



## Artigo de Revisão / Review Paper

# Desafios atuais da modelagem preditiva de distribuição de espécies

*Current challenges of species distribution predictive modelling*

Tereza C. Giannini<sup>1</sup>; Marinez F. Siqueira<sup>2</sup>; André L. Acosta<sup>1</sup>; Francisco C.C. Barreto<sup>3</sup>;

Antonio M. Saraiva<sup>4</sup> & Isabel Alves-dos-Santos<sup>1</sup>

### Resumo

A modelagem preditiva tem sido aplicada para analisar a distribuição geográfica de espécies, a partir de extrapolações das características ambientais dos locais conhecidos de ocorrência. O interesse por esse tipo de modelagem deve-se à necessidade de respostas rápidas e fundamentadas para as ameaças que as espécies têm enfrentado, devido à perda de habitat, invasão de espécies exóticas, mudanças climáticas, entre outros. Este artigo oferece uma visão geral dos avanços recentes no campo da modelagem e visa incentivar a discussão e aplicação desse método, que pode auxiliar tanto na aquisição de conhecimento básico sobre a biologia das espécies, quanto na análise e formulação de políticas para sua conservação.

**Palavras chaves:** nicho ecológico, áreas potenciais de ocorrência, algoritmos de modelagem.

### Abstract

Predictive modelling has been used to analyze species geographic distribution through extrapolation of environmental characteristics of known occurrence areas. Interest in this kind of modelling is derived from the need for rapid and well-grounded answers to the threats faced by species due to habitat loss, exotic species invasion, climate changes, and others. This article provides an overall view of recent advances in modelling and aims to encourage the discussion and application of this method that can help with basic biological knowledge acquisition as well as with public policies aimed at their conservation.

**Key words:** ecological niche, potential areas of occurrence, modelling algorithms.

### Introdução

A perda e fragmentação de habitats e as mudanças climáticas são apenas dois exemplos de alterações ambientais causadas por fatores antropogênicos, com consequências diretas sobre a distribuição das espécies. Essas ameaças crescentes demandam novas tecnologias e ferramentas de análise, para que se possa adquirir ou aprofundar o conhecimento existente sobre as espécies e auxiliar em sua proteção e conservação.

Em linhas gerais, a modelagem preditiva de distribuição de espécies consiste em um processamento computacional que combina dados de ocorrência de uma ou mais espécies

com variáveis ambientais, construindo assim uma representação das condições requeridas pelas espécies (Anderson *et al.* 2003). Alguns algoritmos têm sido aplicados para criar modelos que representam essas condições e que podem ser projetados sobre um mapa que exhibe as áreas potenciais de ocorrência dessas espécies.

A modelagem de distribuição tem sido amplamente utilizada com múltiplos objetivos, tais como: utilização de modelos de distribuição potencial em análises biogeográficas (Siqueira e Durigan 2007); conservação de espécies raras ou ameaçadas (Araújo e Williams 2000; Engler *et al.* 2004); reintrodução de espécies (Hirzel *et al.*

<sup>1</sup> Universidade de São Paulo, Instituto de Biociências, R. do Matão trav. 14, 05508-900, São Paulo, SP.

<sup>2</sup> Instituto de Pesquisa Jardim Botânico do Rio de Janeiro, R. Pacheco Leão 915, 22460-030, Rio de Janeiro, RJ.

<sup>3</sup> Universidade Federal do Espírito Santo, Centro de Ciências Humanas e Naturais, Av. Fernando Ferrari, 29075-910, Vitória, ES.

<sup>4</sup> Universidade de São Paulo, Escola Politécnica, Av. Prof. Luciano Gualberto 380, 05508-970, São Paulo, SP.

<sup>5</sup> Autor para correspondência: giannini@usp.br

2002); perda de biodiversidade (Polasky e Solo 2001); impactos de mudanças climáticas (Peterson *et al.* 2002; Oberhauser e Peterso 2003; Siqueira e Peterson 2003; Thomas *et al.* 2004; Pearson *et al.* 2006; Araújo *et al.* 2006; Araújo *et al.* 2008; Wiens *et al.* 2009); avaliação do potencial invasivo de espécies exóticas (Peterson *et al.* 2003; Peterson e Robins, 2003; Peterson *et al.* 2006a); estudo das possíveis rotas de disseminação de doenças infecciosas (Costa *et al.* 2002; Peterson *et al.* 2006b; Levine *et al.* 2007); auxílio na determinação de áreas prioritárias para conservação (Ortega-Huerta e Peterson 2004; Chen 2009), entre outros.

Este artigo apresenta os fundamentos ecológicos que estão na base da conceituação e aplicação da modelagem preditiva da distribuição de espécies. Na sequência apresenta também um breve histórico de seu desenvolvimento com algumas das principais contribuições ao avanço dessa técnica. Com o objetivo de promover a aplicação deste tipo de estudo, as etapas do processo de modelagem também são discutidas de forma sucinta, e finalmente, são sintetizados os principais desafios ainda existentes neste campo.

### Fundamentos ecológicos da modelagem preditiva de distribuição de espécies

A modelagem preditiva tem sido chamada por alguns autores de modelagem de distribuição geográfica (Austin 2002) e por outros, de modelagem de nicho ecológico (Austin 1992; Soberón e Peterson 2005). A ausência de consenso foi relacionada, entre outros fatores discutidos adiante, com a dificuldade de se definir, de forma consistente, o conceito de nicho (Elith e Leathwick 2009).

A primazia da aplicação do conceito de nicho é atribuída frequentemente a Joseph Grinnel que discutiu, em uma série de artigos (Grinnell 1904, 1917, 1924), o nicho de várias espécies de aves e mamíferos. De acordo com suas descrições, o nicho consistiria não exatamente em um atributo da espécie, mas sim do espaço (Schoener 1989), e definiria as condições ou situações ambientais nas quais a espécie seria capaz de sobreviver e se reproduzir.

Posteriormente, Elton (1927) apresentou o nicho de uma espécie como seu papel dentro da cadeia alimentar, analisando tanto o impacto sofrido pela espécie, quanto o que ela causa em seu ambiente.

A conceituação de nicho ecológico deu um grande salto a partir dos trabalhos de George Evelyn Hutchinson, que definiu o termo nicho como a soma de todos os fatores ambientais que atuam sobre o organismo: “uma região de um hiper-espaço  $n$ -dimensional!” (Hutchinson 1944), sendo que o termo “hiper-espaço” foi substituído posteriormente por “hiper-volume” (Hutchinson 1957). Segundo o autor, definido um número qualquer ( $n$ ) de fatores limitantes (por exemplo, luminosidade, temperatura e recursos alimentares) à sobrevivência do organismo, a amplitude de tolerância a estes fatores, ou seja, os valores mínimos e máximos dos limites de tolerância para cada fator constituiria então um hiper-volume com  $n$  dimensões. Portanto, o espaço abrangido por este hiper-volume representaria a faixa de variação dos fatores ambientais na qual a espécie é capaz de sobreviver e se reproduzir.

Recentemente, buscando uma abordagem quantitativa e mais rigorosa em termos estatísticos, Chase e Leibold (2003) propuseram uma definição de nicho que resgata a conceituação dos três principais autores citados acima (Grinnell 1924; Elton 1927; Hutchinson 1957). O nicho corresponderia às “condições ambientais que permitem à espécie suprir suas necessidades mínimas, de tal forma que a taxa de natalidade de uma população local seja igual ou maior à sua taxa de mortalidade, junto com o conjunto de efeitos *per capita* da espécie nessas condições ambientais<sup>2</sup>”.

Por outro lado, a distribuição geográfica de uma dada espécie foi considerada como uma expressão complexa de sua ecologia e evolução, influenciada por vários fatores que operam em intensidades e escalas diferentes tais como, a história evolutiva da espécie e sua capacidade de dispersão (Brown e Lomolino 2006). Por exemplo, Pulliam (2000) discutiu a complexidade da dinâmica de dispersão diferenciando habitats do tipo fonte e dreno. Os habitats fonte são aqueles nos quais a reprodução local excede a mortalidade, e nos drenos ocorre o oposto. Por definição, os drenos não apresentam condições necessárias para a espécie sobreviver e se reproduzir, embora os indivíduos

<sup>1</sup> No original: “The term niche is here defined as the sum of all the environmental factors acting on the organism; the niche thus defined is a region of an  $n$ -dimensional hyper-space.” (Hutchinson 1944).

