LEAD ARTICLE

EL MODELADO DE NICHOS Y DISTRIBUCIONES: NO ES SIMPLEMENTE "CLIC, CLIC, CLIC"

Robert P. Anderson

INTRODUCCIÓN

Ofrezco este breve ensayo con la idea de resaltar que, a pesar de ser un campo reciente y en pleno desarrollo, el modelado de nicho ecológico (y sus aplicaciones para estudios de distribuciones geográficas de especies) involucra tanto aspectos teóricos importantes como metodologías ya bastante sofisticadas. ensayo trata específicamente sobre el modelado de nichos desde una perspectiva que ha sido llamado Grinnelliana, la cual enfatiza los requisitos de una especie (y no sus efectos sobre otras especies, ni su rol ecológico en una comunidad; Soberón 2007, 2010, Soberón & Nakamura 2009, Peterson et al. 2011, Anderson 2013). Este ensayo nace del "I Simposio de Biogeografía: Actualidad y Retos" que formó parte del XII Congreso Nacional Mastozoología mexicano, realizado en Puebla en octubre del 2014. Lo redacto en castellano como homenaje a tal congreso, con el especial deseo de llegar con mayor facilidad a estudiantes universitarios y otros investigadores que están iniciándose en este campo de estudio, tanto en México como en otros países hispanohablantes. Traducciones al inglés y francés aparecerán al final, con el objetivo de conseguir la mayor difusión posible.

Durante los últimos 15 años el modelado de nicho ecológico (frecuentemente también llamado modelado de distribución de especies) se ha establecido como una de las herramientas más importantes en biogeografía y campos relacionados (Guisan & Thuiller 2005, Wiens & Graham 2005, Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011) pero opino que no está claro para muchos investigadores lo estrechamente enlazados que son el modelado de nicho y la ecología clásica, ni que los métodos requieren que el usuario tome decisiones biológicas y estadísticas importantes (Guisan & Zimmermann 2000, Anderson 2013). En parte, la situación se debe al hecho de que el campo es relativamente nuevo y sigue desarrollándose con alta rapidez. Pero además percibo que proviene del hecho de que la tecnología necesaria para realizar los análisis es relativamente fácil de usar y a la vez económicamente accesible. Por ejemplo, en contraste con lo complicado (en términos mecánicos) que es secuenciar el ADN, el modelado de nicho se puede hacer con un simple ordenador básico en cualquier despacho, sala o cibercafé. Esta facilidad del uso, puede promover el mal uso (Anderson 2012). Por lo tanto, con el deseo de guiar a personas que están iniciándose en el área, aquí reseño algunos de los puntos más importantes—que

según he visto son los más frecuentemente malentendidos. No intento que este corto ensayo sea exhaustivo ni trato de citar con igualdad entre las miles de contribuciones científicas relevantes publicadas en la última década. Sin embargo, para orientar al lector, hago referencia a varias publicaciones que han salido de mi laboratorio y a otras que son las que más me han aclarado asuntos claves.

EL PRIMER PASO: ¿QUÉ ES LO QUE UNO QUIERE MODELAR?

Aunque muchos autores ya lo han dicho (Peterson et al. 2011, Anderson 2012), reitero que el primer paso es decidir si el problema biológico requiere un modelo de: (1) los requisitos ambientales de la especie (que también indica las áreas geográficas adecuadas para ella, es decir su distribución potencial) o (2) las áreas realmente ocupadas por la especie distribución geográfica o rango; Fig. 1a). Cabe destacar que el resultado del primer tipo de modelo puede ser procesado para obtener el segundo pero viceversa. Tal tipo. no procedimiento incluye la consideración de barreras geográficas, diferencias en el contexto biótico (la presencia y abundancia de otras especies que interactúan de manera clave con la especie de interés) v cambios ambientales locales producidos por actividades del ser humano (Anderson & Martínez-Meyer 2004, Urban et al. 2013). Bajo determinados supuestos,

un modelo de los requisitos ambientales de la especie puede ser aplicado a otras regiones u otros tiempos (pasado o futuro) para indicar las regiones adecuadas para dicha especie (Nogués-Bravo 2009, Varela et al. 2009, Elith et al. 2010, Broennimann et al. 2012, Veloz et al. 2012, Guisan et al. 2014, Strubbe et al. 2015). Al contrario, un modelo que indica las áreas ocupadas por una especie no puede ser aplicado a otras regiones u otros tiempos para predecir con éxito su distribución (pero véase Pagel & Schurr 2012, Schurr et al. 2012). Además, modelar o los requisitos ambientales de la especie o su área de distribución tendrá consecuencias relacionadas con otras decisiones importantes después durante el proceso de modelamiento. A continuación menciono varias de estas decisiones, asumiendo que el objetivo es modelar los requisitos ambientales de la especie (y así predecir las áreas adecuadas para ella). Enfatizo aspectos relacionados con los datos y la generación del modelo, pero invito al lector a consultar la literatura (p. ej., Veloz 2009, Franklin 2010. Peterson et al. 2011. Radosavljevic & Anderson 2014) para ponderar los principios claves sobre la evaluación del modelo.

CUATRO ASUNTOS CRÍTICOS

1. Variables ambientales

Para modelar los requisitos ambientales de una especie (su nicho Grinnelliano), existen tres

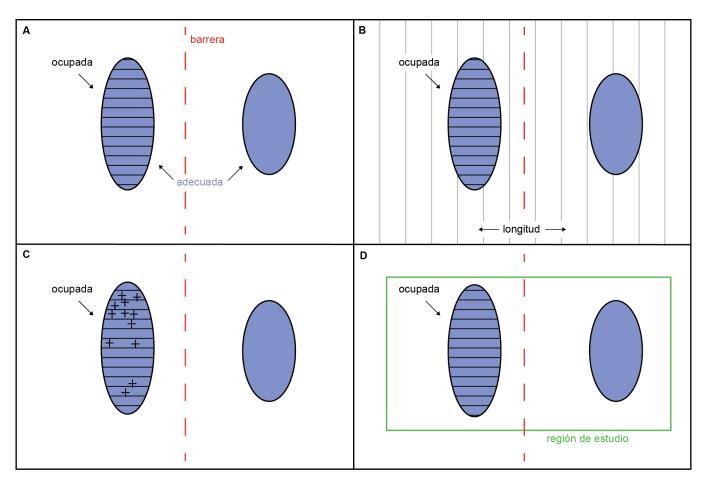


Fig 1. Ilustración de algunos asuntos claves del modelado de nicho ecológico. A) El área ocupada por una especie puede ser menor que el área ambientalmente adecuada para ella, por ejemplo debido a una barrera de dispersión. B) Algunas variables, especialmente las que son puramente espaciales (como latitud, o aquí longitud) pueden ser correlacionadas no con el área adecuada para la especie sino con el área que ocupa (su distribución). C) Los sesgos de muestreo pueden resultar en muestras no representativas de las áreas adecuadas para la especie (y potencialmente también de las condiciones ambientales requeridas por ella). D) Si la región de estudio incluye áreas adecuadas para la especie pero no ocupadas por ella (por ejemplo debido a una barrera de dispersión), el resultado del modelo puede representar un subconjunto ambientalmente sesgado de las condiciones adecuadas para la especie.

