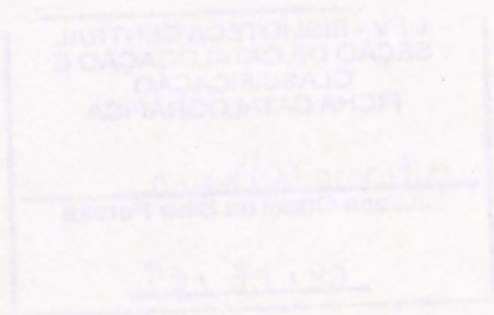


FRANCISCO CANDIDO CARDOSO BARRETO

MODELAGEM DE DISTRIBUIÇÃO POTENCIAL DE ESPÉCIES COMO FERRAMENTA PARA A CONSERVAÇÃO: SELEÇÃO E AVALIAÇÃO DE ALGORITMOS E APLICAÇÃO COM *Heliconius nattereri* FELDER, 1865 (NYMPHALIDAE: HELICONIINAE).

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Entomologia, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2008



Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV

T

B273m
2008

Barreto, Francisco Candido Cardoso, 1976-
Modelagem de distribuição potencial de espécies
como ferramenta para a conservação: seleção e avaliação de
algoritmos e aplicação com *Heliconius natteri* Felder, 1865
(Nymphalidae: Heliconiinae) / Francisco Candido Cardoso
Barreto – Viçosa, MG, 2008.
ix, 57f.: il. ; 29cm.

Orientador: Paulo De Marco Júnior
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. *Heliconius natteri* – Conservação. 2. Espécies em
extinção. 3. Ecologia – Modelos matemáticos
I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 595.789

FRANSCISCO CANDIDO CARDOSO BARRETO

**MODELAGEM DE DISTRIBUIÇÃO POTENCIAL DE ESPÉCIES
COMOFERRAMENTA PARA A CONSERVAÇÃO: SELEÇÃO E AVALIAÇÃO
DE ALGORITMOS E APLICAÇÃO COM *HELICONIUS NATTERERI* FELDER,
1865 (NYMPHALIDAE: HELICONIINAE)**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Entomologia, para obtenção do título de *Doctor Scientiae*.

APROVADA: 12 de Dezembro de 2008.

Prof. Dr. Raul Narciso Carvalho
Guedes

Prof^a. Dra Tânia Gonçalves dos
Santos

Prof. Dr. Rômulo Ribon

Prof^a. Dra Marinez Ferreira de Siqueira
(co-orientadora)

Prof. Dr. Paulo de Marco Júnior
(Orientador)

“The spice must flow...”
Frank Herbert - Duna

AGRADECIMENTOS

Foram quatro longos anos para que essa tese se concretizasse. Muita coisa aconteceu. Minha mudança para Viçosa, os anos que passei no extinto Laboratório de Ecologia Quantitativa desta instituição, os amigos que fiz, as disciplinas que cursei, as experiências profissionais a que tive acesso. As questões pessoais que me fizeram mudar e crescer.

Agradeço à todos os amigos que fiz nesse período, alguns dos quais me acompanham até hoje. Agradeço à Prof. Dra. Flávia Coelho, pelo apoio e amizade. À Rubana Palhares pelo carinho sempre presente, Daniel (Jesus) por dividir suas angústias e alegrias comigo. Ao pessoal de Nova Xavantina, Leandro Juen pela pessoa excepcional que é, Joana Darc Batista pela amizade e longas conversas na hora do lanche da tarde. Vocês dois são um exemplo do que o esforço pessoal pode alcançar. Agradeço à Mirian Cristina de Almeida, por todas as coisas boas que você transmite às pessoas a sua volta (além do fato de sempre rir das minhas piadas). Não posso deixar de agradecer pela amizade e alegria espontânea de Letícia de Almeida. Longe, mas nunca distante. Ao longo desses anos participei de importantes projetos, como por exemplo o Projeto Leptagrion, sob coordenação e execução da Dra. Karina Schmidt Furieri. Muito obrigado por tudo.

Jamais poderia deixar de agradecer uma pessoa muito importante, minha primeira orientada de graduação e grande amiga, Natália Caixeta Barroso. Sem sua amizade sincera e apoio eu não estaria aqui, e você sabe disso. Devo agradecer à todos que conheci por seu intermédio, sua família que me recebeu com carinho quando eu precisei. Às moradoras de sua antiga república, em especial à Gabriela pela alegria que sempre pude desfrutar com sua companhia.

Me mudei em 2007 para a cidade de Ouro Preto, onde fui professor substituto por um ano na UFOP, experiência essencial para minha formação acadêmica e profissional. Lá contei com o suporte e amizade do Prof. Dr. Rômulo Ribon, que rapidamente se tornou um exemplo de hombridade a ser seguido. Obrigado por todas nossas conversas, piadas, ditos populares e kibes à meia noite na Assufop. Saiba que adoto muito de seus conselhos e orientações em minha vida.

Ainda em Ouro Preto conheci minha noiva e futura esposa, Tamyris Cuppari. Desde então minha vida tomou outro tom. É contigo que viverei minha vida, e à você tenho muito a agradecer. Obrigado por todas as coisas boas que você me trouxe, por me re-erguer quando eu mais precisava, por me ajudar em tudo e se fazer essencial no meu dia a dia. É pelo nosso futuro e pela família que contruiremos juntos que eu luto e sempre lutarei.

Em 2008 me mudei para a cidade de Ipatinga. Enfrentei muitas dificuldades para conseguir conciliar minha obrigação como doutorando e minhas obrigações como professor. Mesmo com a orientação de desistir e me dedicar somente à tese – conselho dado pelo meu orientador Prof. Dr. Paulo De Marco Jr. Insisti e se há algo que possa desabonar meu trabalho isto é de inteira responsabilidade minha por não acatá-lo.

Ainda quanto ao meu orientador, o quê posso dizer? Tivemos altos e baixos, discutimos, brigamos, conversamos, nos entendemos. Nunca poderei listar quantas coisas boas e positivas você me trouxe, em todos os aspectos. Sou-lhe imensamente grato por tudo. Todas as oportunidades, todo o incentivo, todo o estímulo que recebi de você foram para me fazer crescer como pessoa e como profissional. Você é o maior exemplo que posso seguir.

Agradeço à outra pessoa muito importante em minha formação profissional, a Dra. Marinez Ferreira de Siqueira, minha co-orientadora e amiga. Aprendi muito em sua companhia e os cursos que damos juntos têm me ajudado a entender mais e mais o que fazemos. Ainda com relação à você, devo estender meus agradecimentos por todo o apoio que o Centro de Referência em Informação Ambiental – CRIA – me proporcionou. Cresci muito com todos vocês. Agradeço também ao IPÊ, Instituto de Pesquisas Ecológicas, por acreditar em mim e oferecer um espaço de qualidade para todas as atividades que realizamos.

Um outro grupo de pessoas deve receber minha atenção: minha família. Minha Mãe, meu Pai e meus irmãos, Alice e Pedro Barreto. Tudo o que faço penso em vocês. Tudo que luto para obter uma posição na vida é para poder oferecer-lhes meu apoio incondicional. Eu amo todos vocês e lamento por não dizer isso mais vezes. Saibam que levo vocês comigo para onde quer que eu vá.

SUMÁRIO

RESUMO.....	vi
ABSTRACT.....	viii
CAPÍTULO UM: Modelagem de Distribuição Potencial de Espécies: Uma comparação entre GARP, MAXENT e SVM.....	1
RESUMO.....	2
ABSTRACT	3
INTRODUÇÃO	4
METODOLOGIA.....	6
RESULTADOS	11
DISCUSSÃO	13
REFERÊNCIAS BIBLIOGRÁFICAS	16
TABELAS	19
FIGURAS	27
CAPÍTULO DOIS: Estratégias para conservação de espécies raras com uso de modelos de distribuição potencial de espécies: um estudo de caso com <i>Heliconius nattereri</i> Felder, 1865 (Nymphalidae: Heliconiinae).....	32
RESUMO.....	33
ABSTRACT	34
INTRODUÇÃO	35
METODOLOGIA.....	37
RESULTADOS	41
DISCUSSÃO	42
REFERÊNCIAS	46
FIGURAS	51
TABELAS	54

RESUMO

BARRETO, Francisco Candido Cardoso, D.Sc., Universidade Federal de Viçosa, Dezembro de 2008. **Modelagem de Distribuição Potencial de Espécies como Ferramenta para a Conservação: Seleção e Avaliação de Algoritmos e aplicação com *Heliconius nattereri* Felder, 1865 (Nymphalidae: Heliconiinae).** Orientador: Paulo De Marco Júnior. Co-orientadores: Marinez Ferreira de Siqueira e Lúcio Antônio de Oliveira Campos.