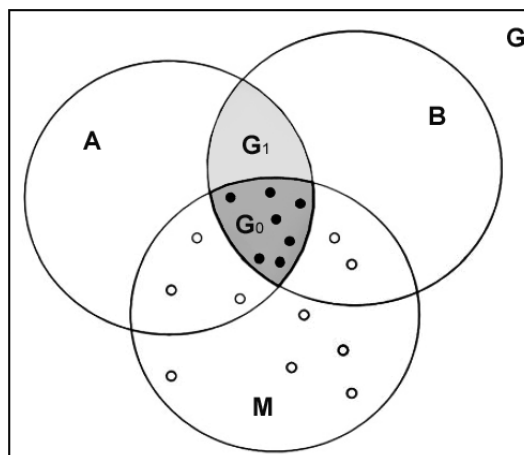
<sup>2</sup> No original: “the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions.” (Chase e Leibold 2003, p. 15)...

possam ocorrer nessas áreas devido aos processos de imigração a partir de suas fontes. Assim, é possível encontrar indivíduos em locais onde as condições não são inteiramente adequadas à sobrevivência e à reprodução da espécie. Por outro lado, as espécies podem apresentar limites de dispersão, e conseqüentemente, não chegar aos locais adequados. Portanto, a dinâmica fonte-dreno prevê que uma espécie possa ser encontrada fora dos limites de seu nicho fundamental, enquanto que, ao contrário, o limite na capacidade de dispersão prevê que certos locais, potencialmente adequados, possam não ser ocupados por determinadas espécies.

Considerando essa dinâmica, Soberón (2010) sugeriu três fatores principais que determinam a distribuição geográfica, sendo os dois primeiros uma função direta do nicho (Fig. 1): 1) condições abióticas (por exemplo, o clima); 2) fatores bióticos (por exemplo, interações entre espécies); 3) capacidade de dispersão, através de sua locomoção ou através de agentes externos, o que determina o grau de acessibilidade da espécie a outras áreas.

Segundo Soberón e Peterson (2005), “os modelos de nicho têm sido validados universalmente através de comparações entre ocorrências no espaço geográfico, embora eles sejam, estritamente falando, modelos de nicho ecológico e não de distribuição geográfica”<sup>3</sup>. Já Phillips (2008) discute que a modelagem, em alguns casos, tenta estimar o quão satisfatório um local é para a espécie considerando-se suas condições ambientais. Ou seja, em termos estatísticos, o resultado da modelagem indica se um local é satisfatório para a espécie e não exatamente se ele está sendo ocupado.

Jiménez-Valverde *et al.* (2008b) afirmaram que o termo modelagem de nicho deveria ser evitado, pois além de ser complexo, implica em considerar os efeitos tanto dos fatores bióticos quanto dos abióticos que atuam sobre o organismo. Segundo os autores, citados anteriormente, os modelos são capazes de projetar simulações da distribuição das espécies no espaço geográfico, mas não oferecem uma descrição do nicho das espécies, pois, embora haja disponibilidade de dados relacionados com condições ambientais, os dados sobre interações ainda são escassos ou de difícil interpretação. Como resultado, a maioria



**Figura 1** — Fatores que determinam a distribuição geográfica de uma espécie. G: região geográfica considerada; A: área onde as condições abióticas são favoráveis para a espécie. Esta é uma região potencial e, provavelmente, não ocupada pela espécie em todas as suas seções; B: região onde as condições bióticas são favoráveis; M: área que a espécie foi capaz de alcançar em um dado período;  $G_1$ : região que pode ser potencialmente invadida porque os dois tipos de condições (A e B) são favoráveis, mas a espécie ainda não foi capaz de alcançá-la;  $G_0$ : área real de distribuição da espécie. Os círculos fechados representam a população fonte e os abertos, a população dreno (Soberón 2010).

**Figure 1** — Factors determining the species geographical distribution. G: geographical region under consideration; A: area in which abiotic conditions are favorable for a species. This is a potential region and, probably, unoccupied by the species in some sections; B: region in which the biotic conditions are suitable; M: area that the species has been able to reach within a given time period;  $G_1$ : region that may potentially be invaded because both types of conditions (A and B) are suitable, but that the species has not yet been able to reach;  $G_0$ : actual area of distribution of the species. The closed circles represent source populations. The open circles represent sink populations (Soberón 2010).

dos exemplos na literatura tem utilizado apenas as variáveis abióticas em suas análises e, assim, a modelagem deveria ser considerada como modelagem de distribuição e não de nicho (Araújo e Guisan 2006; Jiménez-Valverde *et al.* 2008b; Phillips 2008; Elith e Leathwick 2009).

Soberón e Nakamura (2009, p. 19644) sintetizaram a questão da seguinte forma: “Essencialmente, essa é uma técnica usada para estimar áreas de distribuição reais ou potenciais, ou conjuntos de habitats favoráveis para uma dada espécie, com base em sua presença e (às vezes)

<sup>3</sup> No original: “Niche models have universally been validated via comparisons with occurrences in geographic space, even though they are, strictly speaking, models of ecological niches and not geographic distributions.” (Soberón e Peterson 2005, p. 8).

ausência observadas. Estes métodos relacionam “nicho” com “área de distribuição”. As aspas são usadas para indicar que definições rigorosas desses conceitos ainda não foram apresentadas<sup>4</sup>. Assim, frente à falta de definições mais precisas, o termo modelagem preditiva de distribuição de espécies foi escolhido para este trabalho por estar sendo bastante utilizado, principalmente após o trabalho de Lobo *et al.* (2008). Também acreditamos que este termo seja um pouco menos genérico que “ecological niche modeling” e um pouco mais informativo que “species distribution modeling”, ambos extensivamente utilizados na literatura científica.

### Principais contribuições ao desenvolvimento da modelagem preditiva de distribuição de espécies

Segundo Guisan e Thuiller (2005) as técnicas computacionais desenvolvidas para analisar o impacto das variáveis ambientais na distribuição de uma espécie surgiram principalmente na década de 1970, com os trabalhos de Austin (1971) e Nix *et al.* (1977). Nas décadas de 1980-1990 as simulações de Ferrier (1984) e os livros de Verner *et al.* (1986) e Margules e Austin (1991) foram importantes contribuições para promover essa nova abordagem, bem como as revisões de Franklin (1995) e Austin (1998).

O desenvolvimento de algoritmos (sequências finitas de comandos, executados através de códigos escritos em alguma linguagem computacional) específicos para a modelagem preditiva de distribuição foi também fundamental para a evolução dessa técnica. Destacam-se os algoritmos Maxent (Phillips *et al.* 2006) e o Genetic Algorithm for Rule Set Production (GARP) (Stockwell e Noble 1991; Stockwell e Peters 1999). Ambos utilizam apenas dados de presença e apresentam bons resultados com baixo número de pontos de ocorrência (Wisz *et al.* 2008). Outros algoritmos importantes a serem destacados, e que necessitam de dados de presença e de ausência, são o Generalised Linear Models (GLM) (McCullagh e Nelder 1989), Generalised Additive Models (GAM) (Hastie e Tibshirani 1990) e o Random Forest (RF) (Breiman 2001).

<sup>4</sup>No original: “Essentially, it is a technique used to estimate actual or potential areas of distribution, or sets of favorable habitats for a given species, on the basis of its observed presences and (sometimes) absences. These methods relate “niches” to “areas of distribution.” The quotes are used to indicate that rigorous definitions of those concepts have not as yet been presented.” (Soberón e Nakamura 2009, p. 19644).

O desenvolvimento da modelagem preditiva de distribuição de espécies aconteceu concomitante com iniciativas internacionais que visavam, principalmente, a padronização, compartilhamento e disponibilização de dados primários de coleções biológicas, museus e herbários (Graham *et al.* 2004). Atualmente, as principais iniciativas neste sentido são a do BIS/TDWG (Biodiversity Information Standards/Taxonomic Database Working Group - desde 1985), voltado principalmente para o desenvolvimento de padrões para compartilhar dados; ITIS (The Integrated Taxonomic Information System - desde meados da década de 90), com a meta de criar um banco de dados sobre informações taxonômicas; GBIF (Global Biodiversity Information Facility – desde 1996), cujo objetivo primário é o de tornar disponíveis os dados sobre biodiversidade e; IABIN (Inter-American Biodiversity Information Network – desde 1996) com a finalidade de promover colaboração técnica sobre o armazenamento, uso e compartilhamento de informações sobre biodiversidade. No Brasil foi criada a rede *speciesLink*, uma iniciativa do CRIA, com o objetivo geral de disponibilizar dados sobre a biodiversidade brasileira. Outras iniciativas importantes são o programa Biota FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo) e o Sis-Biota Brasil, ambos com o objetivo de inventariar e caracterizar a biodiversidade, respectivamente, desse estado e do país.

Elith & Leathwick (2009) destacaram também a importância dos avanços em métodos que permitiram maior robustez e detalhamento nos modelos geográficos, avanços esses que incluem a geração e disponibilização de variáveis ambientais em formato digital, como por exemplo: 1) variáveis climáticas: Worldclim (Hijmans *et al.* 2005) e Köppen-Geiger “climate classification” (Rubel e Kottek 2010); 2) ecorregiões: WWF (World Wildlife Fund – Olson *et al.* 2001); 3) edáficas: USGS (United States Geological Survey) e 4) modelos de elevação: USGS/EROS (United States Geological Survey Center for Earth Resources Observation and Science). Destaca-se também no Brasil, a iniciativa do IBGE (Instituto Brasileiro de Geografia e Estatística) em prover dados associados a diversas características ambientais (<<http://www.ibge.gov.br>>).