principios importantes con respecto escoger las variables predictivas (Anderson 2013). La primera es usar solamente variables cuyos valores no son afectados por la especie ejemplo. que está modelando (por se precipitación). Dichas variables se califican como "escenopoyéticas" (las que "preparan/construyen el escenario"; Hutchinson 1978: 172, Soberón 2007, 2010, Soberón & Nakamura 2009),

aparentemente procedente de la palabra griega skene (raíz de la latina scena) y la griega poieo o poïesis. Al contrario, el uso de una variable no escenopoyética (aquella cuyos valores sí son afectados por la especie que está siendo modelada) requiere tener en cuenta la retroalimentación entre la variable predictora y la abundancia de la especie siendo modelada a través de relaciones matemáticas más complejas

que las de los algoritmos del modelado de nichos Grinnellianos (Chase & Leibold 2003). Por ejemplo, si una especie de roedor come una especie de gramínea (y así reduce de manera importante la abundancia de ella), la gramínea no es escenopoyética y por lo tanto no es adecuada como una variable predictiva en un modelo Grinnelliano del roedor. Segundo, es necesario evitar el uso de variables puramente espaciales (como latitud y longitud), que se relacionan con la distribución de la especie pero no con sus requerimientos ambientales (Fig. 1b; Austin 2002, Anderson 2013). Tercero, siempre que sea posible, es preferible usar variables a las cuales la especie responde, es decir variables causales (Austin 2002). En las comunes ocasiones en las cuales las variables causales no están disponibles (o ni siguiera son conocidas), el investigador puede utilizar otras variables proxy que son de utilidad porque están correlacionadas con las variables causales. Sin embargo, hay que tener en cuenta que el modelo, al ser proyectado a otras regiones o períodos, será solamente de utilidad si las correlaciones entre las variables proxy y las causales (que no fueron utilizadas) son las mismas que en la región y período de tiempo en los cuales se generó el modelo. [Vale la pena mencionar que para modelar la distribución de una especie (el área que realmente ocupa) también es necesario usar variables escenopoyéticas, pero en este caso sí es útil (y probablemente necesario) incorporar también variables espaciales. Para estos modelos es posible además usar variables proxy;

sin embargo, ya que modelos de este tipo no mantienen la misma utilidad proyectados en el espacio o el tiempo, no es aplicable la aclaración sobre las correlaciones entre las variables *proxy* y causales.]

2. Registros de la especie

Además del requisito de que los registros pertenezcan a la especie de interés y que sus coordenadas no posean demasiado error en relación a la pregunta biológica, es necesario que los registros sirvan como una muestra representativa de las condiciones ambientales requeridas por la especie (Peterson et al. 2011). Sin embargo, para la gran mayoría de taxones en casi todas las regiones del mundo, los de registros conocidos cualquier especie representan el resultado de esfuerzos de muestreo no homogéneos en el espacio los geográfico, cuales pueden además representar un sesgo en el espacio ambiental (Fig. 1c; Reddy & Dávalos 2003, Kadmon et al. 2004, Soberón & Peterson 2004). Por lo tanto, casi siempre es necesario realizar algún análisis metodológico para corregir o disminuir el efecto de dichos sesgos. Tal vez la mejor manera de tomar en cuenta cualquier sesgo es por medio de índices que cuantifiquen el esfuerzo de muestreo para todas las especies que se colectan u observan con las mismas técnicas usadas para detectar la especie de interés (target group en inglés; Anderson 2003, Phillips et al. 2009). En ausencia de esta información, otra técnica propuesta en la literatura es el filtrado de

localidades únicas de presencia de la especie de interés, usualmente en el espacio geográfico (pero véase Varela et al. 2014 para filtrado en el espacio ambiental) con el fin de disminuir la repetitividad de localidades sobre-representadas en el conjunto de datos debido a sesgos de muestreo (Aiello-Lammens et al. 2015). Sin embargo, ya que el objetivo del filtrado es remover el efecto del sesgo sin disminuir la verdadera señal del nicho, es necesario determinar empíricamente la distancia óptima para dicho filtrado (Anderson 2012: 74, Boria et al. 2014). [Notablemente, lo mismo es cierto para análisis que pretenden modelar la distribución de una especie.]

3. Región de estudio

En el caso de técnicas que generan un modelo utilizando un conjunto de localidades contra las cuales se comparan los registros de la especie (la gran mayoría de los algoritmos; Elith & Graham 2009, Peterson et al. 2011) el resultado del modelo puede ser sesgado si el conjunto de datos de comparación incluye regiones adecuadas para la especie, que no son habitadas por ella (Anderson & Raza 2010, Barve et al. 2011). Estos conjuntos de datos de comparación incluyen ausencias, seudoausencias o muestras (usualmente aleatorias) del entorno ambiental disponible (background en inglés; Anderson 2013: 14) y pueden incluir regiones adecuadas pero no habitadas debido a: (1) la falta de dispersión (o la falta de permanencia de poblaciones); (2) interacciones bióticas y/o (3) cambios ambientales causados por el ser humano (Fig. 1d). Por tales razones, tomando en cuenta la información disponible, el investigador debe escoger una región de estudio (entorno geográfico), que concuerde lo más posible con el supuesto de que ninguno de estos tres factores sea responsable de que la especie habite un subconjunto ambientalmente no aleatorio de lo adecuado para ella (Anderson 2013). [Tal supuesto no es relevante para modelos cuyo propósito es caracterizar la distribución de la especie.]

4. Complejidad del modelo

Pasando ya a la generación del modelo, el investigador debe tomar en cuenta el nivel de complejidad del modelo. ajustando configuración del algoritmo para conseguir un modelo que no sea ni demasiado sencillo ni demasiado complejo (Warren & Seifert 2011). Debido а las características de muchos algoritmos existentes en la actualidad y el alto número de variables disponibles, existe al parecer una tendencia hacia excesiva complejidad, que da como consecuencia un modelo llamado "sobreajustado" (Anderson & Gonzalez 2011, Radosavljevic & Anderson 2014). Tales modelos pueden predecir con mayor éxito las localidades conocidas de una especie (y probablemente a la vez su distribución), pero tienden a subestimar sus requisitos ambientales (es decir, su nicho ecológico) y así también las áreas adecuadas para la especie. Doy un par de ejemplos, con los algoritmos BIOCLIM y Maxent.