Esta tese está dividida em dois capítulos. O primeiro capítulo apresenta uma abordagem mais técnica onde, com auxílio de uma espécie criada artificialmente, testei três algoritmos para modelagem da distribuição geográfica potencial de espécies: o MAXENT, GARP (best subsets-OM) e o SVM. O objetivo do primeiro capítulo foi testar qual desses três algoritmos apresentava melhor performance ao tentar prever a distribuição real da espécie artificial. Como a performance ou avaliação dos modelos pode ser calculada através de mais de uma técnica, testei qual dessas métricas apresenta resultados mais próximos da realidade do modelo com base na taxa de acertos e erros de cada um calculadas diretamente pela comparação das previsões de cada modelo em relação à área de ocorrência verdadeira da espécie artificial. Além disso, para o cálculo dessas métricas, é necessário a definição de um ponto de corte que divide o modelo em dois aspectos: 1- Acima desse ponto de corte os erros de omissão e sobreprevisão estão equilibrados e 2- abaixo desse ponto as previsões contêm uma taxa de erro maior que a aceitável. Logo, devido à importância dessa definição, utilizei duas abordagens presentes na literatura da área: uma delas é o ponto de corte definido pela construção de uma curva ROC e a outra é o valor mínimo atribuído pelo modelo para adequabilidade ambiental a um ponto de presença verdadeira. Os resultados obtidos indicam que essa definição é muito importante para avaliar corretamente cada modelo. Com o ponto de corte ROC, o algoritmo que obteve maior taxa de acertos foi o GARP. Com o ponto de corte mínimo, isso mudou e o algoritmo de Máxima Entropia (MAXENT) apresentou melhor performance. O segundo capítulo usa o conhecimento gerado pelo primeiro capítulo e adota o MAXENT para modelar a distribuição geográfica potencial da borboleta *Heliconius nattereri* Felder, 1865 classificada como vulnerável pela lista de espécies ameaçadas de extinção publicada pelo IBAMA em 2003. Essa espécie tem registros de ocorrência nos

estados de Minas Gerais, Espírito Santo e Bahia, e apresenta a característica de ser monofágica em relação aos recursos utilizados por suas lagartas – a espécie pertencente à família Passifloraceae conhecida como *Tetrastylis ovalis* (Vell.) Killip. Essa espécie de maracujá apresenta baixo crescimento vegetativo e é polinizada por morcegos, ocorrendo em uma distribuição relativamente restrita no Brasil, e em baixa abundância local. Devido à essa estreita relação de dependência da borboleta ao seu recurso larval, a distribuição geográfica potencial de *Tetrastylis ovalis* foi utilizada como referência para corte (refinamento) da distribuição potencial de *H. nattereri*. Com os resultados obtidos, que consistem em uma previsão do grau de adequabilidade ambiental para ocorrência da espécie para o Brasil, pode cruzar esse dado com a rede de unidades de conservação em nosso país, listando dessa forma todas as unidades de conservação que apresentam uma chance de abrigar novas populações de *Heliconius nattereri*. Discuto ainda algumas ações que podem ser tomadas para que essa espécie saia do nível de risco em que se encontra, aumentando suas populações nos locais onde possui registros de ocorrência e no encontro de novas populações em diferentes locais. São elas: (1) Cultivo e manejo de sua planta hospedeira *Tetrastylis ovalis*, aumentando assim sua abundância e disponibilidade nas áreas de proteção; (2) Estudos da dinâmica populacional da espécie; (3) Estudos da variabilidade genética da espécie; (4) Translocação de indivíduos, se comprovada necessidade no item (3); (5) Criação de *H. nattereri*, em borboletários sediados na própria unidade de conservação, com uso das informações básicas obtidas nos itens (2) e (3).

ABSTRACT

BARRETO, Francisco Candido Cardoso, D.Sc., Universidade Federal de Viçosa, December, 2008. Species **Potential Distribution Modelling as a tool to conservation: Algorithms' Selection and Evaluation and application with *Heliconius nattereri* Felder, 1865 (Nymphalidae: Heliconiinae)**. Advisor: Paulo De Marco Júnior. Co-advisors: Marinez Ferreira de Siqueira and Lúcio Antônio de Oliveira Campos.

This thesis is divided into two chapters. The first one presents a technical approach in which with an artificial created species, I've tested three different algorithms used to predict the species potential geographic distribution: MAXENT, GARP and SVM. The first chapter's objective was to test the performance of these three algorithms and determine which one is the best to predict the real distribution of the artificial species. Because the performance of the models can be evaluated by many techniques, I've tested which metric presents the most real results based on the errors and successes rates, calculated directly by the comparison of each model prediction in relation to the real occurrence area of the artificial species. Besides, use these metrics, it is necessary the definition of a threshold that divides the model in two aspects: 1- Above this threshold, the omission and commission errors are balanced; and 2- Below this threshold the models have an error rate bigger than the acceptable. Knowing the importance of this definition, I used two different approaches find in the literature: the threshold defined with the construction of a ROC curve and the minimum value of environmental suitability attributed, by the model, to a real presence point. The results indicate that this definition is very important to evaluate correctly each model. When the ROC threshold was used, the algorithm with the best performance was GARP. When the Minimum threshold was used, this situation changed and MAXENT presented best performance. The second chapter uses the knowledge generated by the first chapter and adopts MAXENT to model the potential geographic distribution of the butterfly *Heliconius nattereri* Felder, 1865, classified as vulnerable in the IBAMA endangered species list, published in 2003. This species has occurrence registries in Minas Gerais, Espírito Santo e Bahia, and presents the characteristics of being monophagic about the resource of its larvae – the species of the Passifloraceae family known as *Tetrastylis ovalis* (Vell.) Killip. This passion-fruit species has low vegetative growing and is pollinated by bats,

occurring in a relative restrict distribution in Brazil, and with local low abundance. Because of this strict dependence relation of the butterfly with its larval resource, the potential geographic distribution of *Tetrastylis ovalis* was used as a reference to the refining of the potential distribution of *H. nattereri*. With the results, the prediction of the environmental suitability degree to the species occurrence in Brazil, I was able to cross this data with the conservation units' information, making a list of the conservation units that have a chance of being occupied by unknowing populations of *H. nattereri*. I also discuss some actions that can be made in order to reduce the risk level of this species, by increasing its registry local populations and finding new ones in different places, such as: (1) Growth and management of *H. nattereri*'s host plant, *T. ovalis*, increasing its abundance and availability in the protection areas; (2) Population dynamics studies of the species; (3) Genetic variability studies of the species; (4) Individuals translocation, if the necessity is determined in item (3); (5) *H. nattereri* captive creation in Butterfly Houses in the own conservation unit, with the basic information obtained in (2) and (3) items.

CAPÍTULO UM

Modelagem de Distribuição Potencial de Espécies: Uma comparação entre GARP, MAXENT e SVM.

RESUMO

BARRETO, Francisco Candido Cardoso, D.Sc., Universidade Federal de Viçosa, Dezembro de 2008. **Modelagem de Distribuição Potencial de Espécies: Uma comparação entre GARP, MAXENT e SVM.** Orientador: Paulo De Marco Júnior. Co-orientadores: Marinez Ferreira de Siqueira e Lúcio Antônio de Oliveira Campos.

