

J.M. Lobo

Departamento de Biogeografía y Cambio Global., Museo Nacional de Ciencias Naturales (CSIC), C/ José Gutiérrez Abascal 2, 28006, Madrid
Correio electrónico: mcnj117@mncn.csic.es

35

¿Debemos fiarnos de los modelos de distribución de especies?

Resultados clave

- Los resultados de los modelos de distribución están fuertemente condicionados por la calidad de la información de partida, lo que impide utilizar una muestra representativa capaz de generar predicciones útiles.
- Diversos tipos de factores condicionan la distribución de las especies, siendo necesario discriminar el efecto “puro” de los factores climáticos del que podría atribuirse a otras variables correlacionadas espacialmente con el clima.
- Los modelos de distribución que buscan representar la distribución actual de las especies, necesitan datos de ausencia fiables y predictores capaces de dar cuenta de los factores no climáticos que afectan a las distribuciones.
- Las simulaciones que solo utilizan la información ambiental proveniente de los datos de presencia, generan predicciones geográficas futuras de difícil validación, incapaces de representar la probabilidad de aparición de las especies.

Contexto

La carencia de datos sobre las modificaciones en la distribución y abundancia de los seres vivos debidas a los cambios climáticos, ha propiciado la utilización de diversas técnicas de modelización que, mediante el uso de los datos disponibles y diversos predictores, tratan de extrapolar la trayectoria que experimentarán las especies según diversos escenarios futuros. Estas modelizaciones de carácter correlacional se han impuesto en la literatura, tanto debido a su facilidad operativa como a la carencia de datos fisiológicos, demográficos o ecológicos que permitan obtener predicciones basadas en resultados

empíricos. Desde 1995 el número de trabajos publicados sobre este tema en revistas científicas indexadas ha crecido exponencialmente de modo que, actualmente, alrededor de un trabajo es publicado cada día (Lobo et al. 2010, Hortal et al. 2012) ¿Hasta qué punto son fiables estos pronósticos? ¿Es posible minimizar o limitar los sesgos, los errores y las incertidumbres de estas predicciones?

Sin ánimo de desacreditar las posibilidades de un enfoque correlacional a la hora de ofrecer predicciones útiles para anticipar y mitigar los efectos del cambio climático sobre las especies, revisaré los principales inconvenientes y defectos de esta aproximación con el propósito de arrojar luz sobre sus verdaderas posibilidades. Me apoyaré, para ello, en estudios recientemente publicados, principalmente realizados con datos ibéricos. Si para la gran mayoría de la biodiversidad apenas poseemos información parcial sobre su distribución actual y/o pasada, y hay un gran desconocimiento sobre sus preferencias ambientales y tolerancias fisiológicas ¿Cómo debemos usar estos datos para ofrecer estimas razonables sobre la modificación de los tamaños poblacionales y los límites de distribución ante el cambio climático previsto?

Mirando un mapa

Empecemos por un típico mapa de distribución tal y como se presenta en un atlas cualquiera (Figura 1). Se trata de la distribución de *Osmoderma eremita* (Scopoli), un coleóptero europeo de la familia Scarabaeidae cuyas larvas se desarrollan en árboles huecos y cuyas poblaciones parecen haber disminuido a lo largo de todo su rango de distribución. Por esa razón, la especie está protegida en la mayoría de los países europeos, y se le ha dado la más alta prioridad de acuerdo a la Directiva de Hábitats de la UE. Si examinamos su distribución

Figura 1.

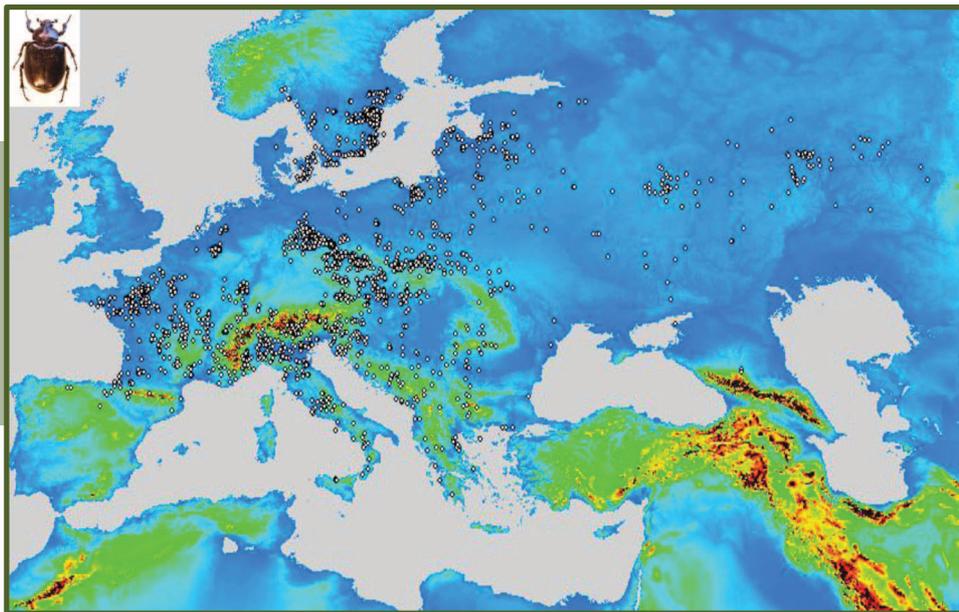


Figura 1. Distribución de *Osmoderma eremita* sobre un mapa topográfico de Europa.

Fuente: Elaborado a partir de Ranius et al. (2005).

(Ranius et al. 2005), observaremos varios detalles que suelen ser comunes en nuestras representaciones de las distribuciones de especies.

Por una parte, nos damos cuenta de que la densidad de los puntos con los que se suelen reflejar las localidades en donde se ha observado la especie es desigual. Esta densidad es alta en algunas zonas de Centroeuropa y los Alpes, pero baja en Europa del Este o Rusia. Evidentemente, la densidad de estos puntos no necesariamente refleja la frecuencia de aparición de la especie, sino que, probablemente, sigue criterios de oportunidad relacionados con el esfuerzo de colecta diferencial realizado en distintas regiones. De este modo, sesgos de prospección asociados a la cantidad de investigadores, densidad de población, disponibilidad de comunicaciones, etc. podrían estar influyendo sobre esta representación. Como en estos mapas y en las bases de datos que los han permitido generar, se obvian frecuentemente aquellas observaciones repetidas pertenecientes a una misma localidad, la variación espacial en el esfuerzo que percibimos debe ser todavía mucho mayor. Es probable que muchos de los puntos individuales de Centroeuropa representen en realidad conjuntos de observaciones.