El algoritmo de BIOCLIM forma un modelo que incluye el rango de valores observado para la especie para cada una de las variables utilizadas de manera independiente. De tal forma se define un subconjunto del espacio multidimensional (el hipervolúmen *n*-dimensional de Hutchinson) dentro del cual las condiciones son adecuadas para la especie (Hutchinson 1957, Booth et al. 2014). Entre más variables (= dimensiones) sean consideras, el modelo resultante será por definición más complejo y así las áreas geográficas predichas como adecuadas para la especie tendrán a ser cada vez más restringidas. Cuando la dimensionalidad sea exagerada, el modelo estará sobreajustado. Por su parte Maxent se basa en un algoritmo matemáticamente muy diferente para generar un modelo de los requisitos de la especie, pero también está sujeto a la posibilidad de un "sobreajuste" (Phillips et al. 2006, Warren & Seifert 2011, Radosavljevic & Anderson 2014). En Maxent la complejidad del modelo se limita por medio de una penalización que en inglés se llama regularization, que se aplica a cada variable incluida en el modelo (Phillips & Dudík 2008, Merow et al. 2013). Entre más alto sea el valor de penalización, más sencillo tendrá a ser el modelo (por ejemplo, tendrá a ser menor el número de variables incluidas en el modelo). Varios estudios recientes han mostrado que modelos con diferentes niveles de complejidad pueden variar tremendamente en sus predicciones de las áreas adecuadas para la especie (Elith et al. 2010, Anderson & Gonzalez 2011, Shcheglovitova & Anderson 2013), resaltando la importancia de tomar en cuenta la complejidad del modelo. [Las mismas normas son relevantes también para los modelos que tienen como objetivo predecir la distribución de la especie.]

CONCLUSIONES

Reitero que los detalles metodológicos van cambiando rápidamente en el campo del modelado de nicho ecológico, pero percibo que la línea de investigación está llegando a cierta madurez en cuanto a principios fundamentales (Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011). Por eso predigo que los cuatro asuntos resumidos aquí seguirán vigentes. Por lo tanto espero que este ensayo inspirado por el simposio realizado en Puebla sirva para ubicar y ayudar a estudiantes y otros investigadores en el 2015 y más allá.

AGRADECIMIENTOS

Agradezco sobre todo a Tania Escalante, quien con gran aporte de los co-organizadores Elkin Noguera-Urbano y César Talonia lideró el "I Simposio de Biogeografía: Actualidad y Retos" dentro del XII Congreso Nacional de Mastozoología mexicano, llevado a cabo en la ciudad de Puebla de Zaragoza en octubre del 2014. Enrique Martínez-Meyer, Claudia Moreno-Arzate, A. Monserrat Cruz-Lira, Ángela Cuervo e

Ignacio Ferro aportaron ayuda logística fundamental. Diego Alvarado-Serrano ofreció sugerencias sobre el contenido del ensayo y corrigió mis numerosos y creativos errores de gramática en castellano en la versión original. Olivier Broennimann y Corentin Bohl hábilmente tradujeron el ensayo al francés. Agradezco también el aporte financiero de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad para asistir al congreso (CONABIO MU-011). La investigación que condujo a las ideas aquí presentadas fue financiada por la National Science Foundation de EE.UU. (NSF DEB-0717357 y DEB-1119915) y por la Blavatnik Family Foundation. Dedico este ensayo a todas las personas interesadas en nichos distribuciones quienes conversaron conmigo sobre estos temas durante el congreso.

LEAD ARTICLE

MODELING NICHES AND DISTRIBUTIONS: IT'S NOT JUST "CLICK, CLICK, CLICK"

Robert P. Anderson

INTRODUCTION

I offer this short essay with the idea of emphasizing that, despite being a relatively new field still under development, ecological niche modeling (and its applications for studies of species geographic distributions) involves both very important theoretical aspects and already fairly sophisticated methodologies. Specifically, the essay concerns the modeling of niches from a perspective that has been called Grinnellian. emphasizing a species' requirements (not its effects on other species, or its ecological role in a community; Soberón 2007, 2010, Soberón & Nakamura 2009, Peterson et al. 2011, Anderson 2013). It derives from the "I Biogeography Symposium: State of the Field and Challenges," which was part of the Mexican XII Congress of Mammalogy held in Puebla in October 2014. In homage to that scientific meeting, I wrote it originally in Spanish, with the particular hope of more easily reaching university students and other researchers entering this field in Mexico and the rest of the Spanish-speaking world. This English translation and the accompanying French one follow the original, with the aim of reaching the largest possible audience.

During the last 15 years, ecological niche modeling (also commonly called species distribution modeling) has become one of the

most important tools in biogeography and related fields (Guisan & Thuiller 2005, Wiens & Graham 2005, Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011), but I feel that it is not clear to many researchers just how tight the theoretical links are between niche modeling and classical ecology, or that the methods require the user to make very important biological and statistical decisions (Guisan & Zimmermann 2000, Anderson 2013). In part, this is because the field is fairly new and continues to be under rapid development. However, I sense that it also has to do with the fact that the technology needed to conduct the analyses is fairly easy to use and at the same time relatively inexpensive. For example, in contrast to how complicated (in physical terms) it is to sequence DNA, niche modeling can be done on a simple standard computer in any office, living room, or cybercafe. This ease of use can promote misuse (Anderson 2012). Therefore, hoping to orient those just starting out in the field, here I outline several of the most important issues—which I have seen to be misunderstood most frequently. The aim of this short essay is not to be exhaustive or try to the thousands of relevant scientific contributions published in the last decade equally. Nevertheless, I do refer the reader to several publications that have come out of my laboratory and to others that have most clarified

key issues for me.

THE FIRST STEP: WHAT IS IT THAT

I WANT TO MODEL?