Os modelos de distribuição potencial de espécies são importantes ferramentas para inferir a distribuição geográfica total de uma espécie de interesse – notadamente espécies raras ou ameaçadas de extinção – prever os efeitos das mudanças climáticas sobre a biota ou mesmo a área adequada a uma espécie invasora, antes que ela se disperse no ambiente. Existe uma grande variedade de técnicas e modelos, o que faz com que o pesquisador que tem um problema a solucionar nem sempre encontre facilidade na escolha do algoritmo mais adequado aos objetivos do trabalho e do melhor método de avaliação da qualidade dos modelos gerados pelo mesmo. A clareza dos objetivos do estudo é indispensável, já que estes servirão de guias na escolha de um algoritmo em detrimento de outro, pesando o efeito e as consequências dessa escolha no modelo de distribuição potencial final. Neste trabalho criei uma espécie artificial, nomeada informalmente de *Cyberconius cupparii*. Meu objetivo ao criá-la foi ter em mãos algo que a maioria de nós não possui: a distribuição geográfica completa de uma espécie e as variáveis ambientais mais importantes para a sua ocorrência, tornando possível uma avaliação criteriosa dos algoritmos mais utilizados atualmente. Os algoritmos escolhidos neste trabalho foram: Maxent – baseado na teoria da máxima entropia; GARP, baseado nos algoritmos genéticos; e o SVM, baseado em redes neurais. Dentre os três, o pior desempenho foi o do algoritmo SVM, cuja implementação precisa ser revisada. Os algoritmos MAXENT e GARP obtiveram ótimos resultados nas análises, embora o primeiro tenha tido maior destaque.

ABSTRACT

BARRETO, Francisco Candido Cardoso, D.Sc., Universidade Federal de Viçosa, December, 2008. **Species Potential Distribution Modelling: A comparison between GARP, MAXENT and SVM.** Advisor: Paulo De Marco Júnior. Co-advisors: Marinez Ferreira de Siqueira and Lúcio Antônio de Oliveira Campos.

The species distribution models are important tools to infer the total geographic distribution (sometimes called fundamental niche) of an interest species – specially rare or endangered species – to predict the climate changes' effects on the biota or the adequate area to exotic species, before their dispersion in the environment. There are lots of models and techniques, what makes the researcher that has a problem to solve become with difficulties about the choice of the algorithm that better fits the work objectives and the best evaluation method of the model's quality. The clarity about the objectives is indispensable, because these must guide the choice of one algorithm instead another, considering the effects and consequences of this choice in the final distribution model. In this work, I created an artificial species, informally named *Cyberconius cupparii*. My objective creating was to hold something that most of researchers don't have: the complete geographic distribution of a species and the environmental characteristics that are important to its occurrence, bringing the possibility of a criterional evaluation of the most used algorithms. The chosen algorithms in this work were: Maxent – based on Maximum Entropy Theory; GARP, based on genetic algorithms; and SVM, based on neural networks. SVM obtained the worst performance, and its implementation should be reviewed. Maxent and GARP had great results on the analysis but the first obtained the best ones.

INTRODUÇÃO

O conhecimento da distribuição geográfica das espécies, e dos fatores que limitam essa distribuição têm sido foco de atenção de duas grandes áreas, a biogeografia e a biologia da conservação (Brown & Gibson 1983; Brown *et al.* 1996; Grehan 1992; MacArthur & Wilson 1967). Nem sempre podemos separar essas duas grandes vertentes do conhecimento ecológico e pesquisadores envolvidos com a conservação das espécies em diferentes ambientes, normalmente trabalham não só com a biogeografia, mas também com disciplinas variadas, como evolução, ecologia de paisagens, filogenética, genética, e até fisiologia comparada. Por meio da divulgação dos trabalhos destes pesquisadores, obtemos informações sobre espécies de interesse, seja relacionado à biologia, ecologia ou fisiologia das mesmas.

Quando essas informações envolvem limites geográficos, normalmente, temos mais detalhes ao tratarmos de grupos de espécies abundantes, geograficamente amplas, de interesse econômico (Arim *et al.* 2006; Baer & Schmid-Hempel 2000; Drake & Bossenbroek 2004) ou relacionadas à saúde pública (Ferreira *et al.* 2001; Gomes *et al.* 1995). Infelizmente, o mesmo não pode ser dito quando tratamos de espécies ameaçadas ou raras (Gaston 1994; Rabinowitz 1981) que geralmente possuem distribuição mais restrita. Além disso, quando trabalhamos com essas espécies, temos que enfrentar as poucas informações disponíveis a respeito de sua biologia, fisiologia e ecologia – pelo menos em países megadiversos como o Brasil. Essa lacuna de informações dificulta a aplicação de ações conservacionistas pois estas ficam limitadas à conservação da espécie de interesse apenas em áreas onde ela já tenha sido registrada. Como a área de distribuição real da espécie de interesse pode estar subestimada devido ao baixo esforço de coleta; expedições ou projetos com objetivo específico de procurar aleatoriamente novas populações possuem baixa taxa de sucesso.

Os modelos de distribuição potencial de espécies são ferramentas que surgiram com a proposta de preencher as lacunas de conhecimento sobre os limites geográficos de espécies de interesse, e ainda ajudam na formulação de novas hipóteses sobre os mecanismos que determinam a distribuição destas espécies (Guisan & Zimmermann 2000). Esses modelos diferem entre si em

relação à sua natureza matemática ou família estatística, mas apresentam um ponto em comum: utilizando pontos de ocorrência, conseguem, com diversas formas de intercruzamento com dados ambientais, prever áreas geográficas com maior grau de adequabilidade ambiental à presença da mesma. Outro ponto em comum aos modelos é a tentativa de representar parcialmente o nicho fundamental da espécie, excluindo desse as interações, fatores históricos, entre outros (Soberon & Peterson 2005). Sendo assim, trabalhamos, na maioria das vezes, mais próximos ao conceito de nicho Grinneliano, que não inclui interações ou recursos, e abrange variáveis, o conjunto de componentes abióticos de ampla escala, que permitem a existência das espécies (Soberon 2007).

Atualmente, muitos trabalhos na área se concentram em questões teóricas, principalmente ligadas à avaliação ou validação desses modelos (Guisan *et al.* 2007a; Guisan *et al.* 2007b) e ao consultarmos com atenção a literatura disponível, vemos ciclos interessantes de utilização ou aplicação de tais modelos com espécies reais (Anderson *et al.* 2002; Anderson & Martinez-Meyer 2004; Costa *et al.* 2002), seguidos por ciclos de artigos sobre a validade dos mesmos (Hirzel *et al.* 2006; Guisan *et al.* 2007b; Graham & Hijmans 2006).

A etapa de avaliação e validação é indispensável na construção dos modelos de adequabilidade ambiental, pois sem elas corremos o risco de aceitar como “verdade” modelos que apresentam erros grosseiros. (Tsoar *et al.* 2007; Allouche *et al.* 2006). A consulta a um especialista na espécie de estudo é também uma etapa importantíssima de avaliação. Ao analisar os modelos gerados, esse profissional pode avaliar se a área prevista condiz com seus conhecimentos.

Para o biólogo conservacionista que tem o primeiro contato com o universo dos modelos, a tarefa de escolher o mais adequado para seu trabalho não é, em definitivo, uma questão simples. São muitos algoritmos, cada um com vantagens e desvantagens particulares, muitas vezes ainda não bem exploradas pelos pesquisadores.

Após a escolha do algoritmo adequado e do desenvolvimento dos modelos, a escolha da técnica de avaliação da confiabilidade dos modelos gerados também se torna uma tarefa árdua. Há uma enorme variedade de métodos de avaliação de modelos de distribuição potencial de espécies mas

ainda existem inúmeras questões importantes não-solucionadas que exigem atenção ou maiores esforços na área de pesquisa. Além disso, alguns pontos dificultam bastante a avaliação dos modelos. Mesmo em situações onde se conheça razoavelmente a distribuição geográfica de uma espécie ou de grupos de espécies (Peterson *et al.* 2002; Peterson *et al.* 2000), não se sabe ao certo as variáveis ambientais que realmente são importantes e/ou limitantes, direta ou indiretamente, à essa distribuição, como por exemplo quando relacionadas aos recursos alimentares.

Por esse motivo, a comparação e avaliação dos modelos gerados com espécies reais pode não ser a melhor estratégia, e o uso de espécies artificiais se torna uma alternativa interessante já que com o grupo de dados criados detemos o conhecimento de sua distribuição geográfica completa, tornando possível avaliar os modelos de modo direto e não-subjetivo (Austin *et al.* 2006; Meynard & Quinn 2007; Hirzel *et al.* 2001).

Neste trabalho, avaliei três algoritmos utilizados no cálculo de distribuição geográfica potencial (MAXENT, GARP e SVM), em relação às suas taxas de erros e acertos e às métricas mais utilizadas para a mensuração da performance de modelos. Além disso, de posse da distribuição real da espécie, também foi possível avaliar os próprios métodos de avaliação mais utilizados, comparando seus resultados com as taxas de erro obtidas. Com isso, discuti alguns aspectos importantes que um pesquisador deve ter em mente no processo de escolha destes algoritmos, assim como das métricas de avaliação dos mesmos, com ênfase nos objetivos pretendidos: a conservação de espécies raras ou ameaçadas de extinção, programas de reintrodução de espécies, previsão de invasão de espécies exóticas, efeito das mudanças climáticas sobre a fauna, etc.