Sin ánimo de ser exhaustivo, diversos trabajos han constatado desde hace tiempo, que los datos disponibles sobre la distribución geográfica de la gran mayoría de los organismos están sesgados. Que cuando se consideran conjuntamente, estos mapas reflejan, básicamente, el esfuerzo de colecta y no la distribución de la riqueza de especies, y que no permiten obtener una imagen fiel de las condiciones ambientales sobre las que pueden vivir las especies (p.ej., Dennis et al., 1999, Hortal et al. 2007 y 2008, Boakes et al. 2010, Rocchini et al. 2011, Ballesteros-Mejía et al. 2013, Yang et al. 2013 o Duputie et al. 2014). En nuestro país diversos estudios han mostrado que estas carencias y sesgos son la regla, sobre todo en grupos con una alta diversidad (Ferrer et al. 2006, Romo et al. 2006, Lobo et al. 2007,

Sánchez-Fernández et al 2011, Medina et al 2013). Las consecuencias de esta falta de calidad en los datos primarios son, a mi juicio, principalmente dos: una que atañe a lo que realmente muestran estas representaciones y otra que afecta a la gestión de los escasos recursos destinados a conocer la influencia del cambio climático sobre la distribución de nuestra fauna y flora. Como se ha demostrado matemáticamente hace poco (Aarts et al. 2012) y hemos ilustrado hace tiempo (Lobo 2008a), los resultados de estas predicciones no hacen sino mostrar la frecuencia de los propios puntos utilizados en el proceso de construcción del modelo (Figura 2), siendo altamente dependientes de la naturaleza y estructura de los datos de partida. Así, en la mayoría de las ocasiones, estamos prediciendo los propios datos incluidos en el modelo y las medidas de validación, inconsistentes y mal utilizadas, que pretenden avalar la coherencia de estas simulaciones (ver más adelante) ofrecen estimas de precisión también falseadas, al basarse en esos mismo datos sesgados. ¿Cómo vamos a obtener valores de probabilidad que reflejen la frecuencia de aparición de una especie si los datos de partida reflejan la frecuencia de colecta? Desde el punto de vista de la gestión, el no reconocimiento de estas carencias en la información biológica, implica destinar fondos y recursos a la creación de simulaciones basadas en la aplicación de técnicas de parametrización complejas sobre datos incompletos y sesgados, en vez de facilitar y promover estructuras estandarizadas de toma de datos biológicos (Observatorios de la Biodiversidad) que nos puedan permitir realizar un seguimiento y unas simulaciones coherentes y eficaces.

Sigamos mirando el mapa (Figura 1) y fijémonos en que existen regiones en las que no se ha observado a la especie y probablemente es cierto, como en el caso del Reino Unido o Córcega, y otras en las que la ausencia puede considerarse dudosa como en algunas zonas de Bulgaria, Ucrania o Azerbaiyán. La ausencia verdadera o altamente probable es un dato imprescindible para poder derivar la distribución de un organismo, pero no

■ **Figura 2.**



▲ **Figura 2.** Predicciones realizadas con MaxEnt (paneles de la derecha) utilizando datos de patrones irreales sin relación aparente con variables ambientales (paneles de la izquierda). Se han usado las opciones por defecto y las mismas variables ambientales utilizadas en el ejemplo proporcionado por el tutorial del programa para predecir estos patrones irreales. En el caso de la palabra “ERROR” el modelo se realizó utilizando únicamente una décima parte de los puntos que constituyen esa palabra. Estos simples ejemplos demuestran que es posible predecir prácticamente cualquier patrón espacial, utilizando algoritmos de modelización complejos y variables ambientales que no tengan relación causal con la variable dependiente.

Fuente: Elaboración propia.

todas las ausencias poseen el mismo valor (ver Lobo et al. 2010). La falta de observación de una especie en una localidad alejada ambiental y espacialmente de aquellas otras en las que se conoce su presencia, aporta poco al conocimiento de los patrones de distribución y sus causas. Información del tipo “no hay elefantes en la Antártida” es poco útil para estimar la distribución de los elefantes. Las ausencias reveladoras son aquellas próximas ambiental y geográficamente a las observaciones conocidas en las que aparece la especie pero ¿cómo obtenerlas?

Si la presencia de una especie en una localidad es una evidencia empírica, su ausencia es, en parte, una cuestión probabilística. Una intensificación progresiva del esfuerzo de colecta permite asegurar, cada vez con mayor certidumbre, que la especie de interés no está presente, pero ello requiere esfuerzos repetidos y generosos. En el caso de nuestro mapa, confiamos en las ausencias del Reino Unido porque sabemos que allí existe una gran tradición de estudio y mucho esfuerzo de colecta. Sin embargo, no estamos tan seguros de las ausencias en el caso de Ucrania o Azerbaiyán. Desafortunadamente, nuestro

análisis no puede ser más detallado porque los Atlas y las bases de datos de biodiversidad no suelen ofrecer información sobre el esfuerzo de colecta desarrollado en cada una de las localidades, y porque observaciones repetidas de una misma especie en una localidad se consideran redundantes y se omiten, generalmente, cuando sin embargo ofrecen una valiosa información sobre la frecuencia de aparición de una especie. Si una especie ha sido observada cien veces en una localidad es factible asumir que la ausencia de una especie hermana en esa misma localidad sea probable, pero ¿qué decir cuando se posee un único dato de presencia? Necesitamos bases de datos exhaustivas que recojan toda la información existente, pero también necesitamos discriminar para cada grupo de organismos donde se encuentran las regiones exploradas y las inexploradas. Es decir, elaborar una cartografía que represente el esfuerzo de colecta realizado o mejor, la probabilidad de obtener inventarios fiables (Rocchini et al. 2011). Así, aquellas localidades relativamente bien prospectadas en las que no se haya observado una especie podrán considerarse ausencias fiables o, en todo caso, podrá asignárseles un peso proporcional al esfuerzo realizado en ellas ¿Qué podemos predecir si no tenemos datos de ausencia fiables y cualquier localidad aparentemente no colonizada podría contener una presencia de la especie? ¿Cómo vamos a realizar predicciones fiables sobre la distribución de una especie si la información disponible refleja sesgos de muestreo desconocidos o no considerados?

