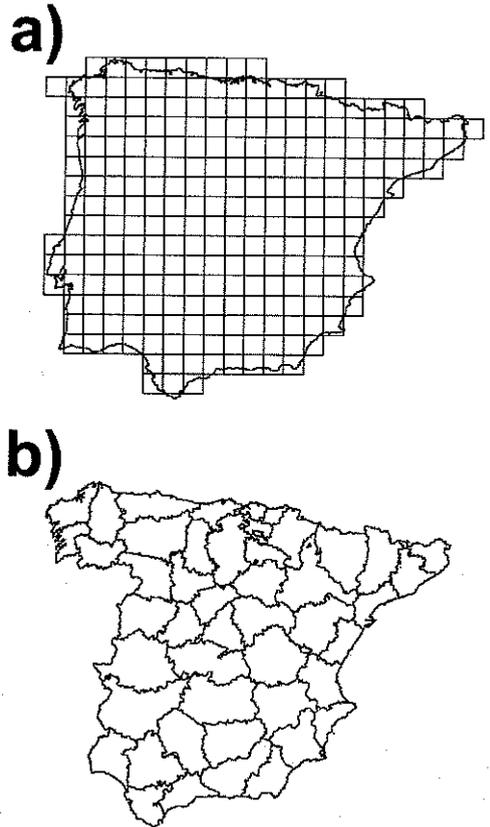


Figura 2: Tipos de mallas de referencia utilizados para designar unidades territoriales (ver texto): a) Métrica (Cuadrículas UTM de 50km); b) Irregular (provincias).

Figure 2: Reference grid types used to designate territorial units (see text): a) Metric (50km UTM squares); b) Irregular (provinces).

Selección de las unidades fiables

Una vez recopilada toda la información corológica disponible y precisados, tanto el territorio

como las unidades espaciales sobre las que vamos a trabajar, es necesario determinar a partir de cuáles de esas unidades vamos a realizar el proceso de modelización. Para ello, en primer lugar, hemos de identificar aquellas unidades en las que los inventarios sean lo suficientemente completos como para obtener medidas fiables del atributo de biodiversidad que queremos modelizar (SOBERÓN *et al.* 2000). Una buena solución a este problema puede ser la utilización de curvas de colecta (SOBERÓN & LLORENTE 1993; COLWELL & CODDINGTON 1994; LEÓN-CORTÉS *et al.* 1998; MORENO & HALFFTER 2000; GOTELLI & COLWELL 2001; Figura 3), las cuales relacionan la incorporación de nuevas especies al inventario con el incremento de alguna medida del esfuerzo de muestreo. A pesar del debate actual sobre su utilidad para predecir la riqueza total de especies en un territorio mediante la extrapolación de la función obtenida hasta alcanzar una asíntota (ver por ejemplo COLWELL & CODDINGTON 1994; GOTELLI & COLWELL 2001; WILLOT 2001; MORENO & HALFFTER 2001), sí que resultan de gran utilidad para determinar el grado de saturación del

inventario de especies en un área determinada y, por tanto, para identificar aquellas localidades cuyos inventarios biológicos pueden ser considerados razonablemente completos. Utilizando esta técnica podemos, además, estimar la intensidad de esfuerzo complementario necesario para obtener un inventario fiable.

Las curvas de colecta necesitan de una unidad de esfuerzo de muestreo que sea espacial y temporalmente homogénea y comparable. En ocasiones, el esfuerzo puede estandarizarse fácilmente considerando unidades como horas/persona o trampas/día. Sin embargo, a partir de la información corológica que puede extraerse de una base de datos como la que hemos descrito, en la que se incluye información heterogénea recolectada con propósitos distintos, no suele ser posible determinar con precisión el esfuerzo realizado, siendo difícil establecer una unidad. Por eso, es necesario encontrar una medida que sea, a la vez, fácilmente extraíble de la base de datos, tan homogénea como sea posible, capaz de ser aplicada a toda la información recopilada y un buen sustituto, cuando no una medida directa, del esfuerzo de muestreo real.

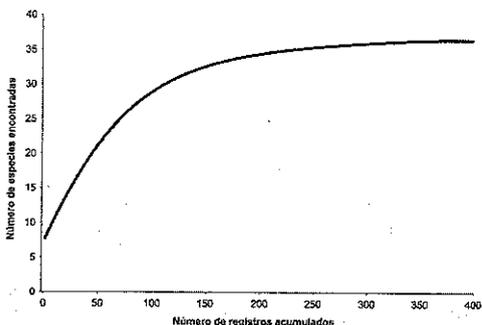


Figura 3: Curva de colecta, en la que se relaciona la cantidad de especies que se van añadiendo al inventario según aumenta el muestreo efectuado. El valor de la asíntota se corresponde con el número total de especies presentes en la zona. La curva es el resultado de la aleatorización en el orden de entrada de las unidades de esfuerzo. Los datos han sido Datos extraídos de la base de datos SCAMAD (no publicados).

Figure 3: Collector's curve, where the number of species that are being added to the inventory as the sampling effort increases are related. Asymptote value corresponds to the total number of species present in the area. The curve is the result of randomizing the order of entrance of effort units. Data extracted from SCAMAD database (unpublished).

Si contamos con información acerca de la fecha o los recolectores, esta medida puede ser, por ejemplo, el número de días de colecta por persona o el número de recolectores diferentes en cada unidad espacial. Desgraciadamente en muchas ocasiones no existe información de este tipo sobre los ejemplares y la variabilidad de fechas y/o colectores es muy baja. Asumiendo que la probabilidad de que una especie aparezca en el inventario de una unidad espacial está positivamente correlacionada con el número de registros de ella que contiene la base de datos, una unidad de esfuerzo más interesante es el propio número de registros de cada unidad espacial. En este caso, entendemos como registro el conjunto de especímenes de una única especie con idéntica información en todos los campos de la base de datos (localidad, altitud, fecha de captura, tipo de hábitat, etc.), de modo que cualquier diferencia en un campo de la base de datos significa la inclusión de esta información en un nuevo registro. Es decir, si se han

colectado durante un día 25 individuos correspondientes a la misma especie, estos constituyen un solo registro; si, en cambio, corresponden a 2 especies o a la misma especie, pero han sido colectados en dos días diferentes, dan lugar a 2 registros. Los registros han demostrado ser un buen sustituto del esfuerzo de muestreo, ya que seleccionar localidades mediante curvas de colecta que utilizan medidas directas del esfuerzo de muestreo como el número de trampas o seleccionarlas utilizando el número de registros informáticos produce resultados muy similares (datos no publicados).

La calidad del inventario en cada unidad espacial puede evaluarse determinando una función exponencial negativa que describa la tasa de especies añadidas al inventario (S_r) con el incremento en el esfuerzo, en nuestro caso el número de registros; r (SOBERÓN & LLORENTE 1993); COLWELL & CODDINGTON 1994; FAGAN & KAREIVA 1997; COLWELL 2000). De acuerdo con SOBERÓN & LLORENTE (1993) y COLWELL & CODDINGTON (1994), esta relación se expresa mediante:

$$S_r = S_{\max} [1 - \exp(-br)]$$

donde S_{\max} , la asíntota, es el número total de especies estimado por cuadrícula, y b es una constante ajustada que controla la forma de la curva. Esta función curvilinear se ajusta a los datos, por ejemplo mediante el método de Quasi-Newton (STATSOFT 1998). Debido a que alcanzar el 100% de la riqueza requiere un número infinito de registros, se calcula el número de registros necesario para una tasa de incremento de especies determinada. Una tasa de 0.01 (una especie añadida al inventario cada 100 nuevos registros; $r_{0.01}$) se calcula mediante la siguiente ecuación:

$$r_{0.01} = 1/b \ln(1 + b/0.01)$$

(SOBERÓN & LLORENTE 1993).

Establecimiento del modelo

Un modelo es, básicamente, una función que relaciona la variable dependiente con una serie de variables explicativas. En términos genera-

les, el proceso de realización de un modelo requiere tres pasos básicos (GUISAN & ZIMMERMAN 2000): (i) la formulación del modelo, (ii) la selección de las variables predictivas y (iii) el ajuste del modelo al objeto mejorar el acuerdo entre los datos de base y las predicciones del modelo.

La formulación del modelo consiste, básicamente, en estimar el tipo de distribución de la variable dependiente a modelizar y conocer cual es la relación de esta con los diversos predictores, lo que permitirá elegir un algoritmo capaz de seleccionar adecuadamente las variables predictoras y ofrecer estimas fiables de sus parámetros. Tanto para determinar qué tipo de distribución de frecuencias tiene el atributo a modelizar, como para encontrar una función (lineal, logarítmica, etc.) que relacione adecuadamente a este con los predictores elegidos, es conveniente realizar un análisis exploratorio preliminar y/o fiarse de los tipos de distribución y relaciones mencionados en la literatura para esas variables. Para variables que resumen la diversidad biológica, se asume generalmente que la variable dependiente posee una distribución de Poisson y que la relación entre esta y las variables independientes es logarítmica (CRAWLEY 1993).

Selección de las variables predictivas

La elección de las variables que vamos a utilizar como predictores es siempre delicada. En general, la distribución de los atributos de la biodiversidad puede considerarse influida por dos tipos de factores básicos (RICKLEFS & SCHLUTER 1993): la adecuación de los requerimientos ecológicos de las especies al medio (*factores ecológicos*) y la actuación de determinados eventos únicos e irrepetibles, como la historia evolutiva de los organismos y la historia geológica de las áreas (*factores históricos*).

A pesar de que no buscamos una relación causal entre los predictores y el atributo que estudiamos, realizar una serie de suposiciones previas acerca de los factores con más probabilidades de afectar a la distribución actual, puede ayudar a seleccionar las variables con mayor capacidad explicativa cuando existe un gran

conjunto de posibilidades. Debido a la frecuente correlación entre las variables ambientales, la utilización de un predictor determinado o su inclusión en un orden concreto, puede oscurecer la influencia de otras variables de mayor capacidad predictiva.