METODOLOGIA

Espécie artificial

Para criação da espécie artificial, nomeada *Cyberconius cupparii*, utilizei três camadas ambientais: o modelo digital de elevação (disponível através do *United States Geological Survey (USGS)*, no endereço <http://eros.usgs.gov/>); a sazonalidade da temperatura (desvio padrão *100) e a sazonalidade da

precipitação (coeficiente de variação). Todas as variáveis foram utilizadas com resolução de 30 segundos de grau (resultando em pixels de cerca de 1 km²).

Os limites da distribuição geográfica de *C. cupparii* foram calculados com base na média dessas variáveis no espaço ambiental: seu 'limite de tolerância' superior é a média somada a seu desvio padrão e o limite inferior é a média menos seu desvio padrão. Fazendo esses cálculos, estabeleci, para cada variável, um intervalo de valores no qual a espécie poderia ocorrer (Tabela 1). Com isso, foi possível delimitar três grandes áreas no Brasil simultaneamente adequadas aos limites de todas as variáveis: uma a norte, uma a nordeste e uma a noroeste do país (Figura 1).

Os pontos de ocorrência para gerar os modelos foram obtidos com uso da extensão *Random Point Generator* 1.3 (de autoria do Dr. Jeff Jenness, livre para download no endereço http://www.jennessent.com/arcview/random_points.htm) no ambiente SIG ArcView 3.2. Dentro da área já delimitada para a ocorrência de *C. cupparii*, foram criados 10.000 pontos de ocorrência (figura 2). Com auxílio de uma planilha eletrônica, sorteei aleatoriamente e com reposição, 10 conjuntos de 75 pontos (Ver figura 3 para um exemplo). Cada um desses conjuntos foi utilizado na geração dos modelos em todos os três algoritmos.

Variáveis ambientais

As variáveis ambientais utilizadas fazem parte do conjunto de variáveis bioclimáticas disponíveis no Worldclim (www.worldclim.org). São, no total, 19 variáveis, compostas pela temperatura e precipitação e uma combinação destas, com extremos máximos e mínimos, sazonalidade, variação anual, diária e médias mensais. Além dessas variáveis, utilizei também outras duas com informações geográficas: o modelo de elevação digital e a inclinação do terreno (Tabela 2).

Todas as variáveis para geração dos modelos possuem resolução de 2,5 minutos de grau (pixels com aproximadamente de 5 km de aresta).

Algoritmos utilizados

MAXENT

O software homônimo Maxent (Phillips *et al.*, 2004; Phillips *et al.*, 2006), baseia-se no princípio da máxima entropia, que diz que a melhor aproximação para uma distribuição de probabilidades desconhecida é aquela que satisfaça qualquer restrição à distribuição. Trata-se de um método para realizar previsões ou inferências a partir de informações incompletas. É aplicado em áreas como astronomia, reconstrução de imagens e processamento de sinais. Ele estima a probabilidade de ocorrência da espécie encontrando a distribuição de probabilidade da máxima entropia (que é a distribuição mais próxima da distribuição uniforme), submetidas a um conjunto de restrições que representam a informação incompleta sobre a distribuição alvo (Phillips *et al.*, 2004; Phillips *et al.*, 2006; Dudik *et al.*, 2004).

SUPPORT VECTOR MACHINES – SVM

Este algoritmo (Máquina de Vetores de Suporte) caracteriza-se por conter um conjunto de métodos de aprendizagem supervisionados relacionados que pertencem a família dos classificadores lineares generalizados (Vapnik, 2005). Os Vetores de Suporte são capazes de tratar amostras de tamanho da ordem de 100.000 instâncias, criando hiperplanos de separação que atuam como classificadores dos dados, gerando probabilidades de presença e ausência (Cristianini & Shawe-Taylor, 2000).

GARP

O GARP, Algoritmo Genético para Predição por Conjunto de Regras (Genetic Algorithm for Rule-set Prediction), inclui diversos algoritmos distintos (por exemplo, BIOCLIM, regressão logística, etc.) para gerar um conjunto de regras candidatas que expliquem a distribuição de uma espécie. Numa abordagem de inteligência artificial interativa ele opera sobre um conjunto de regras gerado, realizando uma “seleção natural”, excluindo regras menos eficientes e criando novos conjuntos de regras a partir das regras “sobreviventes” (Stockwell and Peters, 1999).

Métodos de avaliação

Os métodos de avaliação mais comuns são derivado da chamada matriz de confusão: matriz de acertos e erros associados à previsão dos modelos (Tabela 3). Os itens “a” e “d” são os verdadeiros positivos e verdadeiros negativos, ou seja, as presenças e ausências preditas corretamente. Os possíveis erros dos modelos são os falsos positivos e falsos negativos, itens “b” e “c”, respectivamente (Sargent 1998; Fawcett 2004; Meynard & Quinn 2007).

Os métodos que adotei para avaliar os modelos gerados são: Área sob a curva (AUC), obtida através da curva ROC, acurácia, sensibilidade, especificidade, Kappa e TSS, os cinco últimos derivados diretamente da matriz de confusão (tabela 3).

As fórmulas, segundo Allouche et al (2006), seguem abaixo:

Acurácia:

$$\frac{A + D}{A + B + C + D}$$

Sensibilidade:

$$\frac{A}{A + C}$$

Especificidade:

$$\frac{D}{B + D}$$

True Skill Statistic (TSS):

$$(\text{sensibilidade} + \text{especificidade}) - 1$$

Kappa:

$$\frac{\left(\frac{A + D}{A + B + C + D}\right) - \frac{(A + B)(A + C) + (C + D)(D + B)}{n^2}}{1 - \frac{(A + B)(A + C) + (C + D)(D + B)}{n^2}}$$

A acurácia é a razão entre a soma de verdadeiros positivos e verdadeiros negativos (A+D), e a área total utilizada na criação dos modelos (no caso, em pixels). Ela representa o grau de acerto do modelo, e serve como avaliação da verossimilhança do mesmo.

Podemos considerar a sensibilidade como uma medida que descreve a probabilidade de um pixel de ocorrência verdadeira ser corretamente classificado como uma ocorrência. A especificidade é a probabilidade de um pixel onde a espécie não está presente ser corretamente classificado como uma ausência (Segurado & Araujo 2004; Fielding & Bell 1997).

A AUC é a área sob uma curva ROC. Essa curva é obtida quando representamos graficamente a sensibilidade no eixo y e 1- especificidade no eixo x (Fielding & Bell 1997). Ela é um indicador direto da capacidade discriminatória do modelo e é interpretada diretamente como a probabilidade do modelo de, dado um ponto de verdadeira presença e um ponto de verdadeira ausência, o modelo classificá-los corretamente (Phillips *et al.* 2006).

A estatística Kappa também é obtida através da análise da matriz de confusão. Este índice é interessante pois mede o desempenho de modelos de distribuição de espécies fazendo uso de todas as informações contidas na matriz de confusão (Fielding e Bell, 1997) e indica o grau de concordância do modelo com a realidade. Infelizmente Kappa é sabidamente sensível à prevalência, ou seja, o número de pontos de presença por área.

O True Skill Statistics, ou TSS, foi desenvolvido por Omri Allouche, Asaf Tsoar e Ronen Kadmon (2006). Sua proposta foi criar um método de avaliação de modelos que não sofresse influência da prevalência e do número de pontos utilizados para gerar os modelos analisados, mantendo ainda as vantagens do método anterior.

Pontos de Corte

Todos os métodos de avaliação discutidos acima dependem da definição de um ponto de corte (exceto a AUC, derivada diretamente da curva ROC). Essa escolha pode ser feita de diversas maneiras, de acordo com os objetivos do trabalho. Atualmente, há uma tendência ao uso do Limite de decisão ROC, que atribui a mesma importância aos erros de sensibilidade e especificidade. Outra vertente utiliza o menor valor de adequabilidade ambiental atribuída por um modelo a um ponto de presença verdadeiro, o Limite de decisão Mínimo (Pearson *et al.* 2006; Phillips *et al.* 2006). Neste trabalho, adotamos ambas as

técnicas para verificar o efeito da adoção de um ou outro limite de decisão sobre as métricas de avaliação.