Although many authors already have stated it (Peterson et al. 2011, Anderson 2012), I reiterate that the first step is to decide if the biological problem calls for a model of: (1) the species' environmental requirements (which also indicates the geographic areas suitable for it, that is, its potential distribution) or (2) the areas it actually occupies (its geographic distribution or range; Fig. 1a). It is important to emphasize that the output of the former type of model can be processed to obtain the latter type, but not vice versa. In the first case, processing includes the consideration of geographic barriers, differences in biotic context (the presence and abundance of other species that interact with the focal species in a key way), and local anthropogenic environmental changes (Anderson & Martínez-Meyer 2004, Urban et al. 2013). Under clear assumptions, a model of a species' environmental requirements can be applied to other regions or time periods (past or future) to indicate the areas suitable for it (Nogués-Bravo 2009, Varela et al. 2009, Elith et al. 2010, Broennimann et al. 2012, Veloz et al. 2012, Guisan et al. 2014, Strubbe et al. 2015.) In contrast, a model that identifies the areas that a species occupies cannot be applied to other regions or time periods to predict the species' distribution accurately (but see Pagel & Schurr

2012, Schurr et al. 2012). Furthermore, modeling either the species' environmental requirements, or its distribution, will have consequences associated with important later decisions in the process of modeling. Below, I mention a few of these decisions, assuming that the goal is to model the species' environmental requirements (and hence predict the areas suitable for it). I emphasize issues related to input data and model building, but refer the reader to the literature to consider key principles of model evaluation (Veloz 2009, Franklin 2010, Peterson et al. 2011, Radosavljevic & Anderson 2014).

FOUR KEY ISSUES

1. Environmental variables

To model a species' environmental requirements (its Grinnellian niche), three important principles exist related to selecting predictor variables (Anderson 2013). The first is to use only variables whose values are not affected by the species being modeled (for example, precipitation). Such variables termed are "scenopoetic" (those that "set the stage": Hutchinson 1978: 172, Soberón 2007, 2010, Soberón & Nakamura 2009), apparently from the Greek word skene (root of the Latin scena) and the Greek poieo or poïesis. In contrast, using a non-scenopoetic variable (whose values are affected by the species being modeled) requires taking into account feedback between the predictor variable and the abundance of the

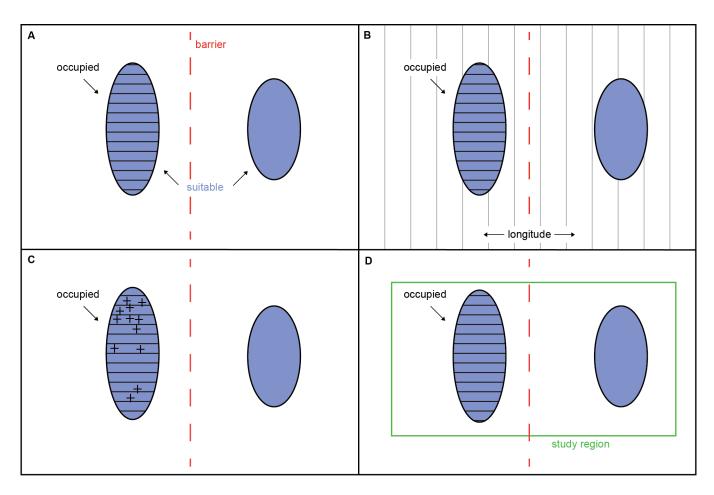


Fig 1. Illustration of a few key issues in ecological niche modeling. A) The area occupied by a species can be smaller than that environmentally suitable for it, for example due to a dispersal barrier. B) Some variables, especially those that are purely spatial (such as latitude, or here longitude) may correlate not with the area suitable for the species but rather with the area it occupies (its distribution). C) Sampling biases can result in samples that are not representative of the areas suitable for the species (and potentially also not of the environmental conditions required by it). D) If the study region includes areas that are suitable for the species but that it does not occupy (for example, due to a dispersal barrier), the model output can represent an environmentally biased subset of the conditions suitable for the species.

species being modeled, via more complicated mathematical formations than those found in algorithms used for modeling Grinnellian niches (Chase & Leibold 2003). For example, if a species of rodent eats a species of grass (and reduces its abundance in an important way), the grass is not scenopoetic and therefore is not valid as a predictor variable in a Grinnellian niche model of the rodent. Second, it is necessary to avoid the use of purely spatial variables (such as

latitude and longitude), which are associated with the species' distribution but not its requirements (Fig. 1b; Austin 2002, Anderson 2013). Third, whenever possible, it is advantageous to use variables that the species responds to—that is, the causal ones (Austin 2002). In the many cases when the causal variables are not available (or even known), the researcher can employ other, proxy variables, which are useful because they are correlated with the causal ones. However, it is

important to note that the model only is valid in other regions or time periods if the correlations between the proxy and (un-used) causal variables remain the same as in the region and time period in which the model was made. [It is worth mentioning that to model a species' distribution (the area it truly occupies) it is also necessary to use scenopoetic variables, but in this case it is useful (and probably necessary) to employ spatial variables as well. For such models, it is also possible to use proxy variables; nevertheless, since such models do not maintain their same function when projected across space or time, the clarification about correlations among proxy and causal variables is irrelevant.]

2. Occurrence records of the species

In addition to the requirement that the occurrence records belong to the species being modeled and that their coordinates do not suffer from excessive error for the biological problem at hand, the records must constitute a representative sample of the environmental conditions required by the species (Peterson et al. 2011). Nevertheless, for the overwhelming majority of taxonomic groups in almost all regions of Earth, the known records of any species represent the result of sampling effort that is heterogeneous across geographic space and which also may include biases in environmental space (Fig. 1c; Reddy & Dávalos 2003, Kadmon et al. 2004, Peterson & Soberón 2004). Therefore, some methodological analysis is almost always required either to correct for or to reduce the effect of such

biases. Perhaps the best way of correcting for a bias is via indices that quantify sampling effort for the overall target group (all the species that are collected or observed with the same techniques through which the focal species is detected; Anderson 2003, Phillips et al. 2009). When such information is unavailable, another technique proposed in the literature is to filter unique occurrence localities for the focal species, usually in geographic space (but see Varela et al. 2014 for filtering in environmental space), with the goal of reducing the repetitiveness of localities that are overrepresented in the dataset due to sampling bias (Aiello-Lammens et al. 2015). Nevertheless, since the objective of filtering is to remove the effect of the bias without diluting the true niche signal, the optimal filtering distance must be determined empirically (Anderson 2012: 74, Boria et al. 2014). [Notably, the same is true for analyses that aim to model the species' distribution.]

3. Study region

For the techniques that make a model using a set of localities that are compared with records of the focal species (the vast majority of algorithms; Elith & Graham 2009, Peterson et al. 2011), the model's output can be biased if the comparison dataset includes regions that are suitable for the species but which it does not inhabit (Anderson & Raza 2010, Barve et al. 2011). Such comparison datasets include absences, pseudoabsences, or (usually random) samples of the environmental background

available (Anderson 2013: 14); they can include suitable yet unoccupied regions due to: (1) dispersal limitations (or local extinctions); (2) biotic interactions; and/or (3) environmental alterations caused by humans (Fig. 1d). For these reasons, the researcher must take into account available information to choose a study region that approximates as much as possible the assumption that these three factors do not cause the species to inhabit an environmentally biased subset of the conditions suitable for it (Anderson 2013). [This assumption is irrelevant for models whose goal is to characterize the species' distribution.]