Obtener ausencias fiables es, por tanto, esencial. Diversos métodos pueden permitirnos obtener esta información de nuestras bases de datos asumiendo que éstas reflejan, aunque sea parcialmente, el esfuerzo de colecta realizado. Las curvas de acumulación en las que se representa la creciente completitud en los inventarios según se incrementa el esfuerzo de colecta son una posibilidad (Chao & Jost 2012), como lo son diversos estimadores no-paramétricos (Hortal et al. 2006). En estos análisis puede utilizarse el número de registros como un sustituto del esfuerzo de colecta (Lobo 2008b) para discriminar las localidades con inventarios relativamente fiables de aquellas otras todavía insuficientemente muestreadas en las que se deben realizar prospecciones adicionales (Hortal & Lobo 2005, Aranda et al. 2011, Medina et al. 2013). La utilización de este procedimiento permite asumir que la ausencia de una especie en una localidad con inventarios fiables sea relativamente cierta y, por tanto, que este dato pueda utilizarse en modelos predictivos de distribución (Lobo & Martín-Piera 2002). Desafortunadamente, todavía no poseemos una herramienta de fácil manejo con capacidad para realizar estas selecciones en unidades espaciales de diferente resolución y sobre amplios territorios.

Por último, nuestro mapa-ejemplo nos permite reconocer también que existen fronteras más o menos difusas más allá de las cuales la especie no ha sido observada; límites en los que la frecuencia de aparición debe aproximarse a cero ¿Cuál puede ser la causa del límite de distribución septentrional? Como estamos tratando con una especie poiquiloterma (cuya temperatura fluctúa según la temperatura ambiental), podemos sospechar que este límite esté relacionado con la disminución del periodo climáticamente favorable, con las temperaturas mínimas o con la propia presencia de árboles en los que puedan desarrollarse las fases larvianas. ¿Por qué no se ha observado la especie en islas como Irlanda, Córcega o Mallorca? ¿Por qué no está presente en el Norte de

África, o en Turquía? ¿Por qué apenas ha colonizado la península ibérica cuando en Italia o Grecia se la observa en localidades meridionales bajo similares condiciones templado-cálidas? La capacidad de colonización, la dispersión desde posibles refugios Pleistocénicos o la alteración del hábitat pueden ser explicaciones razonables a estas preguntas, pero el corolario es que las causas de los límites de distribución no son, muy probablemente, las mismas a lo largo del rango de distribución de una especie. ¿Cómo podemos entonces estimar correctamente los límites de distribución si no acudimos a los factores que permiten explicar cada frontera? ¿Podemos realizar predicciones razonables ante escenarios de cambio climático sin considerar el peso de estos otros factores no-climáticos?

■ Presente, ausente y en equilibrio

Las especies viven en sitios favorables, pero no en todos. Esta afirmación supone: i) que existen unas condiciones ambientales para cada especie en las que el crecimiento demográfico es positivo (su “nicho”) y otras en las el crecimiento demográfico neto es negativo o inexistente, y ii) que los limitantes de dispersión, los procesos históricos y otros efectos contingentes impiden o dificultan constantemente los procesos de colonización.

El primer supuesto soslaya que pueden existir presencias en zonas relativamente desfavorables (Pulliam 2000), pero también el papel de la plasticidad fenotípica, la aclimatación (Wilson & Franklin 2002) o la microevolución por selección natural a condiciones locales (Ellner 2013). Pero además, el razonamiento que lo sustenta es circular. Nuestra carencia de datos fisiológicos y experimentales para la gran mayoría de las especies, nos ha conducido por el camino fácil de medir el “nicho” de una especie a través de las condiciones ambientales existentes en los lugares donde se la encuentra. Pero la distribución no tiene por qué reflejar el “nicho” debido a los procesos contingentes anteriormente mencionados (ver Figura 3). No es de extrañar, entonces, que la especie habite en los sitios favorables, si esta favorabilidad ha sido previamente establecida acudiendo a los sitios habitados.

El segundo supuesto está relacionado con la falta de equilibrio de las distribuciones con las condiciones ambientales y, en especial, con el clima. Este desequilibrio está en la propia base de las cuestiones que han dado lugar a la fundación de la Biogeografía como ciencia (Hortal et al. 2012) ¿Por qué lugares alejados con similares condiciones climáticas poseen floras funcionalmente semejantes pero filogenéticamente dispares? La ausencia de una especie puede deberse a diferentes motivos (Lobo et al. 2010), siendo la falsa ausencia debida a los sesgos de colecta la que más inconvenientes causa a la hora de establecer modelos de distribución fiables (Fig. 3). Las ausencias “contingentes”, que aparecen como consecuencia de la falta de equilibrio en las distribuciones (debido a la dispersión, factores históricos, interacciones bióticas, etc.), son reales y obligan a que los modelos tengan que incluir predictores capaces de dar cuenta de ellas (Sutherst 2014). Si las localidades favorables pero no colonizadas poseen condiciones climáticas particulares, el “nicho” extraído de los datos de distribución será incompleto. Además, cuando la ubicación de estas localidades sigue un gradiente climático o ambiental,

Figura 3.