Os modelos foram gerados utilizando as dez partições de dados de ocorrência de *C. cupparii*, obtive, assim, dez modelos para cada algoritmo. Esses modelos foram sobrepostos, em ambiente SIG, com a área real da espécie sendo possível quantificar seus erros e acertos. Com essas informações, construí uma matriz de confusão (tabela 3), com a qual foi possível calcular os valores dos métodos de avaliação citados acima. Com a distribuição real da espécie em mãos, foi possível também calcular as taxas de erro e acerto dos modelos. Esses valores foram utilizados para avaliar as próprias técnicas de avaliação.

Para analisar as diferenças entre os valores obtidos pelos três modelos, utilizei o teste não-paramétrico de Friedman (ZAR, 1998), pois nenhuma das variáveis analisadas apresentam distribuição normal.

RESULTADOS

Resultados independentes do Limite de decisão adotado

O algoritmo que apresentou a maior AUC média foi o MAXENT, em seguida o GARP e por último o SVM (Friedman ANOVA $X^2(10,2)=12,7$; $p=0,33$) (Figura 4).

Resultados obtidos com o Limite de decisão ROC

O algoritmo que apresentou maior especificidade foi o GARP, seguido pelo SVM e MAXENT (Friedman ANOVA $X^2(10,2)=12,6$; $p=0,001$).

O mais sensível dentre os três algoritmos foi o GARP. O MAXENT foi o segundo e por último o SVM (Friedman ANOVA $X^2(10,2)=7,78$; $p=0,02$).

Os modelos mais acurados foram os produzidos pelo algoritmo GARP seguido pelo SVM e por último o MAXENT (Friedman ANOVA $X^2(10,2)=12,6$; $p=0,001$).

Os maiores valores de estatística kappa foram os do algoritmo GARP, seguido pelo SVM e por último o MAXENT (Friedman ANOVA $X^2(10,2)=12,6$; $p=0,001$).

O maior valor médio de TSS foi obtido no algoritmo GARP. O MAXENT apresentou o segundo maior TSS e o menor valor foi o do algoritmo SVM (Friedman ANOVA $X^2(10,2)=16,8$; $p=0,0002$).

As informações acima estão representadas na tabela 4.

Quanto aos erros e acertos em geral, o algoritmo GARP apresentou as maiores taxas de acerto e menores taxas de erro dentre os três algoritmos. O segundo colocado foi o SVM, seguido pelo MAXENT. Esses valores encontram-se na tabela 5.

Resultados Obtidos com o Limite de decisão Mínimo

Com relação à especificidade, o algoritmo que apresentou os maiores valores foi o MAXENT, enquanto que o SVM e o GARP não diferiram entre si (Friedman ANOVA $X^2(10,2)=12,1$; $p=0,004$).

Dentre os três algoritmos, o que apresentou menores valores de sensibilidade foi o SVM. O MAXENT e o GARP apresentaram valores mais altos e semelhantes entre si (Friedman ANOVA $X^2(10,2)=8,2$; $p=0,001$).

O mais acurado dos algoritmos foi o MAXENT, seguido pelo GARP e pelo SVM (Friedman ANOVA $X^2(10,2)=13,3$; $p=0,024$).

Na estatística Kappa, o maior valor foi obtido pelo algoritmo MAXENT. Em seguida, o GARP e por último o SVM (Friedman ANOVA $X^2(10,2)=13,3$; $p=0,024$).

Quanto ao TSS, o MAXENT apresentou o maior valor, seguido pelo GARP e pelo SVM (Friedman ANOVA $X^2(10,2)=16,1$; $p=0,001$). Estas informações estão representadas na tabela 6.

No que se trata de taxas de erros e acertos, o algoritmo MAXENT apresentou os melhores valores. O GARP obteve resultado bastante próximo ao MAXENT, enquanto que o SVM teve pior performance. Os valores estão apresentados na tabela 7.

Um exemplo dos modelos gerados pelos algoritmos MAXENT, GARP e SVM encontra-se na figura 5.

Todos os resultados obtidos estão apresentados na tabela 8.

DISCUSSÃO

Analisando os resultados posso concluir que a seleção do ponto de corte afeta fortemente não só os modelos gerados e sua interpretação, como todas as métricas utilizadas para a avaliação do modelo em si. O Limite de decisão obtido pela curva ROC é sempre mais alto, o que produz modelos menos inclusivos. O Limite de decisão Mínimo sempre apresenta valores menores que o anterior, gerando resultados mais abrangentes.

Como vimos, com o Limite de decisão ROC chegamos à conclusão de que o algoritmo GARP é o melhor dentre os três analisados e ao adotarmos o Limite de decisão mínimo o MAXENT se mostra com a melhor performance.

É possível observar que, quando adotado o Limite de decisão ROC, a performance geral de todos os algoritmos apresentados decai bastante. Logo, podemos descartar bons modelos simplesmente por termos adotado um limite de decisão muito alto. Quando o limite de decisão mínimo é utilizado, era esperado que o modelo, apesar de maior quantidade de acertos, também apresentasse maior quantidade de erros de sobreprevisão. Isso não aconteceu e os modelos melhoraram perceptivelmente em relação aos erros e acertos.

Com isso, posso concluir que a adoção do limite de decisão mínimo é a mais adequada, exceto quando há uma necessidade por parte do pesquisador em optar por um limite de decisão mais elevado.

Os resultados obtidos neste trabalho reforçam a importância da prevalência, que é o número de pontos por área de estudo. Se pretendemos modelar uma espécie que apresenta poucos registros de ocorrência e uma área de distribuição geográfica supostamente restrita, devemos evitar a utilização de uma máscara de modelagem que abranja uma área geográfica muito grande. Na verdade, apesar de subjetivo, devemos sempre pensar em como o excesso de pixels afetará a forma como esses modelos fazem sua verificação (teste interno) e como avaliaremos os mesmos com técnicas sabidamente dependentes da prevalência.

A estatística Kappa mostrou-se muito afetada pela prevalência, apresentando valores que classificariam os modelos como piores do que o acaso em relação à seu grau de concordância com a realidade. Esse resultado era esperado, como discutido anteriormente por Allouche e colaboradores

(2006). O TSS por outro lado, manteve valores estáveis em ambas as abordagens de limite de decisão.

A AUC é uma medida direta da qualidade do modelo e não sofre muita influência da prevalência. Apesar de ainda ser criticada por alguns pesquisadores, seu uso ainda se justifica devido à sua robustez. Analisando apenas essa medida, o Maxent apresentou os melhores valores, que indicam maior capacidade de classificação correta dos pontos. O fato desse algoritmo atuar como um bom classificador corrobora os resultados obtidos por Graham e Hijmans (2006).

A sensibilidade, diretamente relacionada à capacidade do modelo de indicar presenças onde a espécie está presente também é responsável pelos erros de sobreprevisão, ou falsos positivos. A especificidade, que indica a capacidade do modelo de prever corretamente as ausências é relacionada aos erros de omissão. Ambas as métricas também foram muito afetadas pela escolha do Limite de decisão, se mostrando instáveis para a classificação correta de um modelo.

A relação entre sensibilidade e especificidade pode ser vista ao utilizarmos o limite de decisão baseado na curva ROC (sensibilidade no eixo y e 1-especificidade no eixo x). O ponto de corte (limite de decisão) escolhido dessa forma maximiza a chance de acertos diminuindo os erros ao equilibrar os erros de omissão e de sobreprevisão. Essa característica, que é sua maior qualidade, é também a fonte de sua principal crítica: ele lida com esses dois erros como se tivessem o mesmo 'peso', e sabemos que isso dependerá dos objetivos do trabalho em si.

O support vector machines, SVM, foi o modelo que apresentou pior performance geral. Apesar de deter os segundos maiores valores de kappa e acurácia, quando utilizado o Limite de decisão ROC, o SVM apresentou os maiores desvios-padrões dentre os três algoritmos. Mesmo com os menores limite de decisão, o SVM ainda apresentou baixos valores de AUC, sensibilidade e de TSS, ainda com os maiores desvios observados. Os valores de especificidade também foram altos, ou seja, a chance de cometer erros de omissão é maior. Mesmo utilizando o Limite de decisão Mínimo, o SVM não apresentou melhoras consideráveis.