4. Model complexity

Turning now to model building, the researcher must assess the level of model complexity, tuning the algorithm's settings to achieve a model that is neither too simple nor too complex (Warren & Seifert 2011). Due to the characteristics of many of the currently existing algorithms and the high number of variables available, there seems to be a tendency toward excessive complexity, which results in a model termed "overfit" (Anderson & Gonzalez 2011, Radosavljevic & Anderson 2014). Such models can better predict the known records of the species (and probably its distribution), but they tend to underestimate its environmental requirements (that is, its ecological niche) and therefore also the geographic areas suitable for it. I provide two examples, with the algorithms BIOCLIM and Maxent. BIOCLIM's algorithm forms a model that includes the range of values observed for the species for each of the input variables independently. Hence it defines a subset of multidimensional space (the ndimensional hypervolume of Hutchinson) within which the conditions are suitable for the species (Hutchinson 1957, Booth et al. 2014). The more variables (= dimensions) used, the resulting model will by definition be more complex, and therefore the geographic areas predicted as suitable for the species will tend to be restricted. When the progressively more dimensionality is excessive, the model is overfit. Now considering Maxent, its algorithm models the species' requirements in a mathematically very different way, but it is also subject to the possibility of overfitting (Phillips et al. 2006, Warren & Seifert 2011, Rasosavljevic & Anderson 2014). In Maxent, model complexity is limited by a penalty called regularization, which is applied to each variable included in the model (Phillips & Dudík 2008, Merow et al. 2013). As the strength of the penalty increases, the final model produced will tend to be simpler (for example, tending to include fewer variables). Several recent studies have shown that models with differing levels of complexity can vary tremendously in their geographic predictions of the areas suitable for the species (Elith et al. 2010, Anderson & Gonzalez 2011, Shcheglovitova & Anderson 2013), emphasizing the importance of considering model complexity. [The same issues also apply to models that have the goal of predicting the species' distribution.]

CONCLUSIONS

I reiterate that methodological details continue to change rapidly in the field of ecological niche modeling, but I sense that this line of research is reaching maturity regarding fundamental principles (Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011). Because of this, I predict that the four issues outlined here will continue to hold true. Hence, I hope that this essay inspired by the symposium held in Puebla will serve to orient and help students and other researchers in 2015 and beyond.

Conocimiento y Uso de la Biodiversidad for financial support to attend the meeting (CONABIO MU-011). The research that led to the ideas presented here was financed by the U.S. National Science Foundation (NSF DEB-0717357 and DEB-1119915) and by the Blavatnik Family Foundation. I dedicate this essay to everyone interested in niches and distributions who discussed these topics with me during the meeting.

ACKNOWLEDGMENTS

First and foremost, I thank Tania Escalante. With great support from co-organizers Elkin Noguera-Urbano and César Talonia, she orchestrated the "I Biogeography Symposium: State of the Field and Challenges" at the Mexican XII National Congress of Mammalogy held in the city of Puebla de Zaragoza in October 2014. Enrique Martínez-Meyer, Claudia Moreno-Arzate, A. Monserrat Cruz-Lira, Ángela Cuervo, and Ignacio Ferro provided key logistic assistance. Diego Alvarado-Serrano offered suggestions regarding the content of the essay and corrected my many and creative grammatical errors in the Spanish original. Olivier Broennimann and Corentin Bohl skillfully translated the essay into French. I also thank the Mexican Comisión Nacional para el

LA MODÉLISATION DE NICHE ET DE DISTRIBUTIONS: CE N'EST PAS JUSTE "CLIC, CLIC, CLIC"

LEAD ARTICLE

Robert P. Anderson

INTRODUCTION

Je mets à disposition ce court essai avec l'idée de mettre l'accent sur le fait que, malgré que ce soit un champ de recherche nouveau et encore en développement, la modélisation écologique de la niche (et son application à l'étude de distribution géographique d'espèces) implique à la fois des aspects théoriques importants et des méthodes relativement sophistiquées. En particulier, cet essai traite de la modélisation de niche selon une perspective « Grinnelliènne », qui met l'accent sur les besoins d'une espèce (et non sur son effet sur d'autres espèces, ou son écologique dans une communauté; rôle Soberón 2007, 2010, Soberón & Nakamura 2009, Peterson et al. 2011, Anderson 2013). Il émane du « 1er Symposium de Biogéographie : état des lieux et enjeux » qui fut part du 12ème Congrès Mexicain de Mammologie organisé à Puebla en octobre 2014. Comme hommage à cette réunion scientifique, je l'ai à l'origine écrit en espagnol, avec l'espoir d'atteindre les étudiants et chercheurs commençant dans ce domaine au Mexique et plus largement le monde hispanophone. Cette traduction française et celle en anglais sont fidèles à l'originale, avec comme but d'atteindre une audience aussi large que possible.

Au cours des 15 dernières années, la modélisation écologique de la niche (aussi communément appelée la modélisation de distribution des espèces) est devenu l'un des outils les plus importants en biogéographie et disciplines associées (Guisan & Thuiller 2005, Wiens & Graham 2005, Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011), mais je pense qu'il n'est pas clair pour un grand nombre de chercheurs à quel point les liens théoriques sont étroits entre la modélisation de niche et l'écologie classique, ou que les méthodes obligent l'utilisateur à prendre des décisions biologiques et statistiques très importantes (Guisan & Zimmermann 2000, Anderson 2013). Cela s'explique en partie par le fait que cette discipline est relativement nouvelle et qu'elle continue d'être en cours de développement rapide. Cependant, je crois que cela est aussi lié au fait que la technologie nécessaire pour effectuer les analyses est assez facile à utiliser et relativement peu coûteuse. Par exemple, contrairement à combien il est compliqué (en termes physiques) de séquencer l'ADN, la modélisation de niche peut être fait sur un simple ordinateur standard dans n'importe quel bureau, salon, ou cybercafé. Cette facilité peut promouvoir une utilisation abusive (Anderson 2012). Par conséquent, en espérant orienter ceux qui débutent dans le domaine, j'expose ici

quelques-unes des plus importantes questions—celles que j'ai vues le plus souvent être mal comprises. Le but de ce court essai n'est pas d'être exhaustif ou d'essayer de citer également les milliers de contributions scientifiques pertinentes publiées durant la dernière décennie. Néanmoins, je renvoie le lecteur à quelques publications qui sont sorties de mon laboratoire et à d'autres qui m'ont le plus aidé à clarifier les questions clés.