Em duas revisões recentes sobre modelagem preditiva de distribuição de espécies (Elith & Leathwick 2009; Soberón & Nakamura 2009),



os autores sugeriram que entre os desafios ainda existentes para a modelagem está a inclusão de interações bióticas em seu processo. Assim, merecem destaque as iniciativas dedicadas à disponibilização de dados de interações entre organismos, especialmente plantas e polinizadores. Exemplos importantes são o desenvolvimento de extensões de padrões de troca de dados originalmente propostos pelo BIS/TDWG (Cartolano Junior *et al.* 2007; Saraiva *et al.* 2009b), a iniciativa da IABIN-PTN (Inter-American Biodiversity Information Network - Pollinators Thematic Network) (Saraiva *et al.* 2009a) e a Rede de Polinizadores (Canhos *et al.* 2010). Os dados e análises subsequentes, tornadas acessíveis a partir dessas iniciativas, poderão, num futuro próximo, auxiliar a pesquisa sobre interações, bem como, expandir a aplicação da modelagem em estudos de nicho e distribuição geográfica de espécies.

Ainda existem poucos exemplos na literatura de aplicações da modelagem para analisar organismos que apresentam algum tipo de interação. São encontrados exemplos deste tipo de aplicação em interações de competição (Leathwick e Austin 2001; Anderson *et al.* 2002), predação (Sánchez-Cordero e Martínez-Meyer 2000) e entre plantas e seus polinizadores (Jiménez-Valverde *et al.* 2008a; Godsoe *et al.* 2009; Giannini *et al.* 2010; Giannini *et al.* 2011). Por exemplo, Heikkinen *et al.* (2007) analisaram o efeito da inclusão de dados de interação de espécies mutualistas na modelagem de quatro espécies focais de aves (*Surnia ulul*, *Glaucidium passerinum*, *Strix aluco* e *Aegolius funereus*). Os autores incluíram os dados de ocorrência das espécies mutualistas como variáveis preditoras na modelagem, o que resultou no aumento significativo da acurácia dos modelos obtidos. Já Araujo e Luoto (2007) demonstraram a importância de se incluir dados de interação quando são analisados os possíveis efeitos das mudanças climáticas sobre a distribuição de uma espécie de borboleta (*Parnassius mnemosyne*) e suas plantas associadas. Eles testaram duas hipóteses nulas. Uma delas era de que as interações não tinham um papel significativo em explicar a distribuição atual da espécie de borboleta. A outra hipótese era que as interações não tinham um papel significativo em predições de alteração na distribuição baseadas em mudanças climáticas. Os resultados mostraram que as duas hipóteses eram falsas oferecendo suporte para a proposição de que as interações afetam significativamente a distribuição das espécies.

Considerando a importância das interações entre organismos e a influência das características geográficas sobre essas interações, conclui-se que este é um campo fértil e em aberto no desenvolvimento da área de modelagem preditiva.

### Etapas do processo de modelagem preditiva de distribuição de espécies

A modelagem de distribuição de espécies é um processo complexo que compreende várias etapas, as quais exigem o conhecimento de diversos conceitos e técnicas ligados não só a biologia, mas também a biodiversidade, geografia e clima, e ao processamento de dados requerido antes, durante e depois da obtenção do modelo. As seguintes etapas são sugeridas para executar a modelagem (baseado em Santana *et al.* 2008).

#### 1. Tratamento dos pontos de ocorrência das espécies

A primeira etapa consiste na obtenção, avaliação e preparo dos pontos de presença e ausência (quando disponíveis) das espécies a serem modeladas, o que pode incluir o georreferenciamento dos pontos e a exclusão de dados imprecisos ou duvidosos. Os dados de ausência são muito raros e difíceis de serem obtidos e, por esta razão, alguns algoritmos (Bioclim, Environmental Distance, Maxent) têm sido utilizados na modelagem por demandarem apenas dados de presença (Elith & Leathwick 2009). Os dados podem ser adquiridos *in loco*, através de levantamentos em áreas alvo, fornecendo dados inéditos à análise e/ou através de fontes de dados já publicados em formato digital como o GBIF (<http://www.gbif.org>), speciesLink (<http://slink.cria.org.br>), IABIN (<<http://www.iabin.net>>) e ainda, por meio de literatura acadêmica especializada.

Existem também algumas propostas para a geração de dados de pseudo-ausência para uso com algoritmos que obrigatoriamente precisam desse tipo de dado. Duas dessas propostas podem ser encontradas através da aplicação de ENFA (Ecological-Niche Factor Analysis) (Hirzel *et al.* 2002) ou das estratégias presentes no pacote BIOMOD (Thuiller 2003). Tem sido sugerido que a definição dos pontos de pseudo-ausência deve seguir alguma estratégia e não ser feita aleatoriamente (Engler *et al.* 2004; McPherson *et al.* 2004). Em relação à localização dos pontos, VanDerWal *et al.* (2009) sugeriram que

a performance do modelo é menor quando os pontos de pseudo-ausência são definidos em uma região muito restrita ou muito ampla em relação às presenças registradas, sugerindo que o melhor seria considerar regiões intermediárias para essa escolha. Porém, mais recentemente, Lobo & Tognelli (2011) demonstraram que a localização dos pontos de pseudo-ausência afeta menos o modelo que o número de pontos. De acordo com a análise feita por estes autores, os resultados sugeriram que o aumento de número de pontos de pseudo-ausência aumenta a acurácia dos modelos, sendo sugerida uma relação de 100 pontos de pseudo-ausência para cada ponto de presença. Além disso, um maior número de pontos de pseudo-ausência acarretaria também em uma diminuição na taxa de sobreprevisão, uma tendência que o modelo pode apresentar de determinar certas áreas para as quais não existem ocorrências reportadas, como sendo potencialmente adequadas para a espécie (ver detalhes acerca de sobreprevisão abaixo).

Os dados de coleções biológicas e museus podem apresentar limitações devido à imprecisão da informação ou viés na escolha de áreas de coletas que, às vezes, privilegia locais onde o acesso é mais fácil e menos custoso, ou ainda, onde já se tem certeza da ocorrência da espécie em estudo (Hortal *et al.* 2008; Lobo 2008). Apesar disso, o uso desse tipo de dado é frequentemente justificado devido à falta de levantamentos sistemáticos, além da demanda crescente por mapas de distribuição de espécies e estimativas sobre mudanças potenciais nessa distribuição devido às alterações climáticas (Thuiller *et al.* 2008; Elith e Leathwick 2009).

Visando estimar um número mínimo de pontos de ocorrência para modelagem, Wisz *et al.* (2008) utilizaram três classes em suas análises, com 10, 30 e 100 pontos cada. Os autores concluíram que os algoritmos Maxent e OM-GARP são os melhores para analisar 10 pontos, mas modelos com maior acurácia foram obtidos com pelo menos 30 pontos, considerado pelos autores como um número mínimo. Na época da publicação do referido artigo, o algoritmo OM-GARP estava ainda em fase de teste, mas foi posteriormente disponibilizado no sistema computacional para modelagem, o openModeller (Muñoz *et al.* 2011).

No entanto, outros autores sugeriram números menores de pontos de ocorrência (Stockwell e Peterson 2002) e outros, discutiram a utilização de métodos específicos para análises com número

muito baixo de ocorrências, como no caso de espécies raras (Siqueira *et al.* 2009) ou crípticas (Pearson *et al.* 2007).

## 2. Tratamento e escolha das camadas ambientais

Essa etapa visa identificar, adquirir e converter as camadas ambientais, geralmente utilizadas em formato “raster”, para serem combinadas aos pontos de ocorrência conhecidos das espécies e serem produzidos os modelos. O formato “raster” é comum aos programas de SIG (Sistema de Informação Geográfica) e consiste em bancos de dados georreferenciados exibidos como uma matriz de células, sendo que o tamanho destas células é o que determina a resolução do “raster”. Quanto maior o tamanho da célula, menor a resolução, e vice versa. Em alguns casos, se não estiverem em formato “raster”, é necessário que as camadas sejam convertidas a esse formato por um programa de SIG, como o ArcGIS (ESRI Inc.) ou o DIVA GIS (LizardTech, Inc. and University of California), sendo este último, um programa de código aberto disponível na internet (outros programas gratuitos são o GRASS GIS e o Quantum GIS, ambos da Open Source Geospatial Foundation).

É importante também, definir adequadamente a resolução das camadas ambientais para garantir a qualidade do modelo final. Elementos importantes a serem considerados na escolha são: a precisão dos pontos de ocorrência, as características do terreno e da espécie (Elith e Leathwick 2009) e o tamanho da área geográfica a ser analisada (Chapman *et al.* 2005). Por outro lado, Guisan *et al.* (2007) demonstraram que a resolução das camadas utilizadas tem pouco efeito sobre os modelos, enfatizando a dificuldade que existe em estimar adequadamente esse fator. Mais recentemente, Soberón (2010) demonstrou que o uso de alta resolução justifica-se especialmente se informações bióticas (interações entre espécies, por exemplo) forem consideradas na modelagem – do contrário, caso sejam usados apenas dados abióticos, o uso de resolução mais baixa é o suficiente para produzir bons modelos. Chapman *et al.* (2005) sugerem que para análises em escala local podem ser usadas camadas com 30 segundos de arco, e que para análises em escala continental sejam usadas camadas com 2,5 ou 5 minutos de arco.