O GARP apresentou a maior sensibilidade, especificidade, Kappa e TSS, quando adotamos o limite de decisão ROC. Quando utilizado o limite de decisão mínimo, houve uma redução em sua performance, porém, o modelo manteve-se muito próximo ao MAXENT.

O MAXENT foi o modelo que apresentou maior AUC, que poderíamos chamar de capacidade discriminatória. Seus valores de Kappa e TSS foram baixos, quando o limite de decisão ROC foi utilizado. Quando adotado o limite de decisão mínimo, este algoritmo apresentou uma performance elevada, superior aos demais. Resultados semelhantes foram obtidos por Wisz e colaboradores (2008).

Com esses resultados posso dizer que o SVM merece uma revisão de sua implementação dentro do ambiente de modelagem ao qual está atualmente inserido, o OpenModeller. Os modelos gerados apresentam baixa qualidade em quase todas as métricas utilizadas para avaliação de modelos.

Posso concluir que o GARP e o MAXENT são os melhores modelos de distribuição potencial de espécies, em relação às métricas que utilizei para avaliá-los neste estudo. Esses resultados estão de acordo com os encontrados por Elith e colaboradores (2007).

Logo, o que um pesquisador interessado em modelar sua espécie de interesse precisa é: decidir o grau de incerteza aceitável, no objetivo do trabalho. Qual a natureza da pergunta que ele quer responder: definição de áreas prioritárias para a conservação? Áreas para re-introdução de uma espécie? Uma boa indicação de novas áreas para encontro de espécies raras? Descobrir áreas sensíveis à invasão e permanência de espécies exóticas? Ou, ainda, entender um pouco mais sobre uma espécie com poucos pontos de registro utilizando os dados bioclimáticos e topográficos?

Uma vez que esses objetivos estiverem claros e o efeito e consequências dos erros em cada algoritmo forem pesados em relação ao resultado desejado, não só poderemos escolher um algoritmo em detrimento de outro com um grau de confiança maior que a simples escolha do mais usado ou mais citado na literatura, como também será possível determinar o ponto de corte a ser adotado, sem correr o risco de obter modelos inclusivos ou abrangentes demais, ou descartá-los adotando equivocadamente alguma das muitas métricas de avaliação disponíveis.

REFERÊNCIAS BIBLIOGRÁFICAS

- Allouche O., Tsoar A. & Kadmon R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43, 1223-1232
- Anderson R.P., Gomez-Laverde M. & Peterson A.T. (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography* 11, 131-141
- 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
- Arim M., Abades S.R., Neill P.E., Lima M. & Marquet P.A. (2006) Spread dynamics of invasive species. *Proceedings of the National Academy of Sciences of the United States of America* 103, 374-378
- Austin M.P., Belbin L., Meyers J.A., Doherty M.D. & Luoto M. (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling* 199, 197-216
- Baer B. & Schmid-Hempel P. (2000) The artificial insemination of bumblebee queens. *Insectes Sociaux* 47, 183-187
- Brown J.H. & Gibson A.C. (1983) *Biogeography*. C. V. Mosby, St. Louis.
- Brown J.H., Stevens G.C. & Kaufman D.M. (1996) The geographic range: Size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* 27, 597-623
- 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
- Cristianini, N., and J. Shawe-Taylor. 2000. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press.
- Drake J.M. & Bossenbroek J.M. (2004) The potential distribution of zebra mussels in the United States. *BioScience* 54, 931-941
- Elith, J.; Catherine H. Graham; Robert P. Anderson; Miroslav Dudi'k; Simon Ferrier; Antoine Guisan; Robert J. Hijmans; Falk Huettmann; John R. Leathwick; Anthony Lehmann; Jin Li; Lucia G. Lohmann; Bette A. Loiselle; Glenn Manion; Craig Moritz; Miguel Nakamura; Yoshinori Nakazawa; Jacob McC. Overton; A. Townsend Peterson; Steven J. Phillips; Karen Richardson; Ricardo Scachetti-Pereira; Robert E. Schapire; Jorge Soberón; Stephen Williams; Mary S. Wisz and Niklaus E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129/151.

- Fawcett T. ROC graphs: notes and practical considerations for researchers. 2004. HP Labs, Palo Alto, CA. Technical Report HPL-2003-4.
- Ferreira A.L., Sessa P.A., Varejao J.B.M. & Falqueto A. (2001) Distribution of sand flies (Diptera : Psychodidae) at different altitudes in an endemic region of American cutaneous leishmaniasis in the state of Espirito Santo, Brazil. *Memorias do Instituto Oswaldo Cruz* 96, 1061-1067
- 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
- Gaston K.J. (1994) *Rarity*, 1 ed. Chapman & Hall, London.
- Gomes A.C., Galati E.A.B., Casanova C., Domingos M.F., Marques G.R.A.M. & Neves V.L.F.C. (1995) Analysis of the geographical distribution of Leishmaniasis vectors in the state of São Paulo Brazil. *Boletim de la Direccion de Malariologia y Saneamiento Ambiental* 35, 143-146
- Graham C.H. & Hijmans R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography* 15, 578-587
- Grehan J.R. (1992) Biogeography and conservation in the real world. *Global.Ecol.Biogeogr.Lett.* 2, 96-97
- Guisan A., Graham C.H., Elith J. & Huettmann F. (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13, 332-340
- Guisan A., Zimmermann N.E., Elith J., Graham C.H., Phillips S. & Peterson A.T. (2007) What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs* 77, 615-630
- Guisan A. & Zimmermann N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147-186
- Hirzel A.H., Helfer V. & Metral F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145, 111-121
- Hirzel A.H., Le Lay G., Helfer V., Randin C. & Guisan A. (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199, 142-152
- MacArthur R.H. & Wilson E.O. (1967) *The Theory of Island Biogeography*, 1st edn. Princeton University Press, Princeton.
- Meynard C.N. & Quinn J.F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography* 34, 1455-1469
- Peterson A.T., Ball L.G. & Cohoon K.P. (2002) Predicting distributions of Mexican birds using ecological niche modelling methods. *Ibis* 144, E27-E32
- Peterson A.T., Egbert S.L., Sanchez-Cordero V. & Price K.P. (2000) Geographic analysis of conservation priority: endemic birds and mammals in Veracruz, Mexico. *Biological Conservation* 93, 85-94
- Phillips S.J., Anderson R.P. & Schapire R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231-259

- Rabinowitz D. (1981) Seven forms of rarity. In: *The Biological Aspects of Rare Plant Conservation* (ed Synge H.C.), pp. 205-217. John Wiley.
- Sargent R.G. Verification and Validation of Simulation Models. D.J.Medeiros, E.F.Watson, , J. S. Carson, and M.S.Manivannan. 1998. *Proceedings of 1998 Winter Simulation Conference*.
- Segurado P. & Araujo M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31, 1555-1568
- Soberon J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10, 1115-1123
- Soberon J. & Peterson A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2, 1-10
- Stockwell, D. R. B., and D. Peters. 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographic Information Systems* 13:143-158.
- Tsoar A., Allouche O., Steinitz O., Rotem D. & Kadmon R. (2007) A comparative evaluation of presence-only methods for modelling species distributions. *Diversity and Distributions*
- Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. SpringerVerlag.
- Zar, J.H. 1998. *Biostatistical Analysis*. 4th ed. Prentice Hall. pp. 929.
- Wisz, M. S.; R. J. Hijmans; J. Li; A. T. Peterson; C. H. Graham; A. Guisan; NCEAS Predicting Species Distributions Working Group; 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14, 763–773

TABELAS

Tabela 1. Variáveis ambientais utilizadas na criação da espécie artificial *Cyberconius cupparii*. A coluna “limite inferior” representa o valor mínimo para a existência da espécie na área e é calculado subtraindo-se o desvio-padrão da média. A coluna “limite superior” representa o valor máximo que a espécie tolera e é obtido somando-se o desvio-padrão à média.

Variável	Média	Desvio-Padrão	Limite Inferior	Limite Superior
Sazonalidade da temperatura	1215.98	888.96	327.02	2104.94
Sazonalidade da precipitação	52.67	19.9	32.77	72.57
Modelo digital de elevação	492.48	290.99	201.49	783.47

Tabela 2. Variáveis ambientais utilizadas pelos algoritmos no desenvolvimento dos modelos de distribuição geográfica potencial da espécie *Cyberconius cupparii*. Todas foram utilizadas com resolução de 2,5 minutos de grau (aproximadamente 5km² por pixel) para a construção dos modelos.