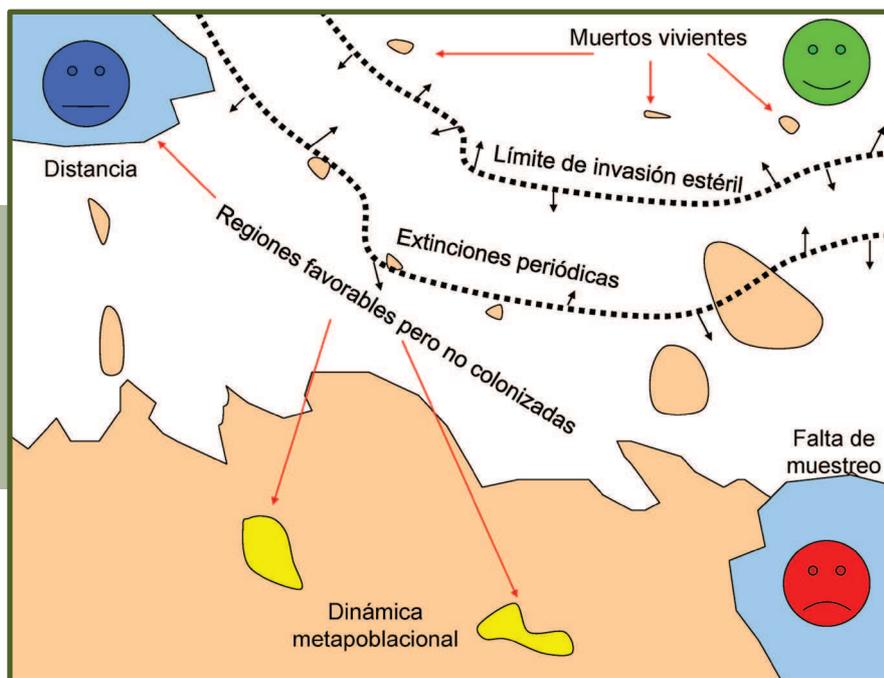


Figura 3. Esquema ejemplificando una hipotética distribución. Dentro del área de distribución más o menos continua de una especie pueden existir zonas favorables inhabitadas debido a la dinámica metapoblacional (en amarillo), zonas favorables que se consideren como ausencias (en azul), bien debido a la falta de prospección (ausencias metodológicas; "smile" rojo) o a la actuación de limitantes de dispersión (ausencias contingentes; "smile" azul). Puede haber incluso presencias en zonas relativamente desfavorables, en las que las tasas de crecimiento demográfico neto son negativas y las poblaciones se mantienen gracias a la inmigración desde las zonas favorables hasta más allá del límite de invasión estéril ("muertos vivos"). Las ausencias climáticas o ambientales ("smile" verde) aparecen a partir de esta zona. Téngase en cuenta que las zonas favorables pero no colonizadas pueden poseer condiciones ambientales o climáticas particulares no representadas en el área ocupada, de modo que el "nicho" derivado de la distribución sea menor que el real.

Fuente: Elaboración propia.

la covariación entre clima y posición espacial impedirá estimar inequívocamente el efecto climático. Si una especie es endémica de las sierras del sur de la península ibérica, la posición espacial y las condiciones climáticas covarían ¿Cuáles son en este caso las variables capaces de explicar su distribución? ¿Las condiciones climáticas imperantes en las localidades con observaciones actuales o los factores históricos y de otro tipo que han promovido su presencia en esta zona? (ver Linares et al. 2011).

La existencia de equilibrio se ha esgrimido frecuentemente como un requisito fundamental para realizar modelos de distribución utilizando variables climáticas (Pearson & Dawson 2003). Las especies deben de estar en un estado de equilibrio o pseudo-equilibrio con las condiciones climáticas. Es decir, deben de haber ocupado la mayor parte de sus condiciones favorables. Este argumento resulta clarificador y engañoso a la vez. Clarificador, porque reconoce que los limitantes de dispersión o las interacciones bióticas afectan a las distribuciones de los organismos y dificultan encontrar parámetros adecuados para expresar correctamente la función del clima. Pero engañoso, porque la falta de equilibrio es probablemente la norma en la distribución de los organismos (Araújo & Pearson 2005, Monahan 2009), y porque otro tipo de factores no climáticos y circunstanciales pueden ser los causantes de los límites de distribución, ofreciendo la falsa impresión de equilibrio.

Tanto la distribución de los organismos unicelulares (Fontaneto & Hortal 2013) como la del hombre están influidas por los limitantes de dispersión: ha habido poblaciones humanas en el Ártico pero no en la Antártida, hasta recientemente. El requerimiento de

equilibrio con las condiciones climáticas es una exigencia incompatible con los datos reales. Erróneamente se asume que el patrón de variación espacial en las distribuciones puede atribuirse en su totalidad a las variaciones climáticas, sin considerar que el aparente efecto del clima puede deberse a otros factores con una estructura espacial similar a la del clima. Es el clásico dilema entre correlaciones y causas (Shipley 2000), magnificado en el caso de las distribuciones de especies por la falta de independencia y la covariación en la estructura espacial de los distintos tipos de predictores relevantes. Algunos autores han tratado de discriminar el efecto puro del clima del debido a otros factores simplemente incluyendo covariables como el rango altitudinal (Lassueur et al. 2006, Austin & Van Niel 2011), utilizando regresiones parciales tras incluir otros predictores relevantes (Aragón et al. 2010a, Márquez et al. 2011, Muñoz et al. 2013, Real et al. 2013) o comparando la capacidad predictiva diferencial de distintos tipos de modelos (García-Valdés et al. 2013). Como era de esperar, el efecto "puro" del clima disminuye de manera importante cuando se utilizan estos procedimientos minimizándose, por tanto, las tasas de cambio en las distribuciones predichas ante escenarios de cambio climático. ¿Cuáles son las predicciones más cercanas a la realidad? Pocos estudios han tratado de evaluar este tipo de modelos correlacionales utilizando datos de distribución provenientes de un periodo distinto. Cuando así se hace (Araújo et al. 2005, Kharouba et al. 2009, Rubidge et al. 2011, Eskildsen et al. 2013, Ko et al. 2013, Watling et al. 2013), los resultados constatan con claridad la limitada capacidad de esta aproximación, pero también la necesidad de incluir otras variables no climáticas, la sobrevaloración de los resultados obtenidos cuando la evaluación se realiza sobre los datos de un único

periodo, o la incapacidad de las medidas de validación al uso para dar cuenta de importantes diferencias en las representaciones geográficas.

Tener en cuenta la falta de equilibrio de las distribuciones con el clima requiere considerar datos de ausencia fiables y predictores capaces de dar cuenta de esas ausencias. No puede ser de otro modo. Pero también requiere poseer una muestra representativa de datos que refleje el gradiente ambiental que se pretende predecir, precaución a la hora de elegir las variables explicativas que acompañarán a las variables climáticas, así como limitar la extensión del área de estudio a la que es menos probable que actúen estos limitantes de dispersión (Acevedo et al. 2012). Un modelo correlacional capaz de predecir con éxito la probabilidad de aparición de una especie en localidades desconocidas, pero incluidas dentro del tiempo y el rango de condiciones de los datos observados (interpolación), no necesita obligadamente predictores realmente causales. Una muestra representativa que incluya datos de presencia-ausencia, abundancia o frecuencia y un mecanismo de validación eficiente pueden ser suficientes.

La variedad de técnicas de modelización y simulación existentes permite obtener predicciones cabales siempre y cuando la muestra sea representativa de la población objeto de estudio. Si podemos predecir cuál será la composición del parlamento efectuando un número reducido de entrevistas bien repartidas entre las distintas características de los votantes (estrato social, provincia, sexo, etc.). ¿Por qué no vamos a poder predecir la distribución de una especie si poseemos datos fiables (presencia y ausencia) bien repartidos a lo largo de su territorio? Ahora bien, realizar un modelo correlacional en el que la función predictiva sea capaz de ofrecer estimas fiables en otro tiempo o más allá de las condiciones de los datos de partida (extrapolación), como es el caso de las predicciones en escenarios climáticos futuros, requiere inevitablemente conocer el verdadero efecto del clima sobre la distribución de las especies. La covariación espacial entre el clima y otro tipo de factores será siempre la norma en estas situaciones, por lo que las representaciones geográficas deberían siempre reflejar la incertidumbre de las predicciones según sea el peso que se puede atribuir al clima: desde el mínimo efecto “puro” de las variables climáticas, al máximo efecto “combinado” que el clima puede representar cuando otras variables predictoras covarían con él y no puede discernirse su efecto incontrovertiblemente.