LE PREMIER PAS: QU'EST-CE QU'ON VEUT MODÉLISER?

Bien que plusieurs auteurs l'aient déjà exprimé (Peterson et al. 2011, Anderson 2012), je répète que le premier pas est de décider si le problème biologique requiert un modèle des : (1) besoins environnementaux de l'espèce (ce qui indique aussi l'aire géographique favorable, c.-à-d. sa distribution géographique potentielle) ou (2) de l'aire qu'elle occupe effectivement (sa distribution géographique observée ; Fig. 1a). Il est important de noter que les prédictions du premier type de modèles peuvent être traitées pour obtenir celles du second type, mais pas le contraire. Dans le premier cas, le traitement inclut la considération de barrières géographiques, de différences de contexte biotique (la présence et abondance d'une autre espèce interagissant de manière avec l'espèce importante focale) les changements anthropogéniques locaux de l'environnement (Anderson & Martínez-Meyer 2004, Urban et al. 2013). Avec des hypothèses claires, un modèle de besoin environnementaux d'une espèce peut être appliqué à une autre région ou à d'autres périodes temporelles (passées ou futures) afin d'indiquer les aires favorables (Nogués-Bravo 2009, Varela et al. 2009, Elith et al. 2010, Broennimann et al. 2012, Veloz et al. 2012, Guisan et al. 2014, Strubbe et al. 2015). Au contraire, un modèle identifiant les aires qu'une espèce occupe ne peut être appliqué à une autre région ou à une autre période temporelle pour prédire la distribution de l'espèce correctement (mais voir Pagel & Schurr 2012, Schurr et al. 2012). De plus, modéliser soit les besoins environnementaux de l'espèce, soit sa distribution, impliquera d'importantes décisions lors du processus de modélisation. Ci-dessous, j'en mentionne qu'elles-unes, en assumant que le but est modéliser de les besoins environnementaux de l'espèce (et donc de prédire les aires favorables). Je souligne les problèmes liés aux données initiales et à la calibration du modèle, mais réfère le lecteur à la littérature quant à la considération des principes clés de l'évaluation des modèles (Veloz 2009, Franklin 2010. Peterson et al. 2011, Radosavljevic & Anderson 2014).

QUATRE PROBLÈMES CLÉS

1. Variables environnementales

Pour modéliser les besoins environnementaux d'une espèce (sa niche « Grinnelliène »), trois importants principes liés à la sélection des variables prédictives existent (Anderson

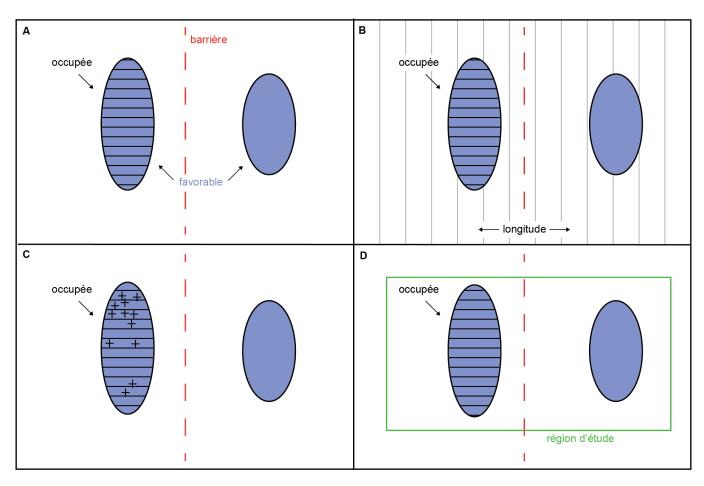


Fig 1. Illustration de quelques problèmes clés lors de modélisation de la niche écologique. A) L'aire occupée par l'espèce peut être plus petite que celle avec un environnement favorable, par exemple dû à une barrière à la dispersion. B) Certaines variables, en particulier celles qui sont purement spatiales (comme la latitude ou ici la longitude), peuvent être corrélées non pas avec l'aire favorable pour l'espèce mais avec l'aire qu'elle occupe (sa distribution). C) Des biais d'échantillonnage peuvent exister dans les échantillons qui ne sont pas représentatifs de l'aire favorable pour l'espèce (et potentiellement aussi des conditions environnementales requises). D) Si la région d'étude inclut des aires favorable pour l'espèce mais qu'elle n'occupe pas (par exemple dû à une barrière à la dispersion), le résultat du modèle peut représenter un sous-ensemble biaisé des conditions environnementales favorables pour l'espèce.

2013). La première est d'utiliser seulement les variables dont les valeurs ne sont pas affectées par l'espèce modélisée (comme les précipitations). De telles variables sont appelées « scénopoïétiques » (celles qui « créent le décor »; Hutchinson 1978 : 172, Soberón 2007, 2010, Soberón & Nakamura 2009), apparemment du Grec skene (racine du mot Latin scena) et du Grec poieo ou poïesis. En revanche, utiliser des

variables non-scénopoïétiques (dont les valeurs sont affectée par l'espèce) requiert de prendre en rétroactions compte les entre variables prédictives et abondance de l'espèce modélisée via des formulations mathématiques compliquées que celles trouvées dans les algorithmes utilisés pour la modélisation de niches Grinnelliènes (Chase & Leibold 2003). Par exemple, si une espèce de rongeur mange une

espèce de graminée (et réduit son abondance de manière importante), la graminée n'est pas scénopoïétique, et de ce fait, non valide comme variable prédictive dans un modèle de niche Grinnelliène. Deuxièmement, il est nécessaire d'éviter l'utilisation de variables purement spatiales (comme la latitude et la longitude), qui sont associées à la distribution, mais pas aux besoins environnementaux de l'espèce (Fig. 1b; Austin 2002, Anderson 2013). Troisièmement, si possible, il est avantageux d'utiliser les variables auxquelles l'espèce répond – c.-à-d. variables causales (Austin 2002). Dans de nombreux cas, quand les variables causales ne sont pas disponibles (ou inconnues), le chercheur peut employer d'autres variables dites « proxy», qui sont utiles car corrélées avec les variables causales. Cependant, il est important de noter que les modèles ne sont valides dans d'autres régions ou périodes de temps que si les corrélations entre variables proxy et variables causales (non-utilisées) restent les mêmes dans la région ou la période de temps à laquelle le modèle a été fait. [Il convient de mentionner que pour modéliser la distribution de l'espèce (l'aire qu'elle occupe effectivement), il est aussi nécessaire d'utiliser variables des scénopoïétiques, mais dans ce cas, il est utile (et probablement nécessaire) d'employer également des variables spatiales. Pour de tels modèles, il est également possible d'utiliser des variables proxy; néanmoins, étant donné que ces modèles ne gardent pas leur fonction quand ils sont projetés dans le temps et l'espace, la clarification

à propos des corrélations entre variables proxy et causales n'est pas pertinente.]