Variável	Média	Desvio Padrão
Temperatura anual media	230.59	29.95
Varição diurna média de temperatura	110.25	17.90
Isotermalidade	68.52	9.58
Sazonalidade da temperatura	1215.98	888.96
Temperatura máxima do mês mais quente	318.75	22.88
Temperatura mínima do mês mais frio	157.20	41.23
Varição anual da temperatura	168.95	31.54
Temperatura média do trimestre mais úmido	237.40	26.90
Temperatura média do trimestre mais seco	219.34	37.28
Temperatura média do trimestre mais quente	254.83	19.33
Temperatura média do trimestre mais frio	226.01	36.62
Precipitação annual	1904.95	664.47
Precipitação do mês mais úmido	297.10	105.55
Precipitação do mês mais seco	76.01	55.77
Sazonalidade da precipitação	52.67	19.90
Precipitação do trimestre mais úmido	765.94	233.37
Precipitação do trimestre mais seco	153.09	147.84
Precipitação do trimestre mais quente	383.73	233.33
Precipitação do trimestre mais frio	609.37	373.34
Modelo digital de elevação	492.48	290.99
Modelo de inclinação do terreno	119.55	130.04

Tabela 3. Matriz de confusão construída a partir dos erros e acertos dos modelos de distribuição potencial. O que define o erro ou acerto é o valor de corte (*limite de decisão*) adotado. 'A' são as verdadeiras presenças, previstas corretamente pelo modelo. 'B' são erros de omissão, onde o modelo previu uma ausência onde há uma presença. 'C' são os erros de sobreprevisão, normalmente difíceis de se avaliar. 'D' são as verdadeiras ausências previstas pelo modelo (Adaptado de Allouche et al, 2006).

	Presença prevista	Ausência prevista
Verdadeira presença	A	B
Verdadeira ausência	C	D

Tabela 4. Comparação entre a acurácia, Kappa, TSS, Sensibilidade e Especificidade nos modelos Maxent, GARP e SVM, quando utilizado o Limite de decisão obtido pela curva ROC. Estão representados os valores médios (n=10) e o desvio padrão dos mesmos.

AVALIAÇÃO	MAXENT		GARP		SVM	
	Média	Desvio	Média	Desvio	Média	Desvio
Acurácia	0.511	0.005	0.617	0.034	0.529	0.053
Kappa	0.008	0.009	0.225	0.070	0.045	0.125
TSS	0.471	0.077	0.562	0.020	0.219	0.389
Sensibilidade	0.962	0.076	0.991	0.005	0.697	0.383
Especificidade	0.509	0.002	0.571	0.022	0.522	0.040

Tabela 5. Taxas de acerto e erro em todos os três algoritmos , quando utilizado o Limite de decisão obtido com a curva ROC. Esses valores foram calculados com base nos números de pixels corretos e errados em relação ao número total de pixels do modelo.

Algoritmo	Taxa de Acerto (%)	Taxa de Erro (%)
MAXENT	51.07	48.93
GARP	61.68	38.32
SVM	52.86	47.14

Tabela 6. Comparação entre a acurácia, Kappa, TSS, Sensibilidade e Especificidade nos modelos Maxent, GARP e SVM, quando utilizado o Limite de decisão Mínimo. Estão representados os valores médios (n=10) e o desvio padrão dos mesmos.

AVALIAÇÃO	MAXENT		GARP		SVM	
	Média	Desvio	Média	Desvio	Média	Desvio
Acurácia	0.87	0.012	0.82	0.026	0.73	0.044
Kappa	0.74	0.025	0.64	0.052	0.46	0.088
Sensibilidade	0.84	0.031	0.86	0.033	0.72	0.067
Especificidade	0.91	0.040	0.80	0.049	0.77	0.068
TSS	0.75	0.025	0.66	0.035	0.49	0.074

Tabela 7. Taxas de acerto e erro em todos os três algoritmos, quando utilizado o Limite de decisão Mínimo. Esses valores foram calculados com base nos números de pixels corretos e errados em relação ao número total de pixels do modelo.

Algoritmo	Taxa de Acerto (%)	Taxa de Erro (%)
MAXENT	87.07	12.93
GARP	82.03	17.97
SVM	73.03	26.97

Tabela 8. Visão geral dos resultados obtidos, onde LR representa o Limite de Decisão ROC, e LM o Limite de Decisão Mínimo.

ALGORITMO	AVALIAÇÃO													
	Acurácia		Kappa		TSS		Sensibilidade		Especificidade		Taxa de Acertos (%)		Taxa de Erros (%)	
	LR	LM	LR	LM	LR	LM	LR	LM	LR	LM	LR	LM	LR	LM
MAXENT	0.511	0.87	0.008	0.74	0.471	0.75	0.962	0.84	0.509	0.91	51.07	87.07	48.93	12.93
GARP	0.617	0.82	0.225	0.64	0.562	0.66	0.991	0.86	0.571	0.80	61.68	82.03	38.32	17.97
SVM	0.529	0.73	0.045	0.46	0.219	0.49	0.697	0.72	0.522	0.77	52.86	73.03	47.14	26.97

FIGURAS

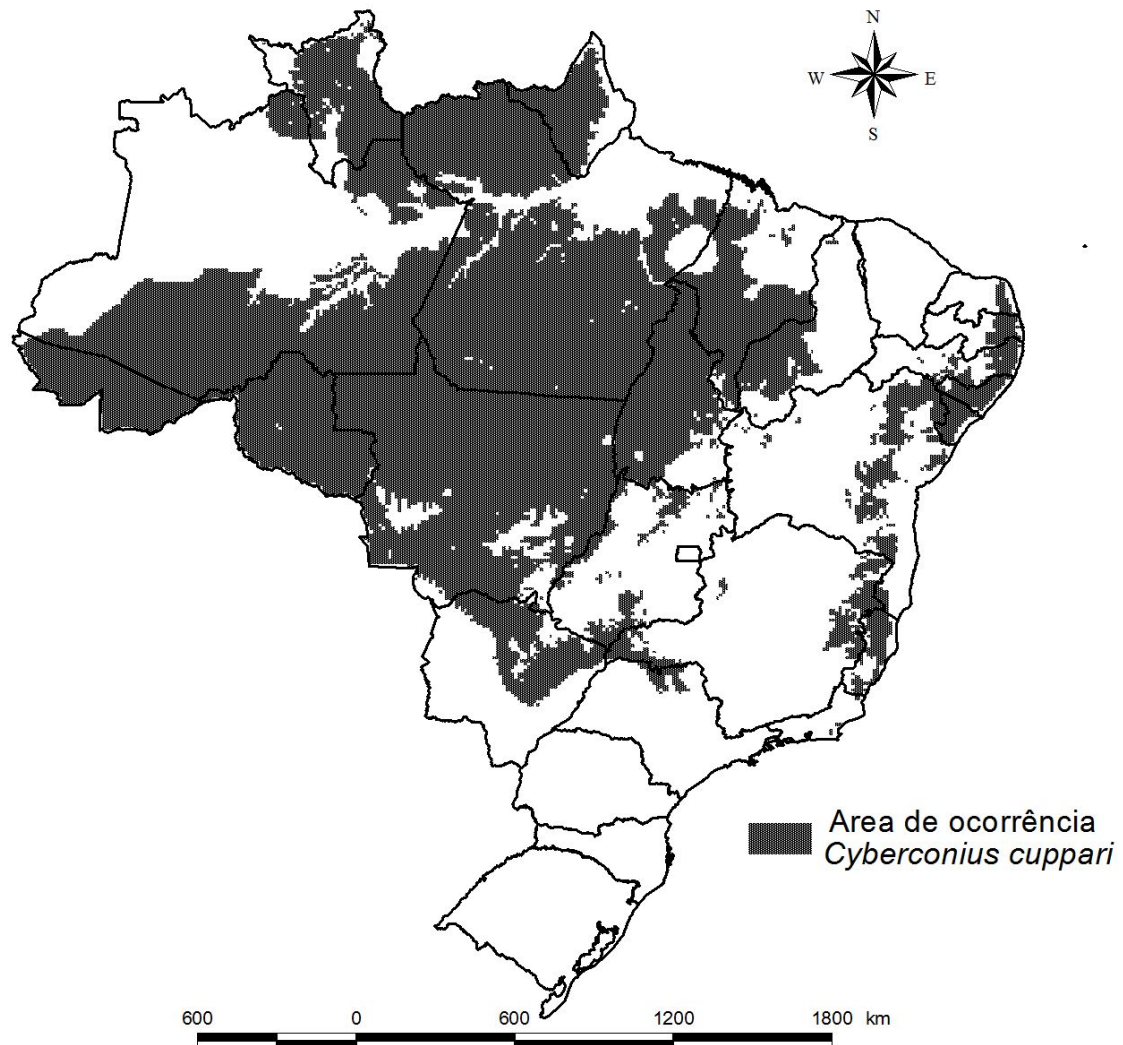


Figura 1. Mapa com a distribuição geográfica total da espécie artificial *Cyberconius cupparii*. A distribuição foi obtida através do corte de 3 variáveis ambientais em relação à sua média e mais um e menos um desvio padrão. As variáveis foram o modelo de elevação digital, a variação da temperatura e o coeficiente de variação da precipitação.