■ Evaluación

Si los resultados de modelos de distribución han de considerarse con mucha precaución ¿Por qué la literatura está repleta de trabajos en los que las predicciones de estos modelos se consideran eficientes? La incongruencia se explica cuando se estudian las medidas utilizadas en el proceso de evaluación. Como el resultado de un modelo de distribución es una variable continua y nuestros datos de partida suelen ser binarios (presencia-ausencia), es necesario encontrar un punto de corte por encima y por debajo del cual se considere que la especie está presente o ausente, respectivamente. El punto de corte ideal está relacionado con la prevalencia de la especie o proporción de localidades habitadas (Acevedo & Real 2012), un parámetro generalmente desconocido que

podría estimarse a partir de un muestreo representativo. La aplicación de este punto de corte permite elaborar una matriz de confusión en la que es posible computar diversas métricas capaces de medir nuestro acierto en las predicciones (Fielding & Bell 1997). El área bajo la curva (AUC) derivada de los diagramas ROC (*receiver operating characteristic*) es, sin duda, la más conocida y utilizada medida para la evaluación de estos modelos. En estas curvas se grafica la relación entre la frecuencia de acierto en las presencias respecto a la frecuencia de falsos positivos (ausencias erróneamente predichas como presencias o errores de comisión) para diferentes puntos de corte, y el área bajo esa relación es considerada una medida de la capacidad del modelo independientemente del punto de corte que se utilice.

El uso de AUC como medida de evaluación en los modelos de distribución tiene varios inconvenientes (Lobo et al. 2008, Jiménez-Valverde 2012), pero el más importante es que la tasa de acierto en las ausencias aumenta a medida que incrementamos el área de estudio. Para una misma especie, aumentar el área de estudio implica disminuir su prevalencia e incrementar las posibilidades de acertar correctamente. Predecir que los elefantes no se encuentran en la Antártida no tiene mérito y si nuestra medida computa este resultado como un acierto nos está ofreciendo una imagen sobrevalorada de poder predictivo. Buena parte del éxito de los modelos de distribución correlacionales, surge con la publicación en 2006 del trabajo de Jane Elith y colaboradores, en el que se compara la eficacia de distintas técnicas utilizando AUC como medida principal de evaluación. Seis años más tarde y utilizando los mismos datos, uno de los coautores de aquel trabajo (Hijmans 2012) reanaliza los resultados, corrigiendo el efecto de la distancia geográfica entre los datos utilizados para realizar y validar el modelo. Los resultados obtenidos muestran que estos modelos son claramente inconsistentes y que su capacidad predictiva dista mucho de ser la que inicialmente se creía, pues las presencias y las ausencias elegidas al azar suelen tener valores de probabilidad similares. Las implicaciones de este estudio pueden ampliarse a otras medidas de evaluación frecuentemente utilizadas como el índice kappa o el *true skill statistics* y, hasta ahora, apenas han tenido repercusión en la literatura, aunque suponen un serio varapalo a la confianza en los resultados de los modelos de distribución de especies.

Una última cuestión de interés en la validación de los modelos de distribución es que su eficacia suele medirse, casi exclusivamente, considerando su poder de discriminación. Es decir, su capacidad para separar las ausencias de las presencias tras elegir un punto de corte sobre los valores continuos predichos. Sin embargo, la efectividad de un modelo puede también medirse considerando si las probabilidades de presencia predichas reflejan la proporción de presencias observadas (calibración o fiabilidad). Una probabilidad predicha de 0.3 en una localidad significaría que en el 30% de las observaciones debería aparecer la especie. En este caso no es necesario acudir a un punto de corte y la calidad del modelo se estima acudiendo a todo el conjunto de valores continuos obtenidos. La calidad de los modelos de distribución de especies apenas ha sido evaluada teniendo en cuenta las distintas métricas que miden la fiabilidad, cuando estas medidas serían mucho más eficaces a la hora de decidir la eficiencia de un modelo que deseamos transferir en el espacio o en el tiempo (Jiménez-Valverde et al. 2013).

■ Probable y potencial

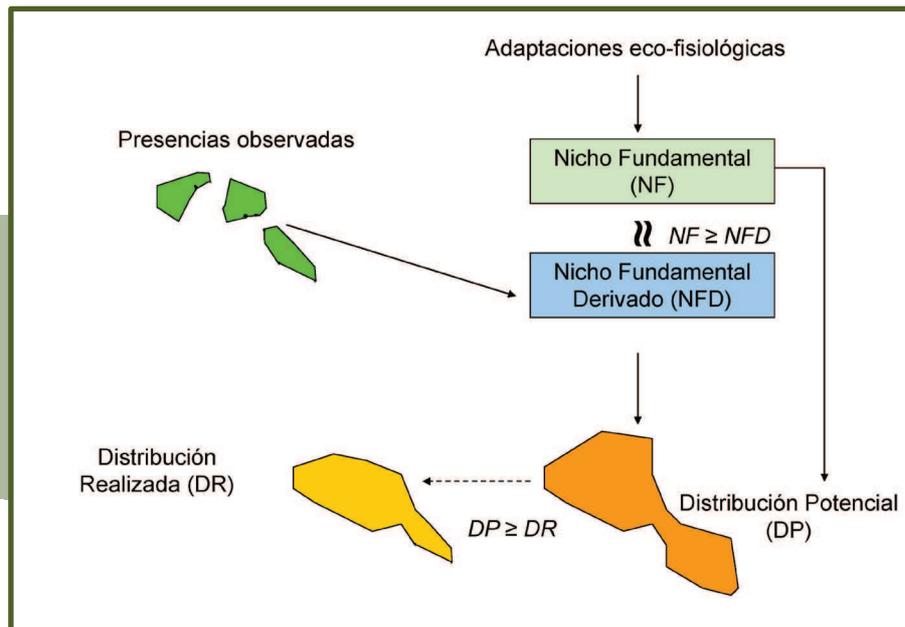
Si las condiciones climáticas de las zonas en las que una especie está ausente pueden ser favorables y esas ausencias son frecuentemente debidas a factores no climáticos ¿necesitamos información fiable de las ausencias para predecir cuál puede ser la distribución en el futuro? Como hemos visto, las ausencias son necesarias para poder considerar los factores que generan el desequilibrio de las distribuciones con el clima. Ahora bien, teniendo en cuenta lo difícil que es encontrar ausencias fiables en algunos grupos o regiones ¿podemos utilizar únicamente la información que proporcionan las presencias para elaborar predicciones de distribución razonables?