Quanto à seleção das variáveis ambientais, de maneira geral, é indicado evitar-se o uso

de camadas de variáveis que sejam altamente correlacionadas (Guisan & Thuiller 2005), sendo recomendada a aplicação de técnicas específicas de seleção de variáveis. Inicialmente foram utilizadas técnicas conhecidas como “jackknife” para a seleção de variáveis, onde os modelos eram gerados e uma camada ambiental era retirada por vez (“leave one out”) e a implicação da remoção desta variável no processo era então avaliada. Este método tem inconvenientes, sendo o mais importante deles a dependência do algoritmo, pois algoritmos diferentes podem responder de forma diferente à remoção das variáveis. Com o advento de novos algoritmos, cada um com diferenças intrínsecas na forma de processar os dados e na forma de apresentar o resultado final, este método foi sendo gradualmente descartado. Alguns autores utilizaram técnicas de análise multivariada como uma forma de avaliar a contribuição das variáveis na variação dos dados como um indicativo de sua importância na distribuição das espécies, sendo um exemplo o uso de Análises de Componentes Principais (Peterson *et al.* 2007). Também é comum a aplicação de índices de correlação para remover as variáveis altamente correlacionadas, diminuindo assim a multicolinearidade entre os dados (Rushton *et al.* 2004; Guisan & Thuiller 2005) ou ainda a aplicação de técnicas de mineração de dados (Hochachka *et al.* 2007) para a seleção de variáveis preditoras. Quanto ao número de variáveis, é reconhecido entre os estatísticos que o tamanho da amostra em testes de regressão logística deve ser no mínimo, 10 vezes maior que o número de variáveis preditoras utilizado (Peduzzi *et al.* 1996) mas ainda não existem estudos suficientes para um consenso sobre as implicações dessa proporção em algoritmos utilizados na modelagem preditiva.

### 3. Escolha dos algoritmos de modelagem

Um dos algoritmos frequentemente utilizados para modelagem preditiva é o Maxent (Phillips *et al.* 2006). Esse algoritmo demonstrou um bom desempenho geral, além de ter sido o que demonstrou melhor resultado com poucos pontos de ocorrência e com dados apenas de presença (Wisz *et al.* 2008), sendo indicado caso se tenha esse tipo de dado. Porém, vários outros algoritmos de modelagem podem ser encontrados em alguns sistemas computacionais, tais como o

openModeller (Muñoz *et al.* 2011), que conta com uma ampla variedade de algoritmos integrados, incluindo o GARP, o ENFA, o Support Vector Machine (SVM), o Bioclim, e o Environmental Distance. Outro sistema disponível é o BIOMOD (Thuiller, 2003; Thuiller *et al.* 2009), acrônimo de Biodiversity Modelling, desenvolvido para a plataforma R (The R Foundation for Statistical Computing). Nesse pacote podem ser encontrados nove algoritmos, entre eles, o GLM, GAM e o Random Forest.

Como ressaltado anteriormente, a escolha do algoritmo deve ser baseada na pergunta do estudo e na disponibilidade de dados de ocorrência (número de registros e dados de presença ou de presença/ausência). Por isso não deve ser considerado que deve haver um consenso sobre o algoritmo a ser utilizado; a escolha deverá ser considerada caso a caso. Recentemente, Elith & Graham (2009), Elith & Leathwick (2009) e Kamino *et al.* (2011) apresentaram revisões onde foram comparados os resultados produzidos por alguns algoritmos, e discutidas as características, vantagens e desvantagens na utilização de cada um deles.

### 4. Avaliação dos modelos gerados

A avaliação da qualidade dos modelos gerados é ainda um tema em evolução, pois a forma como as avaliações são feitas estatisticamente ainda é incompleta e não totalmente resolvida na literatura (Fielding & Bell 1997; Peterson *et al.* 2008; Jimenez-Valverde 2011). A validação de um modelo é também conhecida como teste externo e consiste da confrontação de seus resultados contra dados da distribuição da espécie que não foram utilizados pelo algoritmo no processo de modelagem. Isso distingue esse processo do chamado “teste interno”, também conhecido como verificação, no jargão da modelagem matemática. Basicamente a diferença entre elas é que no caso da validação, os dados de teste nunca são conhecidos pelo algoritmo, enquanto na verificação os dados podem ser usados ora como treino ora como teste, principalmente nos algoritmos que utilizam várias interações para chegar à solução do problema. Discussões sobre a melhor forma para validação de modelos pode ser encontrada em Pearce e Ferrier (2000) e Lobo *et al.* (2008). De qualquer forma, para avaliar a qualidade do modelo, é preciso quantificar os erros que podem acontecer no processo.

#### 4.1 Tipos de erros associados aos modelos preditivos de distribuição de espécies

Para avaliar o resultado da modelagem, dois tipos de erros são mensurados, baseados na matriz de confusão (Phillips *et al.* 2006) (Fig. 2). São eles:

4.1.a) Erro de omissão (subprevisão): esse tipo de erro é geralmente considerado como um erro verdadeiro, representado pela letra C na matriz de confusão (Fig. 2). Contudo, sob algumas circunstâncias, um registro de presença pode não ser muito confiável quanto à sua identificação taxonômica e/ou georrefenciamento. Estes pontos podem representar um “outlier” para o algoritmo, ou seja, pontos com informação ambiental muito fora do padrão gerado pelos demais pontos do conjunto de treino. Nessas circunstâncias, um erro de omissão não seria um erro, e sim uma forma do algoritmo conferir menos importância para pontos considerados “ruins”. Uma consequência direta disso, caso se tenha dúvida sobre a qualidade dos registros de ocorrência da espécie, é que não é recomendado usar taxa de omissão de 0% ao se rodar um modelo. É preferível deixar uma margem de segurança (por exemplo, até 10%) para que o algoritmo possa trabalhar melhor essa questão. Porém, na literatura atual há uma predominância de trabalhos utilizando 5% de taxa de omissão (Elith *et al.* 2006; Phillips *et al.* 2006).

	presença real	ausência real
presença prevista	a	b
ausência prevista	c	d

a = verdadeiro positivo

b = falso positivo

c = falso negativo

d = verdadeiro negativo

a e d são acertos

b e c são erros

taxa de falso positivo = erro de sobreprevisão =  $b/(b+d)$

taxa de falso negativo = erro de comissão =  $c/(a+c)$

sensibilidade =  $a/(a+c)$

especificidade =  $d/(b+d)$

acurácia =  $a+d/(a+b+c+d)$

**Figura 2** — Matriz de confusão, onde a e d são previsões corretas; b é erro de sobreprevisão (falsos positivos) e c é erro de omissão (falsos negativos).

**Figure 2** — Confusion matrix, where a and d are right predictions; b is the overprediction error (false positive) and c is the omission error (false negative).

4.1.b) Erro de sobreprevisão (comissão): este pode ou não ser um erro verdadeiro. Consiste na previsão de ocorrência em uma área na qual não se tem registro da presença da espécie, e é representado pela letra B na matriz de confusão (Fig. 2). Este erro pode ser causado por diferentes situações: a) a área é habitável para a espécie, mas não se tem um esforço amostral suficiente na região para afirmar se a espécie ocorre ou não, indicando lacunas de conhecimento; b) a área é habitável para a espécie, mas fatores históricos ou ecológicos (barreiras geográficas, capacidade de dispersão) ou bióticos (competição, predação) impediram a espécie de chegar ou de se estabelecer na região; c) a área é realmente inabitável, o que consistiria no verdadeiro erro de sobreprevisão. De qualquer forma, não é considerado um erro grave, pois em algumas situações, tais como previsão de áreas de potencial invasão por espécies exóticas, ou rotas de disseminação de vetores de doenças, ou ainda em projeções de distribuições futuras ou no passado, a sobreprevisão, ou uma maior capacidade preditiva do modelo, é necessária.

#### 4.2 Geração de um conjunto independente de dados

Para quantificar os tipos de erros é preciso gerar um conjunto independente de dados. Existem duas formas de fazê-lo: 1) coletando novos dados em trabalho de campo ou levantamento de literatura; 2) dividindo o conjunto de dados em conjuntos de treino e teste antes de realizar a modelagem.

Independente da estratégia escolhida é fundamental ter dois (ou mais) conjuntos de dados diferentes, uma vez que, um conjunto será utilizado para gerar o modelo (dados de treino) e outro irá testá-lo (dados de teste).