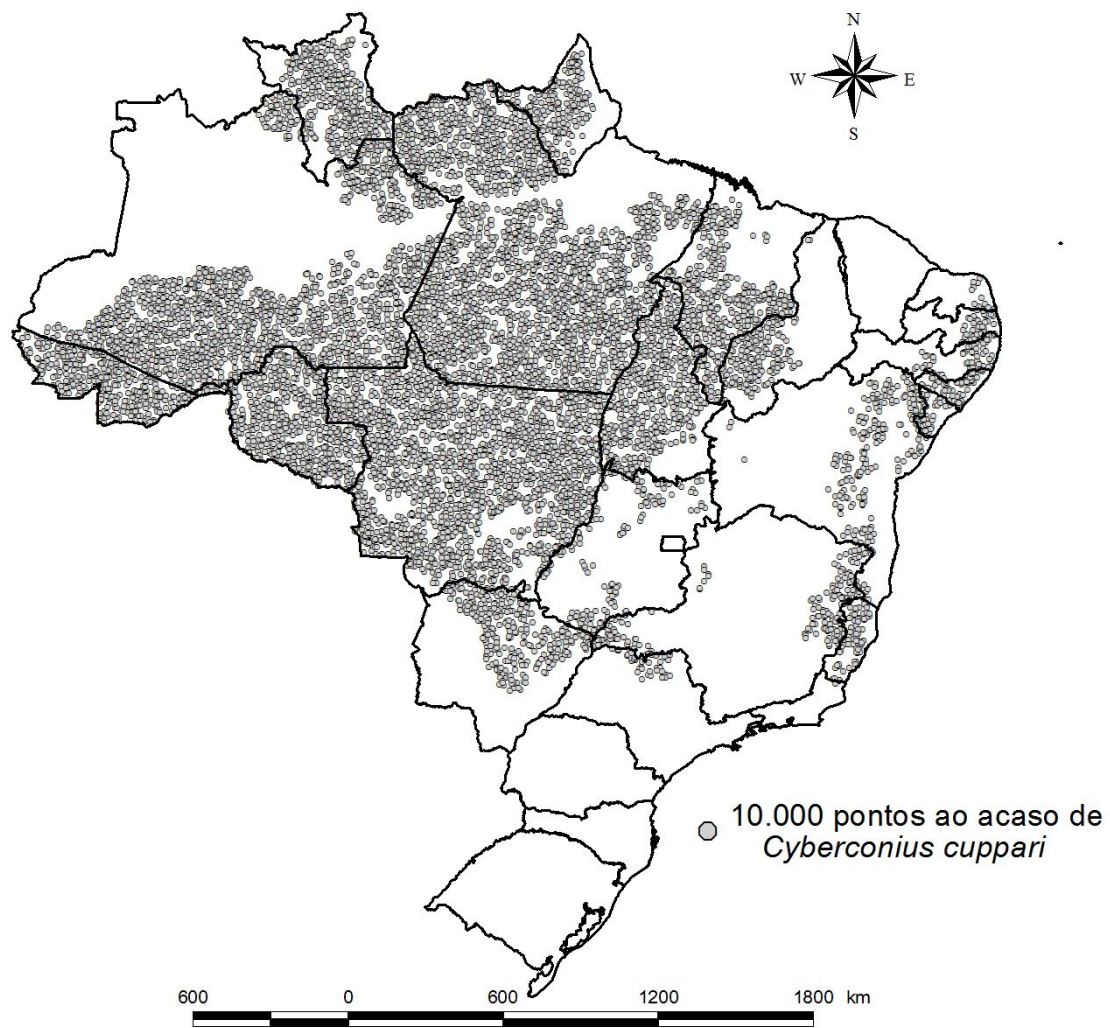


Figura 2. Mapa com os 10.000 pontos sortedos ao caso dentro da área de distribuição da espécie artificial *Cyberconius cuppari*.

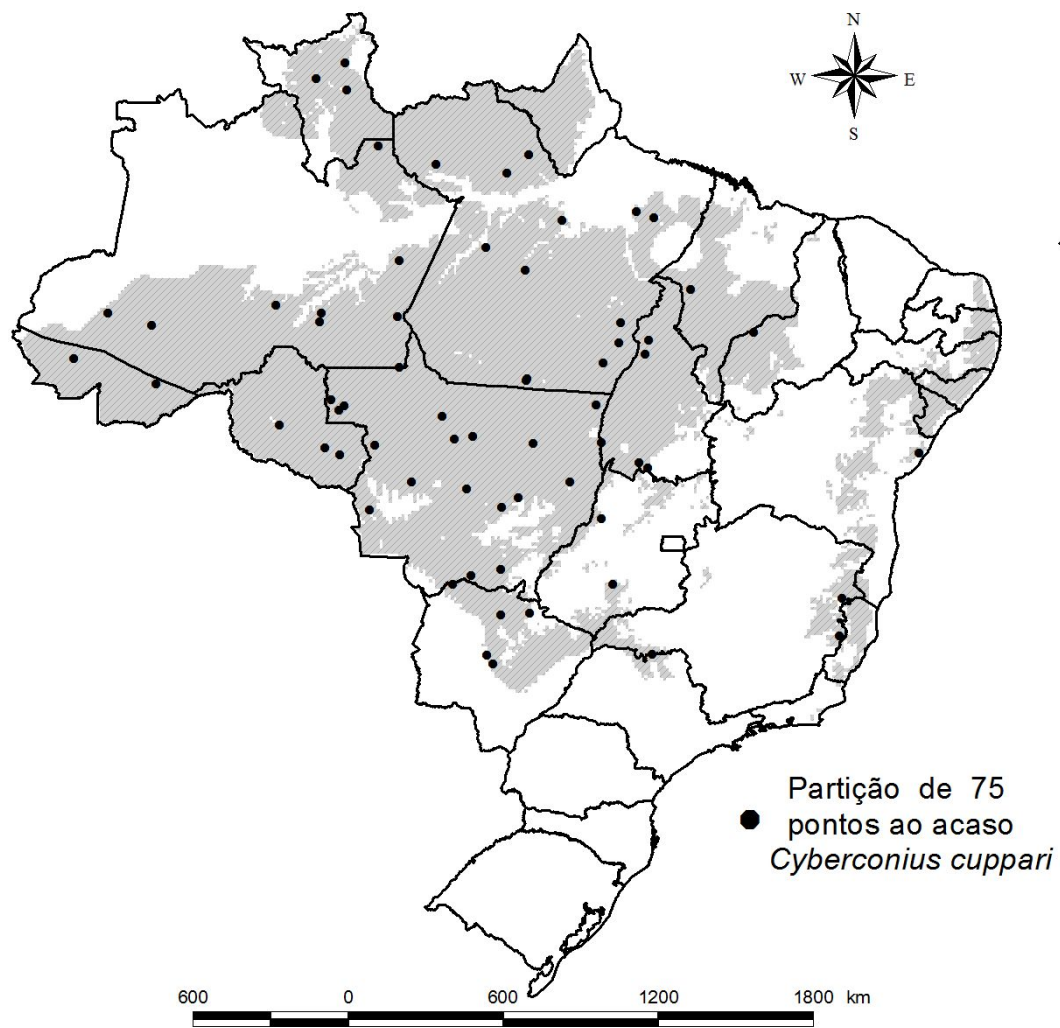


Figura 3. Exemplo de uma das dez partições de pontos ao acaso dentro da distribuição geográfica da espécie artificial *Cyberconius cupparii*. Essas partições foram usadas na criação dos modelos de distribuição potencial.