Una vez elegidas las variables a utilizar, hemos de extraer de nuestra base de datos ambiental los valores de esas variables para cada una de las unidades espaciales de trabajo. El ambiente de cada unidad territorial se puede subdividir en tres grupos de variables. Por una parte, están las características topográficas, climáticas y edáficas, como altitud media, pendiente, exposición, temperatura, precipitación o litología que constituyen, a menudo, variables abióticas con gran capacidad de predicción debido a su relación directa con las adaptaciones fisiológicas de los organismos. Por otra parte están las características bióticas del medio que, generalmente, son más difíciles de conseguir y poseen una influencia notable sobre la distribución actual de las especies. Entre estas variables se encuentran la estructura y composición de la vegetación, la distribución o abundancia de las especies potencialmente hospedadoras o presas del grupo, así como aquellas variables que describen los usos actuales del suelo y que, por tanto, reflejan el grado de antropización. Por último, cabe destacar a las variables capaces de describir la heterogeneidad ambiental del territorio como las diferencias de altitud o de temperatura y la diversidad de ambientes: Como se presume que una mayor variedad de recursos y condiciones permite la supervivencia de especies con tolerancias y adaptaciones distintas, estas variables pueden ser muy importantes a la hora de explicar la riqueza de especies.

La importancia de estos tipos de variables puede diferir según cual sea el atributo de la diversidad biológica que se quiera modelizar. La productividad o la heterogeneidad ambiental suelen ser buenas variables a la hora de predecir la riqueza de especies (ver por ejemplo WOHLGEMUTH 1998), pero la rareza de especies puede explicarse mejor atendiendo a factores edáficos, geomorfológicos o a la presencia de determinados hábitats singulares o relictos

(GASTON 1994). De este modo, el proceso de selección de las variables predictivas a considerar dependerá de nuestro conocimiento previo sobre los factores de mayor relevancia sobre la distribución de la variable a modelizar.

Sin embargo, la distribución de los organismos puede estar condicionada por la particular historia evolutiva de los taxa y de las áreas en las que habitan. Es decir, por factores de tipo histórico (RICKLEFS & SCHLUTER 1993). Aunque este tipo de factores son muy difíciles de resumir en forma de variables convencionales, existen algunas aproximaciones en este sentido (RICKLEFS & SCHLUTER 1993; BIRKS 1996) que podrían mejorar mucho el ajuste de los modelos. A pesar de la dificultad de incluir los procesos históricos en la modelización de la distribución de las especies, se puede asumir que su actuación debe producir frecuentemente un patrón espacial y, en este caso, incorporar la latitud y la longitud como variables predictoras puede producir excelentes resultados (LEGENDRE & LEGENDRE 1998). Si, además de las variables ambientales, se incorporan en el modelo los nueve términos de una ecuación polinomial de tercer grado de la latitud y la longitud (*Trend Surface Analysis*; $b_1LAT + b_2LON + b_3LAT^2 + b_4LAT \times LON + b_5LON^2 + b_6LAT^3 + b_7LAT \times LON + b_8LAT \times LON^2 + b_9LON^3$), es posible incluir en el modelo aquella variación debida a factores históricos o geográficos que produce un patrón espacial (LOBO & MARTÍN-PIERA 2002).

Existen dos problemas en la selección de las variables a utilizar en un proceso de modelización. Por una parte, las variables ambientales son generalmente colineales o, lo que es igual, sus valores están correlacionados entre sí (Figura 4). Esta colinearidad puede sesgar la estimación de los parámetros del modelo, lo que, como discutimos previamente, puede oscurecer las posibles relaciones causales presentes en los datos. Sin embargo, si nuestra intención es tan sólo realizar un pronóstico, es decir, maximizar la varianza explicada, la colinearidad de las variables explicativas no representa un problema (LEGENDRE & LEGENDRE 1998).

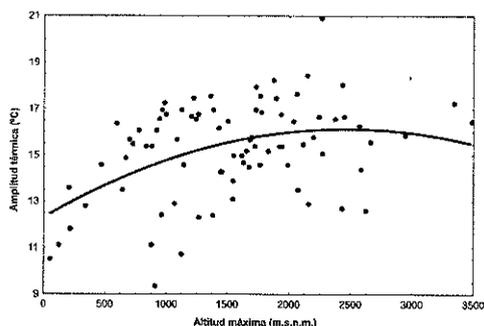


Figura 4: Colinearidad entre dos variables ambientales curvilinealmente relacionadas en la Península Ibérica (ver texto). Datos procedentes de Lobo & Martín-Piera, 2002.

Figure 4: Collinearity among two environmental variables curvilinearly related in the Iberian Peninsula (see text). Data from Lobo & Martín-Piera, 2002.

El segundo problema se refiere a la autocorrelación espacial de las variables biológicas y medioambientales. La autocorrelación en una variable significa que los valores observados en una localidad están influidos por los de las localidades vecinas (Lámina 3). Ello implica una dependencia espacial de las observaciones y supone invalidar la asunción de independencia en la que se basan las herramientas estadísticas clásicas. Como la heterogeneidad espacial en la naturaleza es el resultado de procesos no aleatorios, la autocorrelación espacial es una propiedad intrínseca de las variables ambientales que vamos a utilizar (LEGENDRE & LEGENDRE 1998), por lo que, siguiendo el criterio estadístico de la independencia de las variables estadísticas, nuestro modelo estará inevitablemente sesgado en casi todas las ocasiones. Sin embargo, la autocorrelación espacial es una consecuencia de los procesos que han producido los patrones de distribución espacial de cualquier variable ambiental, por lo que su eliminación disminuye la capacidad predictiva del modelo (SMITH 1994; LEGENDRE & LEGENDRE 1998). Por ello, al desarrollar modelos con datos espacialmente autocorrelacionados el criterio clave debe ser explorar si los errores de la aplicación de la función (los residuos) que resulten del ajuste final del modelo están espacialmente autocorrelacionados. Si los residuos del modelo están autocor-

relacionados es que, al menos, una variable espacialmente estructurada no ha sido incluida en el análisis (CLIFF & ORD 1981; ODLAND 1988). En ese caso debería buscarse dicha variable y lograr un nuevo modelo en el que los residuos no estén espacialmente autocorrelacionados. La inclusión del espacio en el modelo, tal y como comentamos anteriormente, suele ser a menudo una garantía de que se ha incluido el efecto de aquellas variables que tienen una estructura espacial, pero que no han sido consideradas.

Para extraer los valores de las variables por unidad territorial, las propiedades de los SIG son una herramienta muy valiosa, ya que nos permiten relacionar los mapas digitales de cada variable ambiental con el que define las unidades territoriales (Lámina 4). En el caso de las variables continuas, mediante diversas herramientas analíticas implementadas en los SIG, se puede extraer el valor medio de la variable en cada unidad territorial (temperatura máxima, precipitación estival, altura media, pendiente media, etc.) así como sus valores máximo y mínimo (altitud máxima y mínima). Para las variables categóricas resulta más útil extraer la presencia o ausencia de cada categoría, o la proporción de superficie que ocupa en la unidad territorial. La posición (latitud y longitud) de cada unidad territorial se asigna a su centroide (centro de masas o punto más representativo). En el caso de mallas regulares, este se obtiene fácilmente como el centro del polígono que define cada unidad. Para unidades territoriales irregulares, la media de los valores de todos sus vértices produce una posición admisible, excepto en casos extremos, en los que son necesarios procesos más complejos (ver BOSQUE SENDRA 1997).

La heterogeneidad ambiental puede ser representada por la diversidad de diferentes variables, como orientaciones, pendientes, usos del suelo, geología o tipos de formaciones vegetales. Sus valores en cada unidad territorial se pueden estimar mediante el índice de diversidad de Shannon (MAGURRAN 1988):

$$H' = - \sum p_i \cdot \log_2 p_i$$

...donde p_i es la frecuencia relativa de superficie ocupada por cada una de las categorías o rangos de la variable estudiada.

Las variables predictivas deben ser estandarizadas a media cero y varianza uno para eliminar el efecto de las diferencias en la escala de medida de las diferentes variables independientes. El algoritmo utilizado para ello debe ser: valor estandarizado = (valor original - media) / desviación estándar, excepto para el caso de la latitud y la longitud, que, como recomiendan LEGENDRE & LEGENDRE (1998), deben estandarizarse a la media. Es decir, de la siguiente manera: valor estandarizado = valor original - media.

Ajuste del modelo

Una vez obtenidas las variables por unidad territorial, hemos de ajustar los valores del atributo que estudiamos procedentes de las unidades territoriales bien inventariadas a los de las variables ambientales extraídos para esas cuadrículas. El resultado de este ajuste es una función matemática en la que el valor del atributo está determinado por los valores de varios predictores. Un procedimiento ampliamente utilizado para ajustar un modelo consiste en la búsqueda iterativa de las variables a seleccionar. Se trata de un proceso de selección heurístico, es decir, un algoritmo que ajusta secuencialmente predictores a la variable dependiente y que suele producir una buena solución al problema, aunque el resultado generado no tiene porque ser el óptimo. La utilización de los Modelos Lineales Generalizados (GLM; McCULLAGH & NELDER 1989; DOBSON 1999) constituye una excelente opción para modelizar de manera iterativa la variación de un atributo de biodiversidad en función de las variables explicativas ambientales y espaciales más significativas (ver AUSTIN 1980; NICHOLLS 1989; NICHOLLS 1991; TONTERI 1994; AUSTIN *et al.* 1990, AUSTIN *et al.* 1996; HEIKKINEN & NEUVONEN 1997; HORTAL *et al.* 2001; LOBO & MARTÍN-PIERA 2002). Las regresiones lineales o binomiales constituyen un caso más dentro de los GLM en los que la variable dependiente posee un tipo especial de distribución y se asume una relación determi-

nada entre la variable dependiente y las variables explicativas. La utilización de GLM permite, por tanto, realizar cualquier tipo de modelo independientemente de la distribución de la variable a predecir y del tipo de relación que tenga esta con las variables independientes.