A forma mais comum de se dividir o conjunto de dados é através da aplicação de números aleatórios à tabela de registros de ocorrências, geralmente 50% dos dados para gerar o modelo e 50% para testar (Fielding e Bell 1997; Hirzel e Guisan 2002) ou 70% para treino e 30% para teste. Outra forma é dividir os dados espacialmente (geograficamente) e utilizar uma parte como treino e outra como teste (Peterson *et al.* 2008).

Atualmente tem-se usado técnicas de “bootstrap” (Efron 1979) para gerar diferentes conjuntos de dados e avaliar assim a precisão dos modelos de distribuição. Esta técnica envolve a partição dos dados aleatoriamente, com reposição,



em vários conjuntos de treino e teste. Esta técnica de validação tem sido frequentemente utilizada, juntamente com “Cross-validation”, “Subsamples” e “Jackknife”. A “Cross-validation” consiste basicamente em dividir aleatoriamente o banco de dados em  $k$  partições, dessas,  $k-1$  são utilizadas para treinar o algoritmo e a partição restante é utilizada para o teste. O processo é repetido  $k$  vezes. Independente do método, o erro obtido em cada uma das partições é calculado obtendo-se um valor médio da qualidade do modelo gerado. O “jackknife” (ou leave one out) é um tipo especial de “cross-validation”, e é utilizado com amostras pequenas (Pearson *et al.* 2007), onde  $k=N$  (onde  $N$  = número de amostras). O “subsample” consiste em parcionar os dados  $k$  vezes e depois dividir cada partição nas porcentagens de treino e teste. Basicamente, o que definirá a técnica a ser aplicada é o número de dados disponíveis. “Cross-validation” e “subsamples” precisam de uma quantidade maior de pontos, enquanto “bootstrap” e “jackknife” podem ser aplicadas para amostras menores. O programa Maxent (Phillips *et al.* 2006) já traz essas técnicas implementadas, e outros ambientes de modelagem trazem apenas a técnica “bootstrap” e com apenas uma partição, como é o caso do openModeller (Munõz *et al.* 2011).

A validação é uma das etapas mais importantes do processo de modelagem. Sem ela, a interpretação de um modelo perde seu sentido, visto que tudo o que está representado pode estar incorreto ou com graus inaceitáveis de imprecisão.

Os métodos mais comuns para avaliar a qualidade do modelo são baseados na matriz de confusão (Fig. 2), mas ao comparar resultados de modelagens realizadas por diferentes algoritmos, é necessário utilizar também métodos que sejam independentes de limites de corte. Portanto, temos dois tipos de avaliação, a dependente e a independente de um limite de corte específico.

#### 4.3 Estabelecimento do limite de corte

As medidas tradicionais de capacidade discriminatória de um modelo dependem da adoção de um limite de corte arbitrário (“threshold”) o que acaba inserindo mais um fator de subjetividade na interpretação do modelo. Uma boa revisão sobre limites de corte pode ser encontrada em Liu *et al.* (2005).

A escolha desse limite de corte é usualmente baseada no conhecimento da probabilidade da

ocorrência da espécie de interesse (por exemplo, o conhecimento sobre a amplitude da distribuição esperada da espécie, ampla ou restrita) assim como, no julgamento das consequências em tomar decisões erradas no processo e, além disso, está também diretamente relacionada com o propósito da modelagem. Por exemplo, se uma espécie é ameaçada e o objetivo do modelo é identificar áreas potenciais para maximizar o sucesso de sua reintrodução na natureza, é preciso identificar áreas adequadas que minimizem a chance de erro, ou seja, modelos menos inclusivos (com baixa sobreprevisão). Isso requer um limite de corte alto, que selecione apenas as áreas com altos valores de adequabilidade ambiental para a ocorrência da espécie. Já quando o interesse é avaliar o potencial invasivo de espécies exóticas, ou aumentar o conhecimento sobre o nicho de uma espécie pouco coletada, pode-se adotar um limite de corte mais baixo (maiores áreas de sobreprevisão) para aumentar a área de interesse em relação à ocorrência das espécies. O limiar atualmente mais utilizado na literatura é o MPT (Minimum Presence Training ou ainda Lowest Presence Training) (Pearson *et al.* 2007), mas cada valor de corte tem vantagens e desvantagens de uso. A escolha deve sempre estar alinhada com o que se pretende fazer com o modelo gerado, ou seja, depende do propósito da modelagem.

#### 4.4 Tipos de análise

Todo modelo apresenta seus erros e acertos, que são avaliados em conjunto para determinar a qualidade do mesmo. A Figura 2 apresenta a matriz de confusão, ou seja, o esquema que reúne as possíveis formas de acerto e erro em relação às previsões do modelo e à distribuição “real” da espécie na natureza.

Para se calcular os valores da matriz de confusão é necessário aplicar um limite de corte para avaliar o modelo. Depois de estabelecido o limite de corte, testes estatísticos podem ser aplicados para avaliar a qualidade do modelo gerado. Estes testes avaliam a probabilidade do acerto dos pontos de teste ser diferente do acaso em relação à área prevista do modelo gerado. Um modelo pode “acertar” todos os pontos de teste, mas em uma área tão grande que o resultado do teste pode não ser significativo. A recíproca é verdadeira, ou seja, o modelo pode não acertar todos os pontos, mas a área do modelo é tão pequena (em relação

à área de estudo) que a chance de acertar aqueles pontos é significativamente diferente do acaso. Um teste binomial uni-caudal pode ser usado para se determinar se um modelo prediz os pontos de teste significativamente melhor do que o acaso (Anderson *et al.* 2002). Esta medida não deve ser usada diretamente para comparar modelos gerados por diferentes algoritmos porque os mesmos (principalmente algoritmos que geram diferentes tipos de saída, por exemplo, saídas contínuas e categóricas) podem gerar áreas de distribuição muito diferentes, mesmo submetidos ao mesmo tipo de limiar de corte, o que dificulta uma comparação entre eles.

Visando solucionar esse problema surgiu a ideia de usar as curvas características de operação (Fawcett 2003) (ROC), que são amplamente utilizadas na área de controle de qualidade, em processos industriais, na área de saúde, na padronização de doses-resposta, etc. O cálculo da área sob a curva (AUC) fornece uma medida única do desempenho do modelo, independente da escolha prévia de um limite de corte específico, pois a curva é construída a partir de vários limiares (cada limite de corte é responsável por um ponto da curva) e o cálculo da área sob a curva elimina então, a escolha de um limite específico. Este valor (AUC) mede a capacidade discriminatória do modelo, permitindo interpretar seu resultado como a probabilidade de que, ao se sortear dois pontos - um do conjunto de presença e outro do conjunto de ausência - o modelo consiga prever os dois corretamente. Uma das melhores fontes de explicação sobre as curvas características de operação pode ser encontrada no relatório de Fawcett (2003).

A curva ROC é obtida plotando-se a sensibilidade no eixo y e o valor 1-especificidade (ver matriz de confusão, Fig. 2) no eixo x para todos os possíveis limites de corte. A sensibilidade também é conhecida como a taxa de verdadeiros positivos, e representa ausência de erro de omissão. Já a especificidade, ou a variante 1-especificidade, também é conhecida como a taxa de falso positivo, e representa o erro de sobreprevisão. A área abaixo da curva (AUC) é normalmente determinada conectando os pontos com linhas retas e o valor da área é calculado pelo método de trapezóide. Esta área sob a curva (AUC) tem a seguinte interpretação: quanto mais próximo do valor um, ou seja, quanto mais distante o resultado do modelo for da previsão aleatória,

melhor o desempenho do modelo (Rushton *et al.* 2004; Phillips *et al.* 2006).

Com apenas dados de presença, o uso do ROC parece inaplicável, pois sem dados de ausência não há como calcular a especificidade. Este problema é contornado considerando um problema distinto de classificação. Ao invés de tentar avaliar o modelo por sua capacidade de discernir presenças e ausências, sua capacidade de distinguir presença é estimada contra o acaso. Isto é feito sorteando-se, para cada pixel  $x$  na área de estudo, um ponto  $x$ -aleatório e para cada pixel  $x$  dentro da área de distribuição geográfica da espécie é definido um ponto  $x$ -presente. O modelo de distribuição potencial de espécies fará suas previsões para cada pixel correspondente a esses exemplos, mas sem ter informação sobre se este é uma presença ou um ponto escolhido ao acaso na área de estudo (“background” points).

Logo, é possível obter-se previsões tanto para amostras positivas ( $x$ -presenças) quanto negativas ( $x$ -aleatório, pixels selecionados uniformemente ao acaso do “background” da área de modelagem). Juntas, essas informações são suficientes para definir uma curva ROC.

O uso do ROC com dados de presença e ausência difere do cálculo feito apenas com dados de presença em relação ao valor máximo do AUC, que é nesse caso, menor do que 1 (Wiley *et al.* 2003). Se a distribuição da espécie cobre uma taxa  $a$  de pixels, então o máximo que a AUC pode ter é exatamente  $1 - a/2$ . Como não sabemos o valor de  $a$ , não podemos dizer qual o valor ótimo que a AUC pode chegar sob essas circunstâncias (Wiley *et al.* 2003).