2. Données d'occurrence de l'espèce

En plus du besoin que les données d'occurrences appartiennent à l'espèce modélisée et que leurs coordonnées ne souffrent d'erreurs excessives problème biologique abordé, pour le doivent constituer occurrences un échantillonnage représentatif des conditions environnementales requises par l'espèce (Peterson et al. 2011). Néanmoins, pour la grande majorité des groupes taxonomiques, dans la plupart des régions du monde, les données à disposition le résultat d'un effort sont d'échantillonnage hétérogène à travers l'espace géographique, et elles peuvent aussi être biaisées dans l'espace environnemental (Fig. 1c; Reddy & Dávalos 2003, Kadmon et al. 2004, Peterson & Soberón 2004). De ce fait, des analyses méthodologiques sont presque toujours requises pour corriger ou pour réduire l'effet de ces biais. Le meilleur moyen de corriger ce biais est d'utiliser des indices qui quantifient l'effort d'échantillonnage pour l'ensemble des groupes espèces collectées cibles (toutes les observées avec la même technique que l'espèce focale; Anderson 2003, Phillips et al. 2009). Lorsque cette information n'est pas disponible, une autre technique proposée dans la littérature est de filtrer les localités uniques pour l'espèce focale, normalement dans l'espace géographique (mais voir Varela et al. 2014 pour filtrer dans l'espace environnemental), avec comme but de

réduire la redondance de localités qui sont surreprésentées dans les données dû à l'effort d'échantillonnage (Aiello-Lammens et al. 2015). Néanmoins, et comme l'objectif du filtre est d'enlever l'effet du biais sans diluer le signal réel de la niche, la distance optimale pour filtrer doit être déterminée de manière empirique (Anderson 2012: 74, Boria et al. 2014). [Il en va de même pour les analyses ayant pour but de modéliser l'aire de l'espèce.]

3. Région d'étude

Pour les techniques qui construisent un modèle en utilisant un ensemble de localités qui sont comparées aux occurrences de l'espèce focale (la vaste majorité des algorithmes; Elith & Graham 2009, Peterson et al. 2011), les prédictions du modèle peuvent être biaisés si les données de comparaison incluent des régions qui sont favorable à l'espèce mais non occupées (Anderson & Raza 2010, Barve et al. 2011). De telles données de comparaison incluent des absences, des pseudoabsences, ou des échantillons (généralement aléatoires) de l'environnement disponible (« background dataset» en anglais ; Anderson 2013: 14) ; Elles peuvent inclure des régions favorables mais inoccupées à cause: (1) des limites de dispersion (ou des extinctions locales); (2) des interactions biotiques; et/ou (3) des changements de l'environnement dus à l'homme (Fig. 1d). Pour ces raisons, le chercheur doit tenir en compte toute l'information disponible afin de choisir une région d'étude qui respecte au mieux l'assomption selon laquelle ces trois facteurs ne conduisent pas l'espèce à occuper un sousensemble environnemental biaisé des conditions qui lui sont favorables (Anderson 2013). [Cette assomption n'est pas pertinente pour les modèles ayant pour but de caractériser l'aire de l'espèce.]

4. Complexité du modèle

Passant maintenant à la construction du modèle, le chercheur doit estimer le niveau du modèle. complexité adaptant paramètres de l'algorithme pour obtenir un modèle qui ne soit ni trop simple, ni trop complexe (Warren & Seifert 2011). Dû aux caractéristiques de beaucoup d'algorithmes existants et au grand nombre de variables disponibles, une complexité trop grande est plus probable. On parle dans ce cas de modèle « surajusté » (Anderson & Gonzalez 2011, Radosavljevic & Anderson 2014). Un tel modèle peut mieux prédire les occurrences connues de l'espèce (et probablement de sa distribution). mais il tendra à sous-estimer ses besoins environnementaux (c.-à-d. sa niche écologique), ainsi que son aire géographique favorable. Je propose d'illustrer le problème avec deux exemples: les algorithmes BIOCLIM et Maxent. L'algorithme BIOCLIM est un modèle détermine la fourchette des valeurs observées par l'espèce pour chacune des variables indépendamment. Il définit donc une partie de l'espace multidimensionnel (l'hyper-volume de Hutchinson à n dimension) au sein duquel les conditions sont favorables à l'espèce (Hutchinson

1957, Booth et al. 2014). Plus le nombre de variables (= dimensions) est grand, plus le modèle sera par définition complexe, et donc plus l'aire géographique prédite comme favorable pour l'espèce aura tendance à être réduite. Quand la dimensionnalité est excessive, le modèle est surajusté. Dans le cas de Maxent, l'algorithme modélise les besoins de l'espèce de manière mathématiquement différente, mais il est aussi sujet à un possible sur-ajustement (Phillips et al. 2006, Warren & Seifert 2011, Rasosavljevic & Anderson 2014). Dans Maxent, la complexité du modèle est limitée par une pénalité appelée « régularisation », qui est appliquée à chaque variable inclue dans le modèle (Phillips & Dudík 2008, Merow et al. 2013). À mesure que la force de la pénalisation augmente, le modèle final tendra à être plus simple (par exemple, en ayant tendance à inclure de moins en moins de variables). Plusieurs études récentes ont montré que des modèles avec différents niveaux de complexité peuvent produire des prédictions géographiques de l'aire favorable d'une espèce qui varient énormément (Elith et al. 2010, Anderson & Gonzalez 2011, Shcheglovitova & Anderson 2013), mettant l'accent sur l'importance de bien prendre en considération la complexité du modèle. [Le même problème se pose pour les modèles qui ont pour but de modéliser la distribution d'une espèce].

CONCLUSIONS

Je réitère que les détails méthodologiques continuent de changer rapidement dans le domaine de la modélisation écologique de la niche, mais je sens que cette ligne de recherche est en train d'atteindre la maturité en ce qui concerne les principes fondamentaux (Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011). Je prédis donc que les quatre problèmes évoqués ici resteront vrai dans le futur. J'espère donc que cet essai inspiré par le symposium tenu à Puebla servira à orienter et aider les étudiants et chercheurs en 2015 et après.