La carencia de información fiable sobre las ausencias ha propiciado el uso de los denominados *presence-only-models* en los que, en vez de “verdaderas” ausencias se utilizan pseudoausencias, en la mayoría de las ocasiones seleccionadas al azar dentro del territorio considerado (*background absences*). Se trata del clásico procedimiento que compara las condiciones ambientales de los datos de presencia disponibles (uso) respecto a las existentes en todo el territorio considerado (disponibilidad). Son las denominadas *Resource Selection Functions* (Johnson 1980) introducidas en los años 80 para estimar la probabilidad de aparición de una especie en distintos tipos de hábitats. En la actualidad, buena parte de los programas más exitosos en la elaboración de modelos predictivos de distribución realizan esta comparación para ofrecer sus predicciones y la literatura está repleta de estudios que realizan esta práctica. Aunque se conocía que el efecto de la utilización de diversos tipos de pseudoausencias, más o menos alejadas ambiental o espacialmente del universo de las presencias, ejercía una importante influencia sobre las representaciones obtenidas (Chefaoui & Lobo 2008), recientemente se ha demostrado matemáticamente que este tipo de procedimientos no pueden estimar la probabilidad de

aparición de una especie (Hastie & Fithiam 2013) y que sólo reflejan la intensidad de los propios datos de partida (Aarts et al. 2012). Cuando realizamos un modelo de este tipo el resultado final se parece sospechosamente demasiado a los propios datos de partida (Figura 2) y las evaluaciones con esos mismos datos ofrecen poderes de discriminación inevitablemente altos.

Las localidades en la que se ha observado una especie nos ofrecen, sin embargo, una información valiosa sobre las condiciones en las que ésta puede habitar (Figura 4). Sabemos que, en muchos casos, las colectas que hemos realizado están lejos de seguir los protocolos de un muestreo estandarizado clásico (por ejemplo, muestreos al azar o estratificados) y sabemos también que los datos de presencia son, en ocasiones, insuficientes y sesgados. A pesar de ello, y conociendo que existen localidades en las que puede observarse una especie sin que exista un crecimiento poblacional neto positivo (Figura 3), esta información es, en la inmensa mayoría de las ocasiones, la única fuente de conocimiento empírico sobre las condiciones en las que una especie puede habitar. Cuando se representan geográficamente estas condiciones se obtiene una inferencia de distribución potencial: una imagen provisional de las localidades habitables por una especie cuando el conjunto de factores contingentes que impide la colonización de ciertas regiones deja de actuar. Al contrario que en el caso de la distribución “real”, actual o realizada, la cual necesita ausencias fiables y predictores capaces de dar cuenta de esas ausencias para ofrecer estimas creíbles (Jiménez-Valverde et al. 2008), la distribución potencial no puede validarse utilizando datos reales de distribución, pues estos siempre pueden estar influidos por los factores contingentes que impiden la aparición de la especie en localidades *a priori* favorables. Las predicciones acerca de una distribución realizada son estimas de probabilidad (probable es aquello con razones fundadas para suponer que se verificará o sucederá), mientras que las predicciones sobre una distribución

■ Figura 4.



▲ **Figura 4.** Datos fisiológicos y ecológicos básicos que podrían ayudar a determinar el “nicho” de las especies, concretamente el nicho térmico. Sin embargo, las condiciones ambientales de las localidades en las que se ha observado una especie solo nos proporcionan una imagen parcial de ese nicho (NFD) que, cuando se representa en el espacio geográfico, nos ofrece una imagen de la distribución potencial de la especie.

Fuente: Elaboración propia.

potencial son posibles (que pueden suceder o existir). La distribución potencial climática de una especie podría inferirse utilizando datos de presencia-ausencia, pero para ello se necesita minimizar el peso de aquellos factores contingentes capaces de propiciar la ausencia de una especie bajo condiciones favorables (Acevedo et al. 2012). Otra aproximación para estimar, al menos parcialmente, el conjunto de localidades favorables según unas determinadas condiciones ambientales consistiría simplemente en discriminar aquellas localidades con condiciones similares a los lugares de presencia observada. Los datos geográficos ofrecerían una representación parcial del “nicho” de una especie que podría trasladarse al espacio geográfico siguiendo un simple proceso de intersección generalizada de conjunto o *Multidimensional Enveloping* (Jiménez-Valverde et al. 2011)

Como la presencia es, en la mayoría de las ocasiones y para muchos grupos de organismos, la única información disponible, los resultados de aplicar este procedimiento para estimar la distribución potencial de los organismos pueden servir para ofrecer mapas de riesgo en el caso de especies invasoras (Aragón et al. 2010b), decidir la ubicación de nuevas áreas a prospectar (Sánchez-Fernández et al. 2011a), estimar el grado de equilibrio en las distribuciones (Sánchez-Fernández et al. 2012) o predecir los posibles efectos del cambio climático sobre las distribuciones (Aragón & Lobo 2012). En todos estos casos las predicciones no son, en realidad, escenarios que reflejen la probabilidad de aparición de la especie, sino escenarios que representan la posibilidad de que las localidades posean condiciones similares a las que se dan en las localidades en las que se ha observado a la especie. Varias fuentes de incertidumbre afectan a los resultados de estas simulaciones, siendo las principales la identidad y el número de las variables explicativas que se utilicen (Beaumont et al. 2005) o la falta de datos geográficos capaces de maximizar la representación del “nicho” de las especies (Sánchez-Fernández et al. 2011b). En esta aproximación es clave incluir el mayor espectro de localidades con observaciones de presencia disponibles, tal vez incluso incorporando aquellas provenientes de los lugares ocupados en los que no hay crecimiento demográfico neto positivo (Figura 3). Se trata, básicamente, de identificar las condiciones en los límites ambientales y/o geográficos de la distribución, en vez de establecer curvas de respuesta basadas en unas frecuencias de observación a menudo influenciadas por los sesgos de colecta. En este caso, sólo se deben eliminar los registros de presencia claramente equivocados debido a una errónea adscripción taxonómica o a un fallo en la georeferenciación, pero se debería evitar cualquier mecanismo de detección automática de *outliers* que rechace datos capaces de reflejar los límites de tolerancia de las especies.