Uma área igual a 1 representaria o “modelo perfeito”; uma área de 0,5 indicaria que o modelo seleciona ao acaso. De um ponto de vista prático, um teste de validação pode adotar os valores de AUC a seguir como indicadores da qualidade do modelo (Metz, 1986): excelente (0,9 – 1); bom (0,8 – 0,9); médio (0,7 – 0,8); ruim (0,6 – 0,7); muito ruim (0,5 – 0,6).

Atualmente há uma grande discussão na literatura quanto à aplicação da AUC na modelagem preditiva de distribuição (Lobo *et al.* 2008, Jiménez-Valverde 2011). Nestes trabalhos, os autores apontaram os principais problemas associados ao uso da AUC. Destes, os mais importantes são o fato de a análise ROC dar o mesmo peso aos erros de omissão e sobreprevisão (Peterson *et al.* 2008), e do tamanho da área de estudo em relação à distribuição

conhecida da espécie influenciar grandemente na taxa de ausência corretamente predita pelo modelo. Em relação a esta última situação, é possível diminuir a área de estudo para o mais próximo possível da distribuição conhecida (ou esperada) da espécie, visando minimizar o problema. Outros testes estatísticos têm sido propostos como meios alternativos de avaliar o desempenho de modelos, tais como Kappa (Cohen's Kappa Statistic) e o TSS (True Skill Statistics) (Thuiller *et al.* 2009), no entanto, Jiménez-Valverde (2011) salienta que qualquer método que combine erros de omissão e sobreprevisão em uma única medida estão sujeitos aos mesmos problemas conceituais que a AUC. Sendo assim, os autores recomendam cautela no seu uso, mas até o presente momento, o AUC mostra-se o teste mais amplamente utilizado pelos pesquisadores.

A validação de modelos ainda é uma área de pesquisa ativa e novas técnicas podem surgir tão rápido quanto outras podem se tornar obsoletas. Portanto, é necessário manter-se continuamente informado sobre o assunto, acompanhando o ritmo das publicações mais recentes sobre este tópico.

#### 4.5 Validação de campo

Aplicar o modelo de volta no campo é o teste mais robusto que se pode fazer com os resultados de modelagem, visando avaliar a capacidade do modelo em acertar a distribuição da espécie em áreas de alta (como presença) e baixa (como ausência) adequabilidade ambiental. No entanto, alguns cuidados devem ser tomados quanto à validação de campo. O ideal é planejar o levantamento em diferentes níveis de adequabilidade ambiental para se obter informações mais completas sobre o tamanho da área de ocorrência da espécie.

Outra forma de validação é a avaliação do especialista para a espécie modelada. É o que se chama de avaliação biológica e é feita através de interpretação visual (Phillips *et al.* 2006). Esta avaliação é muito importante, pois é bastante comum se atingir um resultado estatisticamente muito bom, mas que biologicamente não é satisfatório.

É importante ressaltar que os modelos são ferramentas intermediárias em um estudo biogeográfico ou ecológico. O uso de informações adequadas e a interpretação correta dos resultados irão determinar a confiabilidade de um estudo baseado em modelos para estimar as características

do relacionamento entre espécies e seu meio. Por isso o apoio de um especialista no grupo focado, tanto para eventualmente corrigir a base de dados de ocorrência quanto para auxiliar na interpretação dos dados, é fundamental.

#### Procedimentos pós-modelagem

Com o objetivo de aumentar a confiabilidade dos modelos gerados, ou visando diminuir a incerteza de se utilizar modelos gerados por diferentes algoritmos, técnicas de projeção combinada ("ensemble forecasting") podem ser aplicadas aos resultados da modelagem oriunda de diferentes algoritmos e/ou de diferentes cenários (diferentes modelos e cenários climáticos, por exemplo) (ver revisão em Araújo e New 2007). Basicamente, esta técnica consiste em gerar um modelo de consenso onde áreas de alta adequabilidade serão as áreas que todos (ou a maioria) dos modelos indicaram como sendo propícias para a ocorrência da espécie. Diferentes estratégias podem ser aplicadas, como por exemplo, análises de componentes principais (PCA) enfocando diferentes algoritmos ou camadas ambientais (Thuiller 2004; Araújo *et al.* 2006), médias ponderadas dos resultados obtidos através dos valores de acurácia (AUC ou TSS ou Kappa) (Thuiller *et al.* 2009) ou estratégias múltiplas com níveis crescentes de complexidade, desde combinação dos mapas resultantes até análises multidimensionais (Diniz-Filho *et al.* 2009, 2010a,b).

Além disso, na literatura existem algumas sugestões de como melhorar o resultado da modelagem retirando-se do modelo as áreas na qual a espécie não ocorre por fatores históricos (barreiras geográficas) (Anderson *et al.* 2003). Pode-se ainda retirar áreas não habitadas devido a interações bióticas (por exemplo, competição) (Anderson *et al.* 2002). O modelo ainda pode ser ajustado utilizando-se dados de uso da terra para excluir áreas desflorestadas ou alteradas, áreas urbanas, ou ainda áreas de cultivo (Anderson e Martinez-Meyer 2004) caso o objetivo da modelagem esteja mais próximo da distribuição real/actual da espécie.

#### Desafios

Os principais desafios que a modelagem preditiva de distribuição de espécies apresenta estão sintetizados a seguir:

Ainda não existe um método efetivo para avaliar comparativamente a acurácia dos modelos preditivos. Apesar de existirem vários algoritmos disponíveis, nem sempre é possível comparar o resultado de cada um deles e estimar as diferenças de uma forma satisfatória. Assim, outras ferramentas de pós-análise deveriam ser desenvolvidas de tal forma a proporcionar métodos mais acurados para qualificar os modelos obtidos.

A modelagem de distribuição é diretamente dependente da qualidade dos dados de ocorrência das espécies. Coletas mais extensas, padronizadas e bem georreferenciadas são fundamentais, especialmente em regiões pouco amostradas. Apesar da existência de várias iniciativas internacionais ainda existe pouca disponibilidade de dados, pois grande parte da informação ainda não está digitalizada ou disponibilizada em portais da internet. Além disso, muitos dados disponibilizados ainda apresentam erros, seja de identificação ou de digitalização, e necessitariam ser corrigidos por especialistas nos grupos enfocados. Neste contexto, também seria muito importante o desenvolvimento e a implementação de filtros de qualidade de dados, tanto taxonômicos como de mapeamento, para uma garantia de maior qualidade nos dados disponibilizados.

Os dados de interação são bastante importantes no processo de modelagem, porém, poucos dados desse tipo estão à disposição. Assim, o desenvolvimento e disponibilização de dados de interação em formato digital é também uma ação fundamental.

A modelagem de distribuição preditiva precisa ser utilizada como um meio e não um fim em si. O propósito deveria ser mais focado na aquisição de mais informação sobre a ocorrência da espécie para que outras técnicas (mais robustas) possam ser também aplicadas. Especificamente em aplicações de cunho conservacionista, esta técnica passa a ser um importante mecanismo de auxílio especialmente na ausência de uma quantidade maior de dados ou na urgência para a tomada de decisão.

Finalmente, o uso de modelagem preditiva de distribuição deveria ser mais divulgado para estimular os pesquisadores, especialmente taxonomistas e ecólogos, que conhecem de forma mais profunda as relações entre espécie e ambiente. Na medida em que mais pesquisadores se debruçarem sobre o tema, com o uso correto da técnica, seu desenvolvimento poderá ser aprimorado.

## Agradecimentos

Os autores agradecem aos dois revisores anônimos que contribuíram com sugestões importantes para o manuscrito. Agradecem também ao CNPq (proc. 385212/2009-7; 575069/2008-2), à FAPESP (proc. 04/15801-0) e à CAPES (proc. 3030-10-5) pelos apoios concedidos a esta proposta.