REMERCIEMENTS

Avant tout je remercie Tania Escalante. Avec le grand support des coorganisateurs Elkin Noguera -Urbano et César Talonia, elle a orchestré le « 1er Symposium de Biogéographie : état des lieux et enjeux » au XIIème Congrès National Mexicain de Mammologie tenu dans la ville de Puebla de Zaragoza en octobre 2014. Enrique Martínez-Meyer, Claudia Moreno-Arzate, A. Monserrat Cruz-Lira, Ángela Cuervo et Ignacio Ferro ont apporté une assistance logistique primordiale. Diego Alvarado-Serrano a offert ses suggestions concernant le contenu de l'essai et corrigé mes nombreuses et créatives erreurs de grammaire dans la version originale espagnole. Olivier Broennimann et Corentin Bohl ont habilement traduit l'essai en français. Je remercie aussi la Commission Nationale Mexicaine pour Connaissance et Utilisation de la Biodiversité pour son support financier m'ayant permis de participer au meeting (CONABIO MU-011). La recherche ayant mené aux idées présentées ici a

été financée par la National Science Fondation des États-Unis (NSF DEB-0717357 and DEB-1119915) et par la Blavatnik Family Foundation. Je dédicace cet essai à toutes les personnes intéressées par les niches et distributions d'espèces et qui ont discuté le sujet avec moi durant le meeting.

REFERENCIAS /REFERENCES / RÉFÉRENCES

- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38: 541–545.
- Anderson RP, Gonzalez I. 2011. Speciesspecific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling* 222: 2796– 2811.
- Anderson RP, Martínez-Meyer E. 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation* 116: 167–179.
- Anderson RP, Raza A. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. Journal of Biogeography 37: 1378–1393.
- Anderson RP. 2003. Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography* 30: 591–605.
- Anderson RP. 2012. Harnessing the world's biodiversity data: promise and peril in

- ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences* 1260: 66–80.
- Anderson RP. 2013. A framework for using niche models to estimate impacts of climate change on species distributions. *Annals of the New York Academy of Sciences* 1297: 8 –28.
- Austin MP. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101–118.
- Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberón J, Villalobos F. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222: 1810–1819.
- Booth TH, Nix HA, Busby JR, Hutchinson MF. 2014. BIOCLIM: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions* 20: 1–9.
- Boria RA, Olson LE, Goodman SM, Anderson RP. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* 275: 73–77.
- Broennimann O, Fitzpatrick MC, Pearman PB,
 Petitpierre B, Pellissier L, Yoccoz NG,
 Thuiller W, Fortin MJ, Randin C,
 Zimmermann NE, Graham CH, Guisan A.

2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* 21: 481–497.

- Chase JM, Leibold MA. 2003. *Ecological niches: linking classical and contemporary approaches*. Chicago: University of Chicago Press.
- Elith J, Graham CH. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66–77.
- Elith J, Kearney M, Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1: 330–342.
- Elith J, Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677–697.
- Franklin J. 2010. *Mapping species distributions:* spatial inference and prediction. Cambridge: Cambridge University Press.
- Guisan A, Petitpierre B, Broennimann O, Daehler C, Kueffer C. 2014. Unifying niche shift studies: insights from biological invasions. Trends in Ecology and Evolution 29: 260–269.
- Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993–1009.
- Guisan A, Zimmermann NE. 2000. Predictive

- habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.
- Hutchinson GE. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415–427.
- Hutchinson GE. 1978. An introduction to population ecology. New Haven, Connecticut: Yale University Press.
- Kadmon R, Farber O, Danin A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications* 14: 401–413.
- Merow C, Smith MJ, Silander JA. 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography* 36: 1058–1069.
- Nogués-Bravo D. 2009. Predicting the past distribution of species climatic niches. Global Ecology and Biogeography 18: 521–531.
- Pagel J, Schurr FM. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. Global Ecology and Biogeography 21: 293–304.
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB. 2011. *Ecological niches and geographic distributions*. Princeton, New Jersey: Princeton University Press. *Monographs in Population Biology*, 49.
- Phillips SJ, Anderson RP, Schapire RE. 2006.

 Maximum entropy modeling of species

geographic distributions. *Ecological Modelling* 190: 231–259.

- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19: 181–197.
- Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175.
- Radosavljevic A, Anderson RP. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography* 41: 629 –643.
- Reddy S, Dávalos LM. 2003. Geographic sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography* 30: 1719–1727.
- Schurr FM, Pagel J, Cabral JS, Groeneveld J, Bykova O, O'Hara RB, Hartig F, Kissling WD, Linder HP, Midgley GF, Schröder B, Singer A, Zimmermann NE. 2012. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography* 39: 2146–2162.
- Shcheglovitova M, Anderson RP. 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecological Modelling* 269: 9–17.

- Soberón J, Nakamura M. 2009. Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19644–19650.
- Soberón J, Peterson AT. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London*, Series B, Biological Sciences 359: 689–698.
- Soberón J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115–1123.
- Soberón JM. 2010. Niche and area of distribution modeling: a population ecology perspective. *Ecography* 33: 159–167.
- Strubbe D, Beauchard O, Matthysen E. 2015.

 Niche conservatism among non-native vertebrates in Europe and North America. *Ecography* 38: 321–329.
- Urban MC, Zarnetske PL, Skelly DK. 2013.

 Moving forward: dispersal and species interactions determine biotic responses to climate change. *Annals of the New York Academy of Sciences* 1297: 44–60.
- Varela S, Anderson RP, García-Valdés R, Fernández-González F. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37: 1084–1091.
- Varela S, Rodríguez J, Lobo JM. 2009. Is current climatic equilibrium a guarantee for the

transferability of distribution model predictions? A case study of the spotted hyena. *Journal of Biogeography* 36: 1645–1655.

Veloz SD, Williams JW, Blois JL, He F, Otto-Bliesner B, Liu Z. 2012. No-analog climates and shifting realized niches during the late Quaternary: implications for 21st-century predictions by species distribution models. *Global Change Biology* 18: 1698–1713.

Veloz SD. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography* 36: 2290–2299.

Warren DL, Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21: 335–342.

CH. 2005. Wiens JJ. Graham Niche conservatism: integrating evolution. ecology, and conservation biology. Annual Review of Ecology, Evolution, and Systematics 36: 519-539.

Robert P. Anderson

Department of Biology, 526 Marshak Science Building City College of the City University of New York 160 Convent Avenue New York, NY 10031 USA

Graduate Center of the City University of New York 365 Fifth Avenue New York, NY 10016 USA

Division of Vertebrate Zoology (Mammalogy) American Museum of Natural History Central Park West at 79th Street New York, NY 10024 USA