■ En conclusión

El repaso efectuado arroja dudas sobre la capacidad de estas técnicas de modelización para ofrecer predicciones útiles ante el cambio climático. Conviene recordar que el establecimiento de correlaciones y el estudio de patrones, es la base de gran parte del conocimiento científico y que, en muy contadas ocasiones, somos capaces de descubrir las causas últimas de un fenómeno natural. Podemos utilizar los modelos

de distribución de especies y debemos usarlos porque, a corto y medio plazo, no vamos a tener otra fuente de conocimiento sobre las respuestas ante el cambio climático para la gran mayoría de la biodiversidad. Eso sí, debemos aplicar estos modelos correlacionales conociendo sus limitaciones e incapacidades y, para ello, es conveniente repasar las lecciones generalmente olvidadas que nos ha proporcionado la larga historia de un empeño científico destinado a conocer y predecir la distribución de los organismos (Sutherland 2014). Buscando la opinión de los que conocen la historia natural de las especies, debemos buscar el efecto único o puro del clima, introduciendo otras variables que tengan sentido biológico o biogeográfico. Debemos utilizar información fiable sobre las ausencias y circunscribir el área de trabajo a aquella en la que los limitantes de dispersión tengan una menor probabilidad de actuar. Si, desgraciadamente, no podemos conseguir ausencias, siempre podremos trabajar únicamente con los datos de presencia, pero evitando aproximaciones que reflejen el propio patrón de distribución de los datos de partida. Y no sería mala cosa mostrar siempre cual es la distribución espacial de la incertidumbre en nuestras predicciones. A fin de cuentas, la estimación del error es, probablemente, una de las señas de identidad del trabajo científico.

El incremento exponencial en la capacidad de cómputo de nuestros ordenadores, junto a la disponibilidad de información ambiental y biológica digitalizada, han significado una oportunidad de estudio y una explosión de trabajos. Buena parte del esfuerzo realizado en este campo hasta ahora, ha tratado de encontrar la técnica o el algoritmo de modelización con mayor capacidad. Tengo el convencimiento de que estas cuestiones no son excesivamente relevantes y que el esfuerzo debe orientarse hacia mejorar la calidad de los datos primarios con los que se alimentan este tipo de modelos. Buena parte de estos datos se encuentran dispersos en la literatura y en las propias “redes neuronales” de los naturalistas. A veces, los propios investigadores conocen las localidades en las que no se ha colectado una especie, a pesar de haber realizado un esfuerzo de colecta importante (nadie publica un trabajo que lleve por título “las especies que no he colectado”). Necesitamos una administración valiente e informada, capaz de promover la digitalización exhaustiva de la información faunística y florística recogida durante décadas y, capaz también, de establecer una red de observatorios sobre la biodiversidad que nos permita recabar nuevos datos y vigilar los cambios.

Se ha sugerido que si se utilizasen estos modelos correlacionales siguiendo otras pautas, la estima de la magnitud del cambio debida al clima sería mucho menor de la que nos ofrecen ahora. Tal vez los mensajes excesivamente alarmistas no sean los más indicados para promover estrategias y actuaciones eficaces. Si una parte de nuestra fauna y flora va a quedar relegada a las cimas de las montañas ibéricas, si los cambios van a ser tan drásticos como los que nos muestran las publicaciones recientes del Plan Nacional de Adaptación al Cambio Climático ¿No tenemos las manos atadas? ¿No exigen estas previsiones respuestas de tanto calibre que es difícil y costoso encontrar estrategias de mitigación eficaces? Nuevos modelos correlacionales capaces de ofrecer estimas más ajustadas al posible papel del clima en la distribución de los organismos, y que proporcionen medidas de incertidumbre, podrían ayudar a proponer medidas de actuación y mitigación dimensionadas ante el impacto del cambio climático.

■ Agradecimientos

Este trabajo se ha beneficiado de las sugerencias y comentarios de Luis M. Carrascal y Joaquín Hortal, así como de los proporcionados por el editor y un evaluador anónimo.

■ Referencias bibliográficas

- Aarts G, Fieberg J, Matthiopoulos J (2012) Comparative interpretation of count, presence–absence and point methods for species distribution models. *Methods in Ecology and Evolution* 3:177-187
- Acevedo P, Real R (2012) Favorability: concept, distinctive characteristics and potential usefulness,” *Naturwissenschaften* 99:515-522
- Acevedo P, Jiménez-Valverde A, Lobo JM, Real R (2012) Delimiting the geographical background in species distribution modelling. *Journal of Biogeography* 39:1383-1390
- Aragón P, Baselga A, Lobo JM (2010b) Global estimation of invasion risk zones for the western corn rootworm *Diabrotica virgifera virgifera*: integrating distribution models and physiological thresholds to assess climatic favourability. *Journal of Applied Ecology* 47:1026-1035
- Aragón PA, Lobo JM, Olalla-Tárraga MA, Rodríguez MA (2010a) The contribution of contemporary climate to ectothermic and endothermic vertebrate distributions in a glacial refuge. *Global Ecology and Biogeography* 19:40-49
- Aranda SC, Gabriel R, Borges PAV, Brito de Azevedo E, Lobo JM (2011) Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores). *Bryologist* 114:611–624
- Araújo MB, Pearson RG (2005) Equilibrium of species’ distributions with climate. *Ecography* 28:693-695
- Araújo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species–climate impact models under climate change. *Global Change Biology* 11:1504-1513
- Austin MP, Van Niel P (2011) Improving species distribution models for climate change studies. Variable selection and scale. *Journal of Biogeography* 38:1-8
- Ballesteros-Mejía L, Kitching IJ, Jetz W, Nagel P, Beck J (2013) Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. *Global Ecology and Biogeography* 22:586-595
- Beaumont LJ, Hughes L, Poulsen M (2005) Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species’ current and future distributions. *Ecological Modelling* 186:250-269
- Boakes EH, McGowan PJK, Fuller RA, Ding C, Clark NE, O’Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biology* 8:e1000385
- Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533-2547