De este modo, el modelo se construye tras estudiar el tipo de distribución del atributo a modelizar y elegir una función base que ligue a este con el conjunto de variables predictivas. Como hemos mencionado antes, en el caso de la riqueza de especies, la distribución suele de Poisson y la relación entre ésta y las variables independientes logarítmica. El ajuste de los modelos desarrollados mediante GLM se comprueba examinando la variación en la desviación del modelo sobre un modelo nulo, en el que se introduce la variable a modelizar sin considerar ninguna variable explicativa. La desviación es, así, una medida de la desviación del modelo sobre la variabilidad original de los datos de la variable dependiente y, como en el caso del coeficiente de determinación (R^2), puede expresarse como el porcentaje de variación de la variable dependiente que es capaz de explicar cada modelo. El cambio en la desviación de cada modelo se comprueba mediante test de la F clásico (McCULLAGH & NELDER 1989; DOBSON 1999).

Con el objeto de tener en cuenta las relaciones no-lineares es conveniente, en un primer paso, que se relacione el atributo a modelizar con las funciones lineal, cuadrática y cúbica de cada variable predictiva por separado. De estas tres funciones para cada variable se selecciona aquella que produce una mayor reducción en la desviación a partir de la del modelo nulo y cuyos términos sean significativos (AUSTIN 1980; MARGULES *et al.* 1987; AUSTIN *et al.* 1996). Si la variable dependiente se relaciona significativamente con la función cuadrática de una variable explicativa ($x + x^2$) es que la relación entre estas variables es parabólica (Figura 5a). Si los tres términos de la función cúbica de una variable explicativa son significativos ($x + x^2 + x^3$) es que la relación entre esta y la variable dependiente es curvilinearmente más compleja (Figura 5b). Después es recomendable comen-

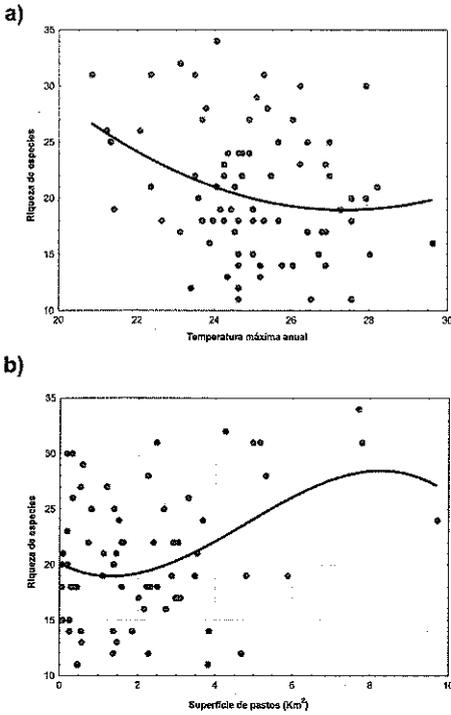


Figura 5: Tipos de relación entre la riqueza de especies y las variables independientes (ver texto): a) Cuadrática o parabólica, en este caso de tipo inverso, con la temperatura máxima anual; b) Cúbica, con la cantidad de superficie de pastos. Datos procedentes de Lobo & Martín-Piera, 2002.

Figure 5: Kinds of relationships between the species richness and the independent variables (see text): a) Quadratic or parabolic, inverse in this case, with maximum annual temperature; b) Cubic, with the amount of grassland area. Data from Lobo & Martín-Piera, 2002.

zar un proceso secuencial en el que se elige primero aquella variable explicativa (bien sea su función lineal, cuadrática o cúbica) que sea responsable del cambio en la desviación más importante. Posteriormente, todas las variables restantes se van añadiendo una a una al modelo (forward stepwise selection), comprobando cada vez la significación en el cambio en la desviación, eligiéndose de nuevo aquella que produce el cambio más significativo en la desviación. Tras cada inclusión significativa, el nuevo modelo se somete a un proceso de selección por pasos hacia atrás (backward stepwise selection), con el objeto de eliminar aquellos

términos que hayan caído por debajo de la significación al añadir nuevas variables al modelo. Este procedimiento se repite iterativamente hasta que no queda sin incluir en el modelo ninguna variable capaz de dar lugar a cambios estadísticamente significativos.

Las interacciones entre las variables explicativas poseen a menudo una alta capacidad predictiva (MARGULES *et al.* 1987). Ello significa que el producto de dos variables cualesquiera puede ser capaz de explicar mejor una variable dependiente que la suma de esas mismas variables, debido al efecto sinérgico de su actuación (Figura 6). Por ello, es conveniente incluir todos los términos de interacción entre las variables explicativas (incluyendo las espaciales), comprobando si la adición puede mejorar el ajuste del modelo obtenido con las variables sencillas. De nuevo en este caso, hay que utilizar un proceso de selección por pasos hacia atrás tras la inclusión de cada interacción significativa. Para añadir las variables espaciales al modelo utilizamos los nueve términos de la ecuación polinomial de tercer grado de latitud y longitud (Trend Surface Analysis), incluyéndolos en un solo paso al modelo y enviándolos también a un proceso de selección por pasos

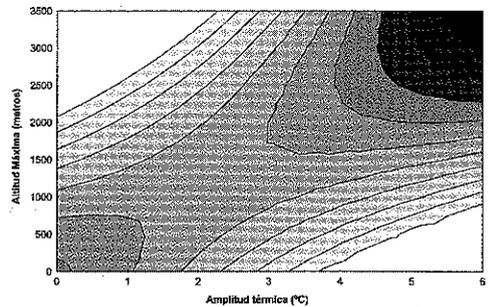


Figura 6: Interacción entre dos variables independientes y la riqueza de especies. Gráfico de contornos en el que los colores (eje z) representan el número de especies (negro: máximo [hasta 40 especies]; blanco: mínimo [desde 5 especies]) para cada valor de altitud máxima (eje y) y amplitud térmica (eje x). Datos procedentes de Lobo & Martín-Piera, 2002.

Figure 6: Interaction among two independent variables and species richness. Contour plot where colours (z axis) mean species number (black: maximum [up to 40 species]; white: minimum [from 5 species]) for each maximum altitude (y axis) and temperature range (x axis) scores. Data from Lobo & Martín-Piera, 2002.

hacia atrás con la intención de eliminar los términos no significativos (ver LEGENDRE 1993 y LEGENDRE & LEGENDRE 1998).

Para elegir la función predictora también pueden incluirse todas las variables explicativas a la vez junto con sus funciones cuadráticas, cúbicas y sus términos de interacción, estimando los modelos o ecuaciones con mayor capacidad predictiva mediante el Criterio de Información de Akaike (McCULLAGH & NELDER 1989), el cual maximiza el porcentaje de la variabilidad que explica el modelo y minimiza su error (la diferencia entre lo que explica el modelo y los valores reales). El objetivo de este criterio es encontrar un compromiso que pondere el ajuste del modelo y la complejidad del mismo, que está en función del número de términos que lo integran (ver MAC NALLY 2000). En general, los procedimientos de selección por pasos son siempre imprecisos, ya que la función que obtenemos no es necesariamente la que más se ajusta a los datos dentro de todas las posibles. Son búsquedas heurísticas que producen buenos resultados pero no tienen porque encontrar la solución óptima. El examen del Criterio de Información de Akaike o la búsqueda exhaustiva de la mejor función utilizando todas las combinaciones posibles de elementos, pueden dar con la o las soluciones óptimas (CORTÉS *et al.* 2000; MAC NALLY 2000; HORTAL & LOBO 2001), pero requieren una gran capacidad de computación.

Validación del modelo

Como hemos discutido anteriormente, la bondad de ajuste que nos proporciona el modelo tan sólo refleja su relación con los datos introducidos en el proceso iterativo, no su validez o utilidad predictiva en el territorio estudiado. Un modelo de este tipo es una herramienta que va a ser utilizada para un propósito determinado. Debemos evaluar en la medida de lo posible todas sus fortalezas y debilidades para saber si puede ser mejorado y si es útil al propósito para el que ha sido construido. De este modo, es necesario evaluar cada modelo, tanto para depurar el proceso de ajuste, como para determinar exactamente cuáles son sus debilidades.

Análisis de los residuos

A partir del primer modelo hallado, NICHOLLS (1989) recomienda llevar a cabo un análisis de los residuos para identificar las observaciones extrañas (outliers): aquellas cuadrículas en las que el valor absoluto de los residuos es mayor que la desviación estándar de los valores predichos. También es preciso seleccionar los puntos con valores elevados de Peso Potencial (PLV o Potential Leverage; NICHOLLS 1989). El PLV es una medida de la distancia de cada observación al centroide del espacio multidimensional definido por las variables incluidas en el modelo. Es necesario explorar cada uno de los outliers y cada una de las observaciones con valores elevados de PLV, al objeto de comprobar si son debidos a datos erróneos, o si, por el contrario, son consecuencia del carácter diferencial de la región en la que se encuentran. Es decir, si son parte fundamental de la variabilidad ambiental del territorio analizado. Mientras que las primeras observaciones deben ser eliminadas, las segundas deben permanecer en la estimación del modelo ya que así se podrá incluir una mayor heterogeneidad ambiental en el ajuste del modelo. Como indicamos antes, los parámetros deben estimarse nuevamente una vez eliminados los verdaderos outliers.

Para comprobar que los residuos del modelo no poseen autocorrelación espacial y que, por tanto, no queda ninguna variable importante por incluir en el modelo (ver Selección de las Variables Predictivas), es conveniente utilizar los tests de autocorrelación de la I de Moran y la C de Geary (LEGENDRE & VAUDOR 1991). En ellos, la autocorrelación se comprueba con respecto a diferentes clases de distancia definidas previamente. Convenientemente, estas clases de distancia deben ser similares a la longitud de las unidades territoriales.

Fiabilidad y poder predictivo del modelo

Tras eliminar los datos erróneos y una vez ajustado el modelo, aplicamos la función predictiva resultante a los valores de las variables explicativas para cada una de las unidades territoriales de la región estudiada, obteniendo de esta manera un mapa pronóstico de la distribución

de los valores del atributo de biodiversidad estudiado. El principal criterio para evaluar el modelo es la manera en que estas predicciones que extrapolamos a todo el territorio son fiables. El mejor método para comprobar la fiabilidad del modelo es el empírico, es decir, llevar a cabo un inventario en algunas de las zonas pobremente muestreadas para comprobar si los valores predichos y reales son similares. Sin embargo, este método es, en muchos casos, demasiado lento y costoso.

Algunos autores dividen las unidades territoriales con datos fiables en dos grupos diferentes, utilizando uno para ajustar el modelo y otro para validar sus resultados. Sin embargo, puede realizarse la validación a partir de los propios datos incluidos en el modelo. Para ello, es posible utilizar un test de Jackknife, en el que se extraen las unidades territoriales una por una, se calculan los parámetros de ajuste del modelo y se compara la predicción con la función resultante en la unidad espacial no considerada con el valor real del atributo en esa misma unidad. Con un conjunto de datos de n unidades espaciales, el modelo se recalcula n veces, extrayendo una unidad territorial diferente en cada ocasión. Cada uno de los modelos basados en las $n-1$ unidades territoriales se aplica al dato excluido para predecir el valor del atributo. La relación entre los valores observados y los estimados puede determinarse mediante un simple coeficiente de correlación. La elección de uno u otro sistema dependerá, principalmente, de la cantidad de datos fiables disponibles. Cuando los datos fiables son escasos, es recomendable no dividir estos en dos grupos, usarlos todos y validar el modelo mediante Jackknife (ver GUISAN & ZIMMERMANN 2000).

Ya se ha mencionado que para obtener una estimación de la variabilidad total de los datos explicada por un modelo calculado mediante GLM, puede determinarse el porcentaje de la desviación total explicada (DOBSON 1999). Sin embargo, este no es un buen estimador de su capacidad predictiva. Para estimar el poder predictivo del modelo es conveniente utilizar la media del error en la predicción (prediction

error; E_i) para cada observación tras haber realizado un procedimiento de Jackknife. Este se define como la distancia relativa entre el valor predicho para el caso i cuando se excluye del proceso de estimación del modelo (P_i) y el valor observado (O_i) (PASCUAL & IRIBARNE 1993). El error porcentual para el caso i será:

$$E_i = \frac{|O_i - P_i|}{O_i} \times 100$$

Como medida del error en la predicción asociado al modelo se utiliza la media de todas las estimaciones de error (Error Predictivo Medio; EPM). La inversa de esta medida ($EPM^{-1}=100-EPM$) puede usarse como estimador de su poder predictivo.

UN EJEMPLO PRÁCTICO

Vamos a utilizar como ejemplo real la modelización de la riqueza (número de especies) de una familia de coleópteros en la Península Ibérica: los escarabeidos coprófagos (Coleoptera, Scarabaeinae) (LOBO & MARTÍN-PIERA 2002). Los datos biológicos provienen de la base de datos BANDASCA (ver estructura en LOBO & MARTÍN-PIERA 1991), que en la actualidad contiene datos de 15.740 registros y 101.996 individuos de las 53 especies ibéricas de la familia Scarabaeinae (MARTÍN-PIERA 2000), procedentes de una revisión exhaustiva de toda la información accesible contenida en la bibliografía y las colecciones de historia natural disponibles. Hemos escogido como ejemplo este grupo de insectos debido a que nuestro conocimiento taxonómico y faunístico sobre ellos es suficiente como para proporcionar datos fiables para algunas localidades ibéricas pero, sin embargo, no tan amplio como para que no sea necesario extrapolarlo al resto de nuestra Península.

La base de datos ambientales utilizada se compone de: i) un Modelo Digital del Terreno de la Península Ibérica con una resolución de 1 km², ii) los valores de los datos climáticos para cada cuadrícula UTM de 50km de lado (cortesía de W. Cramer; CLIMATE database versión 2; <http://www.pik-potsdam.de/~cramer/-climate.htm>), iii) un mapa de usos del suelo,

procedente de una reclasificación en cuatro categorías de las 44 presentes en la información ráster para España y Portugal (282 metros de resolución) proporcionada por la Agencia Europea para el Medioambiente (CORINE Programme 1985-1990; EUROPEAN ENVIRONMENT AGENCY 1996) y, iv) un mapa geológico de la Península Ibérica (INSTITUTO GEOGRÁFICO NACIONAL 1995), del que se digitalizaron los suelos calcáreos, silíceos y arcillosos (1 Km de resolución).

Para modelizar la riqueza de especies de escarabeidos en la Península Ibérica, hemos escogido las 252 cuadrículas UTM de 50X50 Km. (Universal Transverse Mercator) con más de un 85% de superficie terrestre (Figura 10). Hemos escogido la malla UTM, en lugar de la malla Lat/Long, debido a que, en territorios extensos, sus cuadrículas presentan la misma superficie de manera constante. La celda de 2.500 Km² se escogió debido al propósito general del análisis, ya que el uso de una más pequeña (por ejemplo, 100 Km²) podría oscurecer algunos patrones generales gracias al "ruido" producido por procesos locales (micro-ecológicos o de dinámica de poblaciones). También da lugar a un menor número de casos (252), y nos permite poseer información biológica fiable a lo largo de todo el espectro completo de condiciones ambientales.

Como unidad de esfuerzo de muestreo escogimos el número de registros extraídos de BANDASCA para cada unidad territorial. Cuando se mapea su distribución geográfica se obtiene un mapa claramente sesgado (Lámina 5). Se construyeron las curvas de acumulación aleatorizando 500 veces el orden de entrada en el que cada registro es incluido en ellas mediante el programa Estimates 6 (COLWELL 2000). Tras ajustar cada curva, hemos considerado como suficientemente bien muestreadas aquellas cuadrículas en las que el número total de registros es mayor que su valor de $r_{0,01}$ (es decir, es necesario un esfuerzo de muestreo de más de 100 registros para encontrar una especie nueva para el inventario). En total, 82 cuadrículas UTM fueron seleccionadas mediante este criterio (lámina 5). La información biológica y

ambiental de todas estas cuadrículas fue utilizada para elaborar un modelo predictivo que nos permitiera extrapolar la riqueza de especies de Scarabaeinae en el resto del territorio ibérico mal muestreado.

Hemos asumido que la distribución de la riqueza de escarabeidos en la Península Ibérica es consecuencia de la actuación de factores, tanto ecológicos como histórico-geográficos. Debido a ello hemos incluido en primer lugar las variables relacionadas con el clima, la topografía, la geología y el uso del suelo. Este último tipo de variables se incluyó debido a la fuerte influencia que las actividades humanas ejercen sobre la diversidad del grupo y su relación con la ausencia de cubierta arbórea y la presencia de cabaña ganadera. Posteriormente se incluyeron las variables relacionadas con la heterogeneidad ambiental asumiendo, como es frecuente, que la cantidad de especies que es capaz de coexistir en una unidad territorial depende de su diversidad ambiental. Finalmente, se consideró la influencia del espacio, incluyendo los nueve términos de la polinomial de tercer grado de la latitud y la longitud. De este modo, se pretende introducir en el modelo el efecto de cualquier variable ambiental, biótica o histórica no considerada, capaz de producir un patrón espacial.

Se extrajeron así 24 variables continuas para cada una de las 252 cuadrículas ibéricas usando el software SIG Idrisi 2.0 (CLARK LABS 1998): dos variables espaciales (latitud y longitud del centroide de la cuadrícula); dos geográficas (distancia a los Pirineos y cantidad de superficie marítima); tres topográficas (altitud mínima, máxima y media); dos geológicas (cantidad de superficie de suelo básico y ácido); seis climáticas (temperatura mínima y máxima anual, temperatura media anual, precipitación anual media, precipitación estival y número de días de sol al año); cuatro de uso del suelo (área ocupada por cultivos o núcleos urbanos, bosques, arbustos y pastos) y cinco variables de diversidad ambiental (rango de altitudes, rango anual de precipitaciones y temperaturas, diversidad de usos del suelo y diversidad geológica). Todas las variables predictivas se estandariza-

ron a media cero y varianza uno, excepto para el caso de la latitud y la longitud (ver selección de las variables predictivas).

Para ajustar el modelo se introdujeron en el procedimiento iterativo descrito anteriormente, en primer lugar, las funciones lineal, cuadrática y cúbica de todas las variables ambientales. Tras ellas se introdujeron secuencialmente todas las interacciones o productos entre pares de variables (por ejemplo, temperatura máxima \times precipitación media). Finalmente, se introdujo el polinomio espacial descrito para recoger la estructura espacial debida a variables de cualquier tipo no incluidas en el análisis. Al examinar los residuos y valores de PLV derivados del modelo preliminar generado ($n=82$), siete cuadrículas localizadas en las regiones muestreadas de manera más intensa o en las islas fueron identificadas como outliers reales (LOBO & MARTÍN-PIERA 2002), por lo que, al construir el modelo predictivo final de la riqueza de especies fueron eliminadas ($n = 75$; ver Lámina 5).