## Referências

- Anderson, R.P.; Peterson, A.T. & Gómez-Laverde, M. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* 98: 3-16.
- Anderson, R.P.; Lew, D. & Peterson, A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162: 211-232.
- Anderson, R.P. & Martinez-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.
- Araújo, M.B. & Williams, P. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* 96: 331-345.
- Araújo, M.B. & Guisan, A. 2006. Five (or so) challenges for species distribution modeling. *Journal of Biogeography* 33: 1677-1688.
- Araújo, M.B.; Thuiller, W. & Pearson, R.G. 2006. Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* 33: 1712-1728.
- Araújo, M.B. & Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16: 743-753.
- Araújo, M.B. & New, M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42-47.
- Araújo, M.B.; Nogueira-Bravo, D.; Reginster, I.; Rounsevell, M. & Whittaker, R.J. 2008. Exposure of European biodiversity to changes in human-induced pressures. *Environmental Science & Policy* 11: 38-45.
- Austin, M.P. 1971. Role of regression analysis in plant ecology. *The Proceedings of the Ecological Society of Australia* 6: 63-75.
- Austin, M.P. 1992. Modelling the environmental niche of plants: implications for plant community response to elevated CO<sub>2</sub> levels. *Australian Journal of Botany* 40: 615-30.
- Austin, M.P. 1998. An ecological perspective on biodiversity investigations: example from Australian



- eucalypt forests. *Annals of the Missouri Botanical Garden* 85: 2-17.
- Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101-118.
- Breiman, L. 2001. Random forests. *Machine learning* 45: 5-32.
- Brown, J. H. & Lomolino, M.V. 2006. *Biogeografia*. 2nd ed. Funpec, Ribeirão Preto. 692p.
- Canhos, D.A.L.; Canhos, V.P. & Marino, A. 2010. Polinizadores: desafios e oportunidades no desenvolvimento de sistemas de informação online. *Oecologia Australis* 14: 307-316.
- Cartolano Jr., E.A.; Saraiva, A.M.; Correa, P.L.P.; Giannini, T.C. & Giovanni, R. 2007. Uma proposta de esquema de dados de relacionamento entre espécimes. *In: XXXIII Conferencia Latinoamericana de Informática (CLEI 2007) e Taller Latinoamericano de Informática para la Biodiversidad (INBI 2007)*, San Jose. CD-ROM.
- Chapman, A.D.; Muñoz, M.E.S. & Koch, I. 2005. Environmental information: Placing environmental phenomena in an ecological and environmental context. *Biodiversity Informatics* 2: 24-41.
- Chase J.M. & Leibold, M.A. 2003. *Ecological niches: linking classical and contemporary approaches*. The University of Chicago Press, Chicago.
- Chen, Y.H. 2009. Conservation biogeography of the snake family Colubridae of China. *North-Western Journal of Zoology* 5: 251-262.
- Costa, J.; Peterson, A.T. & Beard, C.B. 2002. Ecologic niche modeling and differentiation of populations of *Triatoma brasiliensis neiva*, 1911, the most important Chagas' disease vector in northeastern Brazil (Hemiptera, Reduviidae, Triatominae). *American Journal of Tropical Medicine and Hygiene* 67: 516-520.
- Diniz-Filho, J.A.F.; Bini, L.M.; Oliveira, G.; Barreto, B.S.; Silva, M.M.F.P.; Terribile, L.C.; Rangel, T.F.L.V.B.; Pinto, M.P.; Sousa, N.P.R.; Vieira, L.C.G.; Melo, A.S.; Marco Jr., P.; Vieira, C.M.; Balmires, D.; Bastos, R.P.; Carvalho, P.; Ferreira, L.G.; Telles, M.P.C.; Rodrigues, F.M.; Silva, D.M.; Silva Jr., N.J. & Soares, T.N. 2009. Macroecologia, biogeografia e áreas prioritárias para conservação no cerrado. *Oecologia Brasiliensis* 13: 470-497.
- Diniz-Filho, J.A.F.; Nabout, J.C.; Telles, M.P.C.; Soares, T.N. & Rangel, T.F.L.V.B. 2009c. A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and Molecular Biology* 32: 203-211.
- Diniz-Filho, J.A.F.; Oliveira, G.; Bini, L.M.; Loyola, R.D.; Nabout, J.C. & Rangel, T.F.L.V.B. 2009d. Conservation biogeography and climate change in Brazilian cerrado. *Natureza & Conservação* 7: 8-18.
- Diniz-Filho, J.; Ferro, V.; Santos, T.; Nabout, J.; Dobrovolski, R. & De Marco Jr., P. 2010a. The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae). *Revista Brasileira de Entomologia* 54: 339-349.
- Diniz-Filho, J.; Nabout, J.; Bini, L.; Loyola, R.; Rangel, T.; Nogueira-Bravo, D. & Araújo, M. 2010b. Ensemble forecasting shifts in climatically suitable areas for *Tropidacris cristata* (Orthoptera: Acridoidea: Romaleidae). *Insect Conservation and Diversity*, 3: 213-221,
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7: 1-26.
- Elith, J.; Graham, C.H.; Anderson, R.P.; Dudik, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; Li, J.; Lohmann, L.G.; Loiselle, B.A.; Manion, G.; Moritz, C.; Nakamura, M.; Nakazawa, Y.; Overton, J.M.; Peterson, A.T.; Phillips, S.J.; Richardson, K.S.; Scachetti-Pereira, R.; Schapire, R.E.; Soberon, J.; Williams, S.; Wisz, M.S. & Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.
- Elith, J. & Leathwick, J. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* 13: 265-275.
- Elith, J. & Graham, C.H. 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. & Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics* 40: 677-97.
- Elton, C. 1927. *Animal ecology*. Sidgwick and Jackson, London. 296p.
- Engler, R.; Guisan, A. & Rechsteiner, L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 41: 263-274.
- Fawcett, T. 2003. ROC graphs: Notes and practical considerations for data mining researchers. HP Labs Tech Report HPL-2003-4.
- Ferrier, S. 1984. The status of the Rufous Scrub-Bird *Atrichornis rufescens*: habitat, geographical variation and abundance. PhD Thesis. University of New England, Armidale. 318p.
- Fielding, A.H. & Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation

- presence/absence models. *Environmental Conservation* 24: 38–49.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474-499.
- Giannini, T.C.; Saraiva, A.M.; Alves dos Santos, I. 2010. Ecological niche modeling and geographical distribution of pollinator and plants: a case study of *Peponapis fervens* (Smith, 1879) (Eucerini: Apidae) and *Cucurbita* species (Cucurbitaceae). *Ecological Informatics* 5: 59-66.
- Giannini, T.C.; Lira-Saade, R.; Ayala, R.; Saraiva, A.M. & Alves dos Santos, I. 2011. Ecological niche similarities of *Peponapis* bees and non-domesticated *Cucurbita* species. *Ecological Modelling* 222: 2011-2018.
- Godsoe, W.; Strand, E.; Smith, C.I.; Yoder, J.B.; Esque, T.C. & Pellmyr, O. 2009. Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytologist* 183: 589-599.
- Graham, C.H.; Ferrier, S.; Huettman, F.; Moritz, C. & Peterson, A.T. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19: 497-503.
- Grinnell, J. 1904. The origin and distribution of the chestnut-backed chickadee. *Auk* 21: 375-377.
- Grinnell, J. 1917. The niche-relationship of the California thrasher. *Auk* 34: 427-433.
- Grinnell, J. 1924. Geography and evolution. *Ecology* 5: 225-229.
- Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009.
- Guisan, A.; Graham, C.H.; Elith, J.; Huettmann, F. & The NCEAS Species Distribution Modelling Group. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13: 332-340.
- Hastie, T.J. & Tibshirani, R. 1990. Generalized additive models. Chapman and Hall, Londres. 335p.
- Heikkinen, R.K.; Luoto, M.; Virkkala, R.; Pearson, R.G. & Körber, J.H. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography* 16: 754-763.
- Hijmans, J. R.; Cameron, S. E.; Parra, J. L.; Jones, P. G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.
- Hirzel, A. & Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* 157: 331-341.
- Hirzel, A.H.; Hausser, J.; Chessel, D. & Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83: 2027-2036.
- Hochachka, W.; Caruana, R.; Fink, D.; Munson, A.; Riedewald, M.; Sorokina, D. & Kelling, S. 2007. Data mining for discovery of pattern and process in ecological systems. *Journal of Wildlife Management* 71: 2427-2437.
- Hortal, J.; Jimenez-Valverde, A.; Gomez, J.F.; Lobo, J.M. & Baselga, A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117: 847-858.
- Hutchinson, G.E. 1944. Limnological studies in Connecticut. Part 7. A critical examination of the supposed relationship between phytoplankton periodicity and chemical changes in lake waters. *Ecology* 35: 3-26.
- Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbour Symposium on Quantitative Biology 22: 415-427.
- Jiménez-Valverde, A.; Gómez, J.F.; Lobo, J.M.; Baselga, A. & Hortal, J. 2008a. Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula. *Annales Zoologici Fennici* 45: 200-210.
- Jiménez-Valverde, A.; Lobo, J.M. & Hortal, J. 2008b. Not as good as they seem: the importance of concepts in species distribution modeling. *Diversity and Distributions* 14: 885-890.
- Jiménez-Valverde, A. 2011. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modeling. *Global Ecology and Biogeography*. doi: 10.1111/j.1466-8238.2011.00683.x.
- Kamino, L.H.Y.; Stehmann, J.R.; Amaral, S.; De Marco, Jr., P.; Rangel, T.F.; Siqueira, M.F.; Giovanni, R. & Hortal, J. 2011. Challenges and perspectives for species distribution modelling in the neotropics. *Biology Letters* doi:10.1098/rsbl.2011.0942.
- Leathwick, J.R. & Austin, M.P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82: 2560-2573.
- Levine, R.S.; Peterson, A.T.; Yorita, K.L.; Carroll, D.; Damon, I.K. & Reynolds, M.G. 2007. Ecological niche and geographic distribution of human monkeypox in Africa. *Plos One* 2: e176.
- Liu, C.; Berry, P.M.; Dawson, T.P. & Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385-393.
- Lobo, J.M. 2008. Database records as a surrogate for sampling effort provide higher species richness