- Chefaoui RM, Lobo JM (2008) Assessing the effects of pseudoabsences on predictive distribution model performance. *Ecological Modelling* 210:478-486
- Dennis RLH, Sparks TH, Hardy PB (1999) Bias in butterfly distribution maps: the effects of sampling effort. *Journal of Insect Conservation* 3:33-42
- Duputie A, Zimmermann NE, Chuine I (2014). Where are the wilds things? Why we need better data on species distribution. *Global Ecology and Biogeography* (en prensa)
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151
- Ellner SP (2013) Rapid evolution: from genes to communities, and back again? *Functional Ecology* 27:1087-1099
- Eskildsen A, le Roux PC, Heikkinen RK, Høye TT, Kissling WD, Pöyry J, Wisz MS, Luoto M (2013) Testing species distribution models across space and time: high latitude butterflies and recent warming. *Global Ecology and Biogeography* 22:1293-1303
- Ferrer X, Carrascal LM, Gordo O, Pino J (2006) Bias in avian sampling effort due to human preferences: an analysis with Catalanian birds. *Ardeola* 53:213-227
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38-49
- Fontaneto D, Hortal J (2013) At least some protist species are not ubiquitous. *Molecular Ecology* 22:5053-5055
- García-Valdes R, Zavala MA, Araújo MB, Purves DW (2013) Chasing a moving target: projecting climate change-induced shifts in non-equilibrium tree species distributions. *Journal of Ecology* 101:441-453
- Hastie T, Fithian W (2013) Inference from presence-only data; the ongoing controversy. *Ecography* 36:864-867
- Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93:679-688
- Hortal J, Borges PAV, Gaspar C (2006) Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology* 75:274-287
- Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117:847-858
- Hortal J, Lobo JM (2005) An ED-based protocol for optimal sampling of biodiversity. *Biodiversity and Conservation* 14:2913-2947
- Hortal J, Lobo JM, Jiménez-Valverde A (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). *Conservation Biology* 21:853-863.
- Hortal J, Lobo JM, Jiménez-Valverde A (2012) Basic questions in Biogeography and the (lack of) simplicity of species distributions: Putting Species Distribution Models in the right place. *Natureza & Conservação* 10:108-118
- Jiménez-Valverde A (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21:498-507
- Jiménez-Valverde A, Acevedo P, Barbosa AM, Lobo JM, Real R (2013) Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecology and Biogeography* 22:508-516
- Jiménez-Valverde A, Lobo JM, Hortal J (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions* 14:885-890
- Jiménez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragón P, Lobo JM (2011) Use of niche models in invasive species risk assessments. *Biological Invasions* 13:2785-2797
- Johnson DH (1980) The comparison of usage and availability measurements for evaluating resource preference. *Ecology* 61:65-71
- Kharouba HM, Algar AC, Kerr JT (2009) Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment. *Ecology* 90:2213-2222
- Ko CY, Ko CJ, Lin RS, Lee PF (2013) Influences of temporal independence of data on modelling species distributions. *Basic and Applied Ecology* 14:309-319
- Lassueur T, Joost S, Randin CF (2006) Very high resolution digital elevation models: do they improve models of plant species distribution? *Ecological Modelling* 198:139-153
- Linares JC, Carreira JA, Ochoa V (2011) Human impacts drive forest structure and diversity. Insights from Mediterranean mountain forest dominated by *Abies pinsapo* (Boiss.). *European Journal of Forest Research* 130:533-542
- Lobo JM (2008a) More complex distribution models or more representative data? *Biodiversity Informatics* 5:14-19

- Lobo JM (2008b) Database records as a surrogate for sampling effort provides higher species richness estimations. *Biodiversity and Conservation* 17:873-881
- Lobo JM, Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF (2007) How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Diversity and Distributions* 13:772-780
- Lobo JM, Jiménez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33:103-114
- Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17:145-151
- Lobo JM, Martín-Piera F (2002) Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. *Conservation Biology* 16:158-173
- Márquez AL, Real R, Olivero J, Estrada A (2011) Combining climate with other influential factors for modelling climate change impact on species distribution. *Climatic Change* 108:135-157
- Medina NG, Lara F, Mazimpaka V, Hortal J (2013) Designing bryophyte surveys for an optimal coverage of diversity gradients. *Biodiversity and Conservation* 22:3121-3139
- Monahan WB (2009) A mechanistic niche model for measuring species' distributional responses to seasonal temperature gradients. *PLoS ONE* 4:e7921
- Muñoz A-R, Márquez AL, Real R (2013) Updating Known Distribution Models for Forecasting Climate Change Impact on Endangered Species. *PLoS ONE* 8:e65462
- Pearson RG, Darwon TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12:361-371
- Pulliam RH (2000) On the relationship between niche and distribution. *Ecology Letters* 3:349-361
- Ranius T, Aguado LO, Antonsson K, Audisio P, Ballerio A, Carpaneto GM, Chobot K, Gjurašin B, Hanssen O, Huijbregts H, Lakatos F, Martin O, Neculiseanu Z, Nikitsky NB, Paul W, Pirnat A, Rizun V, Ruicnescu A, Stegner J, Süda I, Szwako P, Tamutis V, Telnov D, Tsinkevich V, Versteirt V, Vignon V, Vögeli M, Zach P (2005) *Osmoderma eremita* (Coleoptera, Scarabaeidae, Cetoniinae) in Europe. *Animal Biodiversity and Conservation* 28:1-14
- Real R, Romero D, Olivero J, Estrada A, Márquez AL (2013) Estimating How Inflated or Obscured Effects of Climate Affect Forecasted Species Distribution. *PLoS ONE* 8:e53646
- Rocchini D, Hortal J, Lengyel S, Lobo JM, Jiménez-Valverde A, Ricotta C, Bacaro G, Chiarucci A (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography* 35:211-226
- Romo H, García-Barros E, Lobo JM (2006) Identifying recorder-induced geographic bias in an Iberian butterfly database. *Ecography* 6:883-875
- Rubidge EM, Monahan WB, Parra JL, Cameron SE, Brashares JS (2011) The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. *Global Change Biology* 17:696-708
- Sánchez-Fernández D, Lobo JM, Abellán P, Millán A (2011a) How to identify future sampling areas when information is biased and scarce: An example using predictive models for species richness of Iberian water beetles. *Journal for Nature Conservation* 19:54-59
- Sánchez-Fernández D, Lobo JM, Hernández-Manrique OL (2011b) Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. *Diversity and Distributions* 17:163-171
- Sánchez-Fernández D, Lobo JM, Millán A. & Ribera I (2012) Habitat type mediates equilibrium with climatic conditions in the distribution of Iberian diving beetles. *Global Ecology and Biogeography* 21:988-997
- Shipley B (2000) *Cause and Correlation in Biology*. Cambridge, Cambridge University Press.
- Sutherst R (2014) Pest species distribution modelling: origins and lessons from history. *Biological Invasions* 16:239-256
- Watling JI, Bucklin DN, Speroterra C, Brandt LA, Mazzotti FJ, Romañach SS (2013) Validating Predictions from Climate Envelope Models. *PLoS ONE* 8:e63600
- Wilson RS, Franklin CE (2002) Testing the beneficial acclimation hypothesis. *Trends in Ecology and Evolution* 17:66-70
- Yang W, Ma K, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *Journal of Biogeography* 40:1415-1426