Sólo 8 variables ambientales (distancia a los Pirineos, temperatura mínima, máxima y media anual, precipitación total anual, días de sol anuales, área de pastizal y diversidad de usos del suelo) y los términos cuadrático y cúbico de la latitud son significativos (Tabla 1). Como la función lineal de los días de sol anuales proporcionó el cambio más importante en la desviación, esta variable fue la primera incluida en el modelo. Tras ella, se seleccionó la función lineal de la altitud máxima, seguida de la función cuadrática de la superficie de pastizal, aunque tan sólo su término cuadrático produce un cambio significativo en la desviación. En el siguiente paso, la función cúbica de la diversidad de usos del suelo es la única variable de diversidad ambiental añadida al modelo, siendo su término cuadrático eliminado del modelo (Tabla 2). De los términos de interacción, el efecto conjunto de la cantidad de superficie ocupada por bosques y la diversidad geológica, del área terrestre y la altitud máxima, de la latitud y el rango anual de precipitaciones, y de la superficie de rocas básicas y la diversidad geológica entran de manera iterativa en el

modelo. Su inclusión elimina el término cúbico de la diversidad de usos del suelo (Tabla 2). Finalmente, al añadir al modelo los términos espaciales significativos (los términos cuadrático y cúbico de la latitud), se produce la salida del modelo del término lineal de los días de sol anuales y de la interacción entre la latitud y el rango anual de precipitaciones (Tabla 2). El modelo final es, de este modo:

$$S = \exp [c + \text{Altitud Máxima} + (\text{Superficie de pastos})^2 + \text{Diversidad de usos del suelo} + (\text{Superficie de bosques} \times \text{Diversidad geológica}) + (\text{Área terrestre} \times \text{Altitud máxima}) + (\text{Rocas básicas} \times \text{Diversidad geológica}) + \text{Latitud}^2 + \text{Latitud}^3],$$

...donde S es el número total de especies de escarabeidos y c es el intercepto. Este modelo logra explicar el 62.4% de la desviación total (bondad de ajuste) (LOBO & MARTÍN-PIERA 2002). Los residuos de este modelo siguen una distribución normal, el gráfico de los residuos frente a los valores predichos forma una nube homogénea alrededor del centro y los errores estándar de los coeficientes son bajos. Por lo que respecta a la autocorrelación espacial de los residuos, ninguno de los valores de la I de Moran en las 8 diferentes clases de distancia utilizadas fueron significativos ($p = 0.05$) aplicando incluso una corrección de Bonferroni. De este modo puede decirse que los residuos del modelo no estaban espacialmente autocorrelacionados. Sin embargo, la representación de los valores predichos frente a los observados mostró una sobreestimación del número de especies en las cuadrículas con bajos valores de riqueza y una subestimación en las celdas con valores elevados (Figura 7). Este problema del modelo permanece aun al eliminar todas las observaciones que pudieran ser consideradas potencialmente como outliers. Los resultados de un análisis Jackknife sobre el modelo final mostraron una gran correspondencia entre los valores observados de riqueza y los predichos para las 75 cuadrículas. La correlación entre ellos fue positiva y significativa, siendo el $EPM-1 = 84.1\%$, lo que muestra que el modelo

Variables	Abreviatura	Términos	Desviación	g.l.	C-Desv	F	signo
Modelo Nulo			99.61	74			
ESPACIALES							
Longitud	LON						
Latitud	LAT	LAT ² + LAT ³	74.16	72	25.45	24.71***	--
GEOGRÁFICAS							
Área terrestre	A	A	98.46	73	1.15	0.85	
Distancia a los Pirineos	Dp	Dp	93.58	73	6.03	4.70*	+
TOPOGRÁFICAS							
altitud mínima	e	e	99.45	73	0.16	0.12	
altitud máxima	E	E	97.47	73	2.15	1.61	
altitud media	em	em	99.46	73	0.05	0.04	
GEOLÓGICAS							
rocas calizas	Cr	Cr	98.64	73	0.98	0.72	
rocas silíceas	Ar	Ar	98.16	73	1.45	1.08	
CLIMÁTICAS							
temperatura mínima anual	t	t	97.96	73	1.65	1.23	
		t + t ²	86.77	72	11.19	9.28**	--
temperatura máxima anual	T	T	93.71	73	5.90	4.60*	+
temperatura media anual	tm	tm	96.23	73	3.38	2.56	
		tm + tm ²	89.02	72	10.59	8.57**	--
precipitaciones anuales	P	P	99.52	73	0.09	0.06	
		P + P ²	95.72	72	3.88	2.92	
		P + P ² + P ³	93.01	71	6.60	5.04*	--
precipitaciones estivales	Ps	Ps	99.29	73	0.22	0.34	
días de sol anuales	Ds	Ds	81.52	73	18.10	16.21***	+
USOS DEL SUELO							
superficie cultivada y urbana	U	U	99.02	73	0.59	0.43	
superficie de bosques	F	F	99.60	73	0.01	0.01	
superficie de arbustos	S	S	99.21	73	0.40	0.30	
superficie de pastos	G	G	96.13	73	3.48	2.84	
		G + G ²	94.55	72	5.06	3.86	
		G + G ² + G ³	92.73	71	6.88	5.27*	--
DIVERSIDAD AMBIENTAL							
rango de altitudes	AR	AR	97.71	73	1.91	1.42	
rango anual de temperaturas	TV	TV	99.30	73	0.31	0.23	
rango anual de precipitaciones	PV	PV	99.41	73	0.20	0.15	
diversidad de usos	LUD	LUD	98.23	73	1.38	1.02	
		LUD + LUD ²	95.31	72	4.30	3.25	
		LUD + LUD ² + LUD ³	91.25	71	8.36	6.51*	--
diversidad geológica	GD	GD	99.00	73	0.61	0.45	

Tabla 1: Variables ambientales explicativas estudiadas en las 75 cuadrículas UTM 50x50 Km de la Península Ibérica muestreadas adecuadamente (Figura 10b). Las variables espaciales se escogieron eliminando los términos no significativos de una ecuación polinomial de tercer grado de la latitud y la longitud mediante una selección por pasos hacia atrás. Se calculó tanto la desviación, como el cambio en la desviación (C-Desv) a partir de un modelo nulo para la riqueza total de especies de coleópteros coprófagos. Las funciones lineal, cuadrática o cúbica de cada variable fueron seleccionadas si producían un cambio estadísticamente significativo en la desviación con una probabilidad menor de 0.05. El signo de las columnas corresponde al signo del término de cada función. *** p<0.001; ** p<0.01; * p<0.05.)

Table 1: Environmental explanatory variables considered in the 75 adequately sampled 50x50 km UTM grid squares of the Iberian Peninsula (Figure 10b). Spatial variables were chosen by removing the non-significant terms from a third-degree polynomial equation of latitude and longitude by a backward stepwise selection. Deviance and change in deviance (C-Desv) from a null model for total dung beetle species number. The linear, quadratic or cubic functions of each variables were selected if they accounted for statistically significant change in the deviance with a probability lower than 0.05. The sign of the columns corresponds to the sign of the term of each function. *** p<0.001; ** p<0.01; * p<0.05.

Paso	Variable	Desv	g.l.	C-Desv	P	Desviación explicada (%)	Coefficientes	Err.Est.
Modelo Nulo		99.61	74					
PASO 1								
	DS	81.52	73	18.10	16.21***	0.1817		
PASO 2								
	E	71.45	72	10.07	10.14**	0.3177	0.062	0.026
PASO 3								
	G	67.59	71	3.86	4.06*	0.3546		
	+ G2	66.93	70	0.66	0.69	0.3608	0.014	0.006
	- G	67.11	71	0.18	0.19	0.3591		
PASO 4								
	LUD	65.12	70	1.99	2.14	0.3782	0.114	0.041
	+LUD2	62.70	69	2.42	2.66	0.4013		
	+LUD3	59.27	68	3.43	3.93	0.4340		
	-LUD2	59.54	69	0.27	0.31	0.4314		
PASO 5								
	F x GD	54.13	68	5.41	6.80*	0.4831	-0.079	0.025
PASO 6								
	A x E	49.19	67	4.94	6.73*	0.5303	0.093	0.026
	-LUD3	49.21	68	0.02	0.03	0.5301		
PASO 7								
	LAT x PV	45.30	67	3.91	5.78*	0.5674		
PASO 8								
	Cr x GD	41.77	66	3.54	5.59*	0.6012	0.054	0.029
PASO 9								
	+ LAT ²	41.76	65	0.01	0.02	0.6013	-0.021	0.009
	+ LAT ³	37.34	64	4.42	7.57**	0.6434	-0.009	0.002
	- DS	37.40	65	0.06	0.11	0.6429		
	-LAT x PV	39.36	66	1.96	3.29	0.6241		
Intercepto							2.940	0.050

Tabla 2: Sumario del proceso de selección por pasos de las variables para construir un modelo de regresión múltiple para la riqueza de especies de coleópteros coprófagos. El modelo emplea una distribución de Poisson en la asunción del error y una función de ligado logarítmica y utiliza los valores de riqueza de especies y los datos ambientales de las 75 cuadrículas UTM de 50 km adecuadamente muestreadas (Figura 10b) tras siete outliers (ver texto). Los códigos de las variables son los de la Tabla 1 y el cambio en la desviación (C-Desv) tras incluir un término en el modelo se identificó mediante un test *F*, con una probabilidad menor que 0.05. Sólo están destacados los coeficientes y errores estándar (*Err. Est.*) de las variables significativas incluidas en el modelo final. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 2: Summary of the stepwise selection of variables to build a multiple regression model for species richness of dung beetles. The model employs a Poisson distribution of error assumption and a log link, and uses the species richness scores and environmental data of the 75 adequately sampled 50 km UTM grid squares (Figure 10b) after omitting seven outliers (see text). The variable codes as in Table 1. The change in deviance (C-Desv) after including a term in the model has been tested by an *F*-ratio test with a probability lower than 0.05. Only the coefficients and standard errors (*Err. Est.*) for parameters of the variables included in the final model are highlighted. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

predictivo es razonablemente bueno a pesar de su tendencia a reducir la diferencia entre los valores extremos de riqueza (LOBO & MARTÍN-PIERA 2002). Este defecto nos indica que las regiones con mayor número de especies son, probablemente, más ricas aun de lo que predice el modelo. Los resultados del modelo final para el conjunto de la Península Ibérica se representan en la figura 8 e, interpolados, en la lámina 3b.

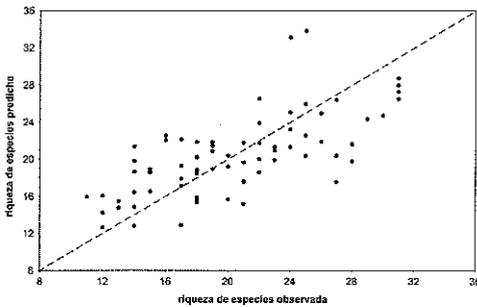


Figura 7: Relación entre los valores de riqueza de especies observados y predichos por el modelo final (datos de LOBO & MARTÍN-PIERA 2002). La línea discontinua representa la igualdad entre ambos valores.

Figure 7: Relationship between the observed and predicted species richness scores according to the final model (data from LOBO & MARTÍN-PIERA 2002). Discontinuous line represents equality between both scores.

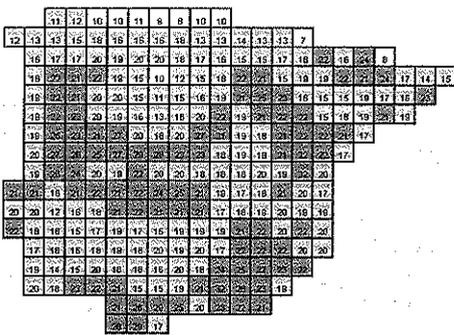


Figura 8: Valores predichos de la riqueza de especies de Scarabaeinae para las 252 cuadrículas UTM de 50x50 km (ver texto y Figura 10) de la Península Ibérica, de blanco (0-10 especies) a gris claro (11-20 especies) y gris oscuro (más de 21 especies). Modificado de LOBO & MARTÍN-PIERA 2002.

Figure 8: Predicted species richness of Iberian Scarabaeinae species richness for the 252 UTM 50x50 km Iberian grid squares (see text and Figure 10): white (0-10 species) to light grey (11-20 species) and dark grey (more than 21 species). Modified from LOBO & MARTÍN-PIERA 2002.

DISCUSIÓN

Como indican los resultados que hemos presentado, es posible producir una estimación geográfica fiable de la riqueza de especies de un grupo determinado. Para ello, es necesario recopilar exhaustivamente los datos faunísticos y corológicos disponibles y estudiar el nivel de muestreo a lo largo y ancho del territorio estudiado. Cuando poseemos conocimiento adecuado de la riqueza en un número suficiente de unidades territoriales, podemos extrapolar este valor al resto del territorio, modelizando el atributo a partir de los datos ambientales contenidos en una base de datos SIG y de la inclusión en el modelo de la estructura espacial no relacionada con el ambiente. Los mapas pronosticados de esta manera pueden permitirnos identificar los principales patrones espaciales de muchos atributos de la biodiversidad (LEVIN 1992), además de su evidente importancia para el desarrollo de estrategias de conservación a escala geográfica.

En algunos países con una larga tradición naturalista, la información corológica de calidad puede ser lo suficientemente abundante como para permitir realizar estimas fiables para determinados grupos, especialmente plantas vasculares y vertebrados, a partir de la información ya existente, aunque esté espacialmente sesgada. Sin embargo, en la mayor parte de los territorios del planeta, y para la gran mayoría de los grupos de seres vivos, el conocimiento existente es tan escaso y sesgado que hace imposible la tarea de modelizar su distribución. En ese caso, es posible realizar modelos predictivos menos fiables basados únicamente en las presencias constatadas de las especies (ver el apartado Modelos Predictivos y Biodiversidad) o desarrollar un proceso de muestreo previo al proceso de modelización, que nos permita recoger los datos mínimos para poder realizar estimas insesgadas de su distribución. Este muestreo ha de estar dirigido a capturar todas las condiciones ambientales diferentes desde la percepción del medio por el taxón que estamos estudiando, así como a muestrear aquellas áreas que, aunque sean

ambientalmente similares, estén segregadas en el espacio, pudiendo albergar faunas diferentes debido a diferentes procesos históricos.

Otros análisis similares efectuados sobre la riqueza de especies de Scarabaeinae en Portugal (nivel de significación 0.01; desviación explicada = 85.4%; EPM-1 = 90.8%; HORTAL *et al.* 2001) y Francia (nivel de significación 0.05; desviación explicada = 86.2%; EPM-1 = 82.3%; LOBO *et al.* 2002) obtuvieron mejores resultados tanto para el total de la desviación explicada como para la EPM-1 que los obtenidos por LOBO & MARTÍN-PIERA (2002), ejemplificados aquí. Esto puede ser debido a la mayor homogeneidad ambiental de ambos territorios con respecto a la totalidad de la Península Ibérica, destacando que, como es de esperar, los patrones producto de procesos heterogéneos con mayor influencia de factores contingentes son más difíciles de modelizar. En todos los casos, la inclusión de la estructura espacial mejoró los modelos basados únicamente en variables ambientales.

Los coleópteros coprófagos son uno de los grupos taxonómicamente mejor conocidos dentro de los insectos y, debido a su vistosidad y facilidad de captura, uno de los más colectados por entomólogos profesionales y aficionados detrás de las mariposas. Todas las especies de Scarabaeinae presentes en la Península Ibérica son conocidas y pueden ser identificadas (MARTÍN-PIERA 2000), y existe una gran tradición de estudio científico sobre este grupo. Desafortunadamente, este panorama no es el mismo para la gran mayoría de grupos de organismos y regiones del globo. Con un gran trabajo de inventario de la biota pendiente en todo el mundo y en la propia Península Ibérica (RAMOS *et al.* 2001) y con la necesidad imperiosa de identificar la distribución geográfica de la riqueza de especies debido a la llamada "crisis de la biodiversidad", ¿cómo podemos describir la distribución espacial de la biodiversidad en un tiempo razonable? Hay demasiados grupos no estudiados, demasiadas regiones inexploradas, muy pocos recursos económicos y muy pocos taxónomos. Muestrear todos los territorios poco o nada

explorados e identificar y clasificar todas las especies desconocidas parece poco práctico a corto y medio plazo (MAY 1990; EHRlich 1992; SYSTEMATICS AGENDA 2000 1994; COTTERILL 1995; WILLIAMS & HUMPHRIES 1996; BLACKMORE & CUTLER 1996; MARTÍN-PIERA 1997; MARTÍN-PIERA & LOBO en prensa). Debemos utilizar otras estrategias.

Es necesario diseñar procedimientos para identificar de manera rápida y barata las áreas de mayor diversidad, por lo que parece más razonable:

- desarrollar nuevas metodologías para el diseño de muestreos a escala regional que nos permitan maximizar nuestra capacidad de elaborar inventarios fiables. Es decir, encontrar más especies con menores costes en tiempo y dinero.
- elaborar extensas bases de datos corológicas, con protocolos generales de intercambio de datos que permitan su uso y comparación de manera fácil y generalizada vía internet (BISBY 2000a y b; EDWARDS *et al.* 2000a y b; SMITH *et al.* 2000),
- y encontrar técnicas de modelización predictiva fáciles de utilizar basadas en variables ambientales y espaciales.

Para afrontar esta última tarea es necesario realizar estudios comparados entre las diversas metodologías capaces de producir pronósticos de atributos de diversidad en forma de mapas (modelización heurística, modelización del nicho, búsqueda exhaustiva de modelos, técnicas de interpolación espacial como kriging y co-kriging, etc.) (MANEL *et al.* 1999 a, b y 2001; GUIGAN & ZIMMERMANN 2000) a tantas escalas espaciales y sobre tantos grupos y regiones como sea posible. Debemos analizar las fortalezas y debilidades de cada técnica a cada escala en cada tipo de región y cada grupo, lo que nos permitirá conocer de manera precisa qué método y escala son los más adecuados para cada caso y objetivo concreto, en términos de simplicidad, precisión y tiempo. Una vez estemos en posesión de esta informa-

ción, los muestreos regionales podrán ser diseñados para alcanzar niveles de conocimiento faunístico o florístico suficientes como para permitir un pronóstico fiable con el menor coste posible.

AGRADECIMIENTOS

Este trabajo está dedicado a la memoria del Dr. Fermín Martín Piera, recientemente fallecido.

Gran parte de las ideas que aquí se exponen son también suyas.

Este trabajo se enclava dentro del proyecto DGCYT "Faunística Predictiva: Análisis comparado de la efectividad de distintas metodologías y su aplicación para la selección de reservas naturales" (REN-2001-1136/GLO). Joaquín Hortal disfruta de una beca predoctoral Museo Nacional de Ciencias Naturales/CSIC/Comunidad de Madrid.

REFERENCIAS BIBLIOGRÁFICAS

- ALLEN, C.R., PEARLSTINE, L.G. & KITCHENS, W.M. 2001. Modelling viable mammal populations in gap analyses. *Biological Conservation* 99: 135-144.
- AUGUSTIN, N.H., CUMMINS, R.P. & FRENCH, D.D. 2001. Exploring spatial vegetation dynamics using logistic regression and multinomial logit model. *Journal of Applied Ecology* 38: 991-1006.
- AUSTIN, M. P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42: 11-21.
- AUSTIN, M. P., NICHOLLS, A. O. & MARGULES, C. R. 1990. Measurement of the realised qualitative niche: environmental niches of five Eucalyptus species. *Ecological Monographs* 60: 161-177.
- AUSTIN, M. P., PAUSAS, G. J. & NICHOLLS, A. O. 1996. Patterns of tree species richness in relation to environment in south-eastern New South Wales, Australia. *Australian Journal of Ecology* 21: 154-164.
- BEARD, K.H., HENGARTNER, N. & SKELLY, D.K. 1999. Effectiveness of predicting breeding bird distributions using probabilistic models. *Conservation Biology* 13: 1108-1116.
- BIRKS, H. J. B. 1996. Statistical approaches to interpreting diversity patterns in the Norwegian mountain flora. *Ecography* 19: 332-340.
- BISBY, F. A. 2000a. The quiet revolution: biodiversity informatics and the internet. *Science* 289: 2309-2312.
- BISBY, F. A. 2000b. Response to Databases tailored for biodiversity conservation. *Science* 290: 2074.
- BLACKMORE, S. & CUTLER, D. (eds.) 1996. Systematic Agenda 2000. The challenge for Europe. Proceedings of the Workshop organized by the European Science Foundation Systematic Biology Network, the Linnean Society of London, the Rijksherbarium/Hortus Botanicus, Leiden University, and the Systematics Association. Linnean Society of London.
- BOJORQUEZ-TAPIA, L.A., AZUARA, I., ESCURRA, E. & FLORES-VILLELA, O. 1996. Identifying conservation priorities in México through geographic information systems and modelling. *Ecological Applications* 5: 215-231.
- BOSQUE SENDRA, J. 1997. *Sistemas de Información Geográfica. Segunda Edición*. Editorial Rialp, Madrid.
- BUCKLAND, S. T. & ELSTON, D. A. 1993. Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology* 30: 478-495.
- BUSBY, J.R. 1991. BIOCLIM - A bioclimate analysis and prediction system. In Margules, C.R., Austin, M.P. (eds.). *Nature Conservation: Cost effective biological surveys and data analysis*. CSIRO, Melbourne.
- CARPENTER, G., GILLISON, A. N. & WINTER, J. 1993. DOMAIN: A flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation* 2: 667-680.

- CLARK LABS 1998. Idrisi 2.0. Software SIG y manual. Clark University.
- CLIFF, A. D. & ORD, J. K. 1981. Spatial Processes. Models and Applications. Pion Limited.
- COLWELL, R. K. & CODDINGTON, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London B* 345: 101-118.
- COLWELL, R. K. 2000. EstimateS, statistical estimation of species richness and shared species from samples. Version 6. Guía del Usuario y aplicación disponible en <http://viceroy.eeb.uconn.edu/estimates>.
- CORTÉS, M., VILLACAMPA, Y., MATEU, J., & USÓ, J. L. 2000. A new methodology for modelling highly structured systems. *Journal of Environmental Modelling and Software* 15: 461-470.
- COTTERILL F. P. D. 1995. Systematics, biological knowledge and environmental conservation. *Biodiversity and Conservation* 4: 183-205.
- CRAWLEY, M. J. 1993. GLIM for Ecologists. Blackwell Scientific Publications.
- DENNIS, R.L. 2001. Progressive bias in species status in symptomatic of fine-grained mapping units subject to repeated sampling. *Biodiversity and Conservation* 10: 483-494.
- DENNIS, R. L. & EALES, H. T. 1999. Probability of site occupancy in the large heath butterfly *Coenonympha tullia* determined from geographical and ecological data. *Biological Conservation* 87: 295-301.
- DOBSON, A. 1999. An introduction to Generalized Linear Models. Chapman & Hall/CRC.
- EDWARDS, J. L., LANE, M. A. & NIELSEN, E. S. 2000a. Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289: 2312-2314.
- EDWARDS, J. L., LANE, M. A. & NIELSEN, E. S. 2000b. Response to Databases tailored for biodiversity conservation. *Science* 290: 2073-2074.
- EHRlich, P. R. 1992. Population biology of checkerspot butterflies and the preservation of global biodiversity. *Oikos* 63: 6-12.
- EUROPEAN ENVIRONMENT AGENCY 1996. Natural Resources CD-Rom. European Environment Agency, Bruselas.
- FAGAN, W. F. & KAREIVA, P. M. 1997. Using compiled species list to make biodiversity comparisons among regions: a test case using Oregon butterflies. *Biological Conservation* 80: 249-259.
- FRESCINO, T.S., EDWARDS, T.C. & MOISEN, G.G. 2001. Modelling spatially explicit forest structural attributes using generalized additive models. *Journal of vegetation science* 12: 15-26.
- GARCÍA HERNÁNDEZ, E. & BOSQUE SENDRA, J. 2001. Bases de datos cartográficas de cobertura global accesibles on-line. *Geofocus (Recursos)*, 1, p. 5-10. Disponible en <http://geofocus.rediris.com>.
- GASTON, K. J. 1994. *Rarity*. Chapman & Hall, London.
- GASTON, K. J. 1996. Species richness: measure and measurement. En K. J. Gaston (ed.) *Biodiversity. A biology of numbers and difference*. pp. 77-113. Blackwell Science.
- GOTELLI, N. J. & COLWELL, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4: 379-391.
- GRIFFITHS, G. H., EVERS HAM, B. C. & ROY, D. B. 1999. Integrating species and habitat data for nature conservation in Great Britain: data sources and methods. *Global Ecology and Biogeography* 8: 329-345.
- GUI SAN, A. & ZIMMERMANN, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186
- HANSKI, I. 1999. *Metapopulation Ecology*. Oxford University Press, Oxford.
- HEIKKINEN, R. K. & NEUVONEN, S. 1997. Species richness of vascular plants in the subarctic landscape of northern Finland: modelling relationships to the environment. *Biodiversity and Conservation* 6: 1181-1201.
- HENGVELD, R. 1997. Impact of Biogeography on a population-paradigm shift. *Journal of Biogeography* 24: 541-547

- HIRZEL, A.H., HELFER, V. & METRAL, F. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145: 111-121.
- HIRZEL, A.H. 2001. When GIS come to life. Linking landscape and population ecology for large population management modelling: the case of Ibex (*Capra ibex*) in Switzerland. Tesis doctoral, Université de Lausanne. Disponible en <http://www.unil.ch/biomapper/>.
- HORTAL, J., LOBO, J. M. & MARTÍN-PIERA, F. 2001. Forecasting insect species richness scores in poorly surveyed territories: the case of the Portuguese dung beetles (Col. Scarabaeinae). *Biodiversity and Conservation* 10: 1343-1367.
- HORTAL, J. & LOBO, J. M. 2001. A preliminary methodological approach to model the spatial distribution of biodiversity attributes. En: J. Mateu & F. Montes (eds.) *Spatio-temporal modelling of environmental processes: proceedings of the 1st Spanish Workshop of Spatio-temporal Modelling of Environmental Processes*: pp. 211-239. Publicacions de la Universitat Jaume I, Col·lecció "Treballs d'Informàtica i Tecnologia" 10, Castelló de la Plana.
- HUSTON, M.A. 1994. *Biological Diversity. The coexistence of species on changing landscapes*. Cambridge University Press, Cambridge.
- INSTITUTO GEOGRÁFICO NACIONAL 1995. Atlas nacional de España. Volúmenes 1 y 2. Centro Nacional de Información.
- IVERSON, L. R. & PRASAD, A. M. 1998a. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs* 68: 465-485.
- IVERSON, L. R. & PRASAD, A. M. 1998b. Estimating regional plant biodiversity with GIS modeling. *Diversity and Distributions* 4: 49-61.
- JOHNSTON, C. A. 1998. *Geographic Information Systems in Ecology*. Blackwell Science.
- LAWTON, J. H. 1999. Are there general laws in ecology? *Oikos* 84(2): 177-192.
- LAWTON, J. H., PRENDERGAST, J. R. & EVERSHAM, B. C. 1994. The numbers and spatial distributions of species: analyses of British data. En: P. L. Forey, C. J. Humphries & R. I. Vane-Wright (eds.). *Systematics and Conservation Evaluation, Systematics Association Special Volume No. 50*, pp 177-195. Clarendon Press, Oxford.
- LEATHWICK, J.R. 1998. Are new Zealand's Nothofagus species in equilibrium with their environment? *Journal of Vegetation Science* 9: 719-732
- LEGENDRE, P. & VAUDOR, P. 1991. The R Package: Multidimensional analysis, spatial analysis. Département de sciences biologiques, Université de Montréal. Disponible en <http://www.fas.umontreal.ca/>.
- LEGENDRE, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659-1673.
- LEGENDRE, P. & LEGENDRE, L. 1998. *Numerical Ecology*, 2nd edition. Elsevier.
- LEÓN CORTÉS, J.L., SOBERÓN-MAINERO, J. & LLORENTE-BOUSQUESTS, J. 1998. Assessing completeness of Mexican sphinx moth inventories through species accumulation functions. *Diversity and Distributions* 4: 37-44.
- LEVIN, S. A. 1992. The problem of pattern and scale in ecology. *Ecology* 73(6): 1943-1967.
- LOBO, J. M. & MARTÍN-PIERA, F. 1991. La creación de un banco de datos zoológico sobre los Scarabaeidae (Coleoptera: Scarabaeoidea) ibero-baleares: una experiencia piloto. *Elytron* 5: 31-38.
- LOBO, J. M. & MARTÍN-PIERA, F. 2002. Searching for a predictive model for Iberian dung beetle species richness based on spatial and environmental variables. *Conservation Biology* 16: 158-173.
- LOBO, J. M., CASTRO, I. & MORENO, J. C. 2001. Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. *Biological Journal of the Linnean Society* 73: 233-253.
- LOBO, J. M., LUMARET, J. P. & JAY-ROBERT, P. 2002. Modelling the species richness distribution of French dung beetles and delimiting the predictive capacity of different groups of explanatory variables (Coleoptera, Scarabaeidae). *Global Ecology and Biogeography*, 11: 256-257.
- LOBO, J. M. 2000. ¿Es posible predecir la distribución geográfica de las especies basándose en variables ambientales? En: F. Martín-Piera, J. J. Morrone & A. Melic (eds.) *Hacia un proyecto*

- CYTED para el inventario y estimación de la diversidad entomológica en Iberoamérica: PrIBES 2000. m3m-Monografías Tercer Milenio, vol. 1, pp. 55-68. Sociedad Entomológica Aragonesa (SEA), Zaragoza, 326 pp.
- LOYN, R.H., MCNABB, E.G., VOLODINA, L. & WOLLIG, R. 2001. Modelling landscape distributions of large forest owls as applied to managing forest in north-east Victoria, Australia. *Biological Conservation* 97: 361-376.
- 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., BUCKTON, S.T. & ORMEROD, S.J. 1999a. Alternative methods for predicting species distribution: an illustration with Himalaya river birds. *Journal of Applied Ecology* 36: 734-747.
- MANEL, S., DIAS, & ORMEROD, S.J. 1999b. 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., WILLIAMS, H.C. & ORMEROD, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38: 921-931.
- MAGURRAN, A. E. 1988. *Ecological diversity and its measurement*. Princeton University Press.
- MARGULES, C. R., NICHOLLS, A. O. & AUSTIN, M. P. 1987. Diversity of Eucalyptus species predicted by a multi-variable environment gradient. *Oecologia* 71: 229-232.
- MARTÍN-PIERA, F. & LOBO, J. M. Database records as a sampling-effort surrogate to predict spatial distribution of insects in either poorly or unevenly surveyed areas. *Acta Zoologica Ibérica e Macaronésica*, en prensa.
- MARTÍN-PIERA, F. 1997. Apuntes sobre biodiversidad y conservación de insectos: Dilemas, ficciones y ¿soluciones?. *Boletín de la Sociedad Entomológica Aragonesa* 20: 25-55.
- MARTÍN-PIERA, F. 2000. Familia Scarabaeidae. En: F. Martín-Piera & J.I. López-Colón (eds.) *Fauna Ibérica* 14. Coleoptera, Scarabaeoidea I. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid.
- MAY, R. 1990. Taxonomy as destiny. *Nature* 347: 129-130.
- MCCARTHY, M.A., LINDENMAYER, D.B. & POSSINGHAM, H.P. 2001. Assessing spatial PVA models of arboreal marsupials using significance test and Bayesian statistics. *Biological Conservation* 98: 191-200.
- MCCULLAGH, P. & NELDER, J. A. 1989. *Generalized Linear Models* (2nd ed). Chapman & Hall.
- MILLER, R. I. 1994. *Mapping the Diversity of Nature*. Chapman & Hall.
- MITCHELL, N. D. 1991. The derivation of climate surfaces for New Zealand, and their application to the bioclimatic analysis of the distribution of kauri (*Agathis australis*). *Journal of the Royal Society of New Zealand* 21: 13-24.
- MOISEN, G.G. & EDWARDS, T.C. 1999. Use of generalized linear models and digital data in a forest inventory of northern Utah. *Journal of Agricultural, Biological, and Environmental Statistics* 4: 372-390.
- MORENO, C. E. & HALFFTER, G. 2000. Assessing the completeness of bat biodiversity inventories using species accumulation curves. *Journal of Applied Ecology* 37: 149-158.
- MORENO, C. E. & HALFFTER, G. 2001. On the measure of sampling effort used in species accumulation curves. *Journal of Applied Ecology* 38, 487-490.
- NICHOLLS, A. O. 1989. How to make biological surveys go further with generalised linear models. *Biological Conservation* 50: 51-75.
- NICHOLLS, A. O. 1991. Examples of the use of Generalised Linear Models in analysis of survey data for conservation evaluation. En: C. R. Margules & M. P. Austin (eds.) *Nature Conservation: Cost-effective Biological Surveys and Data Analysis*. pp 54-63. CSIRO.
- ODLAND, J. 1988. *Spatial Autocorrelation*. Sage Publications Inc.

- 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.
- PARKER, V. 1999. The use of logistic regression in modelling the distributions of bird species in Swaziland. *South African Journal of Zoology* 34: 39-47.
- PASCUAL, M. A. & IRIBARNE, O. O. 1993. How good are empirical predictions of natural mortality? *Fisheries Research* 16: 17-24.
- PEARCE, J. & FERRIER, S. 2000. An evaluation of alternative algorithm for fitting species distribution models using logistic regression. *Ecological Modelling* 128: 127-147.
- PEREIRA, J.M.C. & ITAMI, R.M. 1991. GIS-based habitat modelling using logistic multiple regression: a study of the Mt. Graham Red Squirrel. *Photogram Eng. Remote Sen.* 57: 1475-1486.
- PETERSON, A. T., SOBERÓN, J. & SÁNCHEZ-CORDERO, V. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285: 1265-1267.
- PICKETT, S.T.A., KOLASA, G. & JONES, C.G. 1994. *Ecological understanding: the nature of theory and the theory of Nature*. Academic Press, New York.
- PURVIS, A. & HECTOR, A. 2000. Getting the measure of biodiversity. *Nature* 405(6783): 212-219.
- RABINOWITZ, D., CAIRNS, S. & DILLON, T. 1986. Seven forms of rarity and their frequency in the flora of the British Isles. In *Conservation biology: the science of scarcity and diversity*. Soulé, M.E. (ed.) Sinauer associates, Sunderland, MA, pp182-204.
- RAMOS MA, LOBO JM & ESTEBAN M. 2001. Ten years inventorying the Iberian fauna: results and perspectives. *Biodiversity and Conservation* 10: 19-28.
- RICKLEFS, R.E. & SCHLUTER, D. (eds.) 1993. *Species diversity in ecological communities*. University of Chicago Press.
- SCOTT, J. M., DAVIS, F. W., CSUTI, B., NOSS, R., BUTTERFIELD, B., GROVES, C., ANDERSON, H., CAICCO, S., D'ERCHIA, F., EDWARDS, T.C., ULLIMAN, J. & WRIGHT, R.G. 1993. Gap analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs* 123: 1-41.
- SCOTT, J. M. & JENNINGS, M. D. 1997. A description of the National GAP Analysis Program. Informe técnico del U.S. Geological Survey, publicado en <http://www.gap.uidaho.edu/About/Overview/GapDescription/default.htm>.
- SCOTT, J. M. 1998. Large-area mapping of biodiversity. *Annals of the Missouri Botanical Garden* 85 (1): 34-47.
- SMITH, P. A. 1994. Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters* 4: 47-61.
- SMITH, A. T., BOITANI, L., BIBBY, C., BRACKETT, D., CORSI, F., DA FONSECA, G. A. B., GASCON, C., GIMENEZ DIXON, M., HILTON-TAYLOR, C., MACE, G., MITTERMEIER, R. A., RABINOVICH, J., RICHARDSON, B. J., RYLANDS, A., STEIN, B., STUART, S., THOMSEN, J. & WILSON, C. 2000. Databases tailored for biodiversity conservation. *Science* 290: 2073.
- SOBERÓN, J. M. & LLORENTE, B. J. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* 7: 480-488.
- SOBERÓN, J.M. & LLORENTE, J.B. & OÑATE, L. 2000. The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. *Biodiversity and Conservation* 9: 1441-1466.
- STOCKWELL, D. R. B. & PETERS, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13(2): 143-158.
- STOCKWELL, D.R.B. & PETERSON, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1-13.
- STATSOFT 1998. *STATISTICA for Windows*. Computer program manual. StatSoft, Inc. Tulsa, OK.
- SYSTEMATICS AGENDA 2000 1994. *Systematics Agenda 2000: Charting the Biosphere*. Technical Report. American Society Plant Taxonomist, Society of Systematics Biologist, Willi Hennig Society, Association of Systematics Collection.

- TEIXERA, J., FERRAND, N. & ARNTZEN, J.W. 2001. Biogeography of the golden-striped salamander *Chioglosa lusitanica*: a field survey and spatial modeling approach. *Ecography* 24: 618-623.
- TONTERI, T. 1994. Species richness of boreal understorey forest vegetation in relation to site type and successional factors. *Annali Zoologici Fennici* 31: 53-60.
- VAYSSIÈRES, M.P., PLANT, R.E. & ALLEN-DIAZ, B.H. 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11: 679-694.
- WALKER, P. A. & COCKS, P. A. 1991. HABITAT: A procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters* 1: 108-118.
- WHITTAKER, R. J., WILLIS, K. J. & FIELD, R. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography* 28: 453-470.
- WILLIAMS, P. H. & HUMPHRIES, C. J. 1996. Comparing character diversity among biotas. En: K. J. Gaston (ed.) *Biodiversity: A biology of numbers and difference*. pp 54-76. Blackwell Science.
- WILLOTT, S. J. 2001. Species accumulation curves and the measure of sampling effort. *Journal of Applied Ecology*, 38: 484-486.
- WOHLGEMUTH, T. 1998. Modelling floristic species richness on a regional scale: a case study in Switzerland. *Biodiversity and Conservation*: 159-177.
- ZIMMERMANN, N.E. & KIENAST, F. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science* 10: 469-482.

* Las láminas en color correspondientes a este trabajo se encuentran al final del volumen.