- estimations. *Biodiversity and Conservation* 17: 873-881.
- Lobo, J.M.; 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, J.M. & Tognelli, M.F. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation* 19: 1-7.
- Margules, C.R. & Austin, M.P. 1991. *Nature conservation: cost effective biological survey and data analysis*. CSIRO, Canberra. 207p.
- McCullagh, P. & Nelder, J.A. 1989. *Generalized linear models*. Chapman and Hall, Londres. 511p.
- Mcpherson, J.M.; Jetz, W. & Rogers, D.J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41: 811-823.
- Metz, C.E. 1986. ROC methodology in radiologic imaging. *Investigational Radiology* 21:720-733.
- Muñoz, M.E.S.; Giovanni, R.; Siqueira, M.F.; Sutton, T.; Brewer, P.; Pereira, R.S.; Canhos, D.A.L. & Canhos, V.P. 2011. openModeller: a generic approach to species' potential distribution modeling. *Geoinformatica* 1: 111-135.
- Nix, H.; McMahon, J. & Mackenzie, D. 1977. Potential areas of production and the future of pigeon pea and other grain legumes in Australia. *In*: Wallis, E.S. & Whiteman, P.C. (eds.). *The potential for pigeon pea in Australia*. Proceedings of Pigeon Pea (*Cajanus cajan* (L.) Millsp.). Field Day University of Queensland, Queensland. Pp. 1-12.
- Oberhauser, K. & Peterson, A.T. 2003. Modeling current and future potential wintering distributions of eastern North American monarch butterflies. *Proceedings of the National Academy of Sciences* 100: 14063-14068.
- Olson, D.M.; Dinerstein, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.V.N.; Underwood, E.C.; D'Amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C.; Loucks, C.J.; Allnutt, T.F.; Ricketts, T.H.; Kura, Y.; Lamoreux, J.F.; Wettengel, W.W.; Hedao, P. & Kassem, K.R. 2001. *Terrestrial ecoregions of the world: a new map of life on earth*. *BioScience* 51: 933-938.
- Ortega-Huerta, M.A. & Peterson, A.T. 2004. Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. *Diversity and Distributions* 10: 39-54.
- Pearce, J. L. & Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* 128: 127-147.
- Pearson, R.G.; Thuiller, W.; Araújo, M.B.; Martinez-Meyer, E.; Brotons, L.; McClean, C.; Miles, L.; Segurado, P.; Dawson, T.C. & Lees, D.C. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33: 1704-1711.
- Pearson, R.G.; Raxworthy, C.J.; Nakamura, M. & Peterson, A.T. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102-117.
- Peduzzi, P.; Concato, J.; Kemper, R.; Holford, T.R. & Feinstein, A.R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49: 1373-1379.
- Peterson, A.T.; Ortega-Huerta, M.A.; Bartley, J.; Sánchez-Cordero, V.; Soberón, J.; Buddemeier, R.H. & Stockwell, D.R.B. 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416: 626-629.
- Peterson, A.T.; Papes, M.; Kluza, D.A. 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science* 51: 863-868.
- Peterson, A.T. & Robins, C.R. 2003. Using ecological-niche modeling to predict Barred Owl invasions with implications for Spotted Owl conservation. *Conservation Biology* 17: 1161-1165.
- Peterson, A.T.; Papes, M.; Reynolds, M.G.; Perry, N.D.; Hanson, B.; Regnery, R.L.; Hutson, C.L.; Muizniek, B.; Damon, I.K. & Carroll, D.S. 2006a. Native-range ecology and invasive potential of *Cricetomys* in North America. *Journal of Mammalogy* 87: 427-432.
- Peterson, A.T.; Lash, R.R.; Carroll, D.S. & Johnson, K.M. 2006b. Geographic potential for outbreaks of Marburg hemorrhagic fever. *American Journal of Tropical Medicine and Hygiene* 75: 9-15.
- Peterson, A.T.; Papes, M. & Eaton, M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30: 550-560.
- Peterson, A.T.; Papes, M. & Soberon, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213: 63-72.
- Peterson, A.T.; Barve, N.; Bini, L.M.; Diniz Filho, J.A.F.; Jimenez-Valverde, A.; Lira-Noriega, A.; Lobo, J.; Maher, S.; Marco Jr, P.; Martinez-Meyer, E.; Nakazawa, Y. & Soberon, J. 2009. The climate envelope may not be empty. *PNAS* 106: E47.
- Phillips, S.J.; Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

- Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson *et al.* (2007). *Ecography* 31: 272-278.
- Polasky, S. & Solow, A.R. 2001. The value of information in reserve site selection. *Biodiversity and Conservation* 10: 1051-1058.
- Pulliam, H.R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3: 349-361.
- Rubel, F. & Kotteck, M. 2010. Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift* 19: 135-141.
- Rushton, S.P.; Ormerod, S.J. & Kerby, G. 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology* 41: 193-200.
- Sánchez-Cordero, V. & Martínez-Meyer, E. 2000. Museum specimen data predict crop damage by tropical rodents. *Proceedings of the National Academy of Sciences* 97: 7074-7077.
- Santana, F.S.; Siqueira, M.F.; Saraiva, A.M. & Correa, P.L.P. 2008. A reference business process for ecological niche modeling. *Ecological Informatics* 3: 75-86.
- Saraiva, A.M.; Ruggiero, M.; Adams, L.D.; Sellers, E. & Correa, P.L.P. 2009a. Pollinators information network for the Americas – IABIN-PTN. *In: e-biosphere Conference 09. Conference Abstracts. Londres. Pp. 79-80.*
- Saraiva, A.M.; Cartolano Junior, E.A.; Giovanni, R.; Giannini, T.C. & Correa, P.L.P. 2009b. Exchanging specimen interaction data using Darwin Core. *In: Weitzman, A.L. (ed.). Proceedings of TDWG, Montpellier. CD-ROM.*
- Schoener, T. W. 1989. The ecological niche. *In: Cherrett, J.M. (ed.). Ecological concepts. Blackwell, Oxford. Pp. 79-113.*
- Siqueira, M.F. & Peterson, A.T. 2003. Consequences of global climate change for geographic distributions of cerrado tree species. *Biota Neotropica* 3: 1-14.
- Siqueira, M.F. & Durigan, G. 2007. Modelagem da distribuição geográfica de espécies lenhosas de cerrado no Estado de São Paulo. *Revista Brasileira de Botânica* 30: 239-249.
- Siqueira, M.F.; Durigan, G.; De Marco Jr., P. & Peterson, A.T. 2009. Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *Journal for Nature Conservation* 17: 25-32.
- Soberón, J. & Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics* 2: 1-10.
- Soberón, J. & Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and assumptions. *PNAS* 106: 19644-19650.
- Soberón, J. 2010. Niche and area of distribution modeling: a population ecology perspective. *Ecography* 33: 159-167.
- Stockwell, D.R.B. & Noble, I.R. 1991. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation* 32: 249-254.
- Stockwell, D. & Peters, D. 1999. The GARP modeling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13: 143-158.
- Stockwell, D.R.B. & Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1-13.
- Thomas, C.D.; Cameron, A.; Green, R.E.; Bakkenes, M.; Beaumont, L.J.; Collingham, Y.C.; Erasmus, B.F.N.; Siqueira, M.F.; Grainger, A.; Hannah, L.; Hughes, L.; Huntley, B.; van Jaarsveld, A.S.; Midgley, G.F.; Miles, L.; Ortega-Huerta, M.A.; Peterson, A.T.; Phillips, O.L. & Williams, S.E. 2004. Extinction risk from climate change. *Nature* 427: 145-148.
- Thuiller, W. 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9: 1353-1362.
- Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology* 10: 2020-2027.
- Thuiller, W.; Albert, C.; Araújo, M.B.; Berry, P.M.; Cabeza, M.; Guisan, A.; Hickler, T.; Midgley, G.F.; Paterson, J.; Schurr, F.M.; Sykes, M.T. & Zimmermann, N.E. 2008. Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics* 9: 137-152.
- Thuiller, W.; Lafourcade, L.; Engler, R. & Araújo, M.B. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32: 369-373.
- Vanderwal, J.; Shoo, L.P.; Graham, C. & Williams, S.E. 2009. Selecting pseudoabsence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling* 220: 589-594.
- Verner, J.; Morrison, M.L. & Ralph, C.J. 1986. *Wildlife 2000: modelling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison. 480p.*



- Wiens, J.A.; Stralberg, D.; Jongsomjit, D.; Howell, C.A. & Snyder, M.A. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19729-19736.
- Wiley, E.O.; Mcnyset, K.M.; Peterson, A.T.; Robins, C.R. & Stewart, A.M. 2003. Niche modeling and geographic range predictions in the marine environment using a machine learning algorithm. *Oceanography* 16: 120-127.
- Wisz, M.S.; Hijmans, R.J.; Li, J.; Peterson, A.T.; Graham, C.H.; Guisan, A. & NCEAS. 2008. Predicting species distributions working group. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763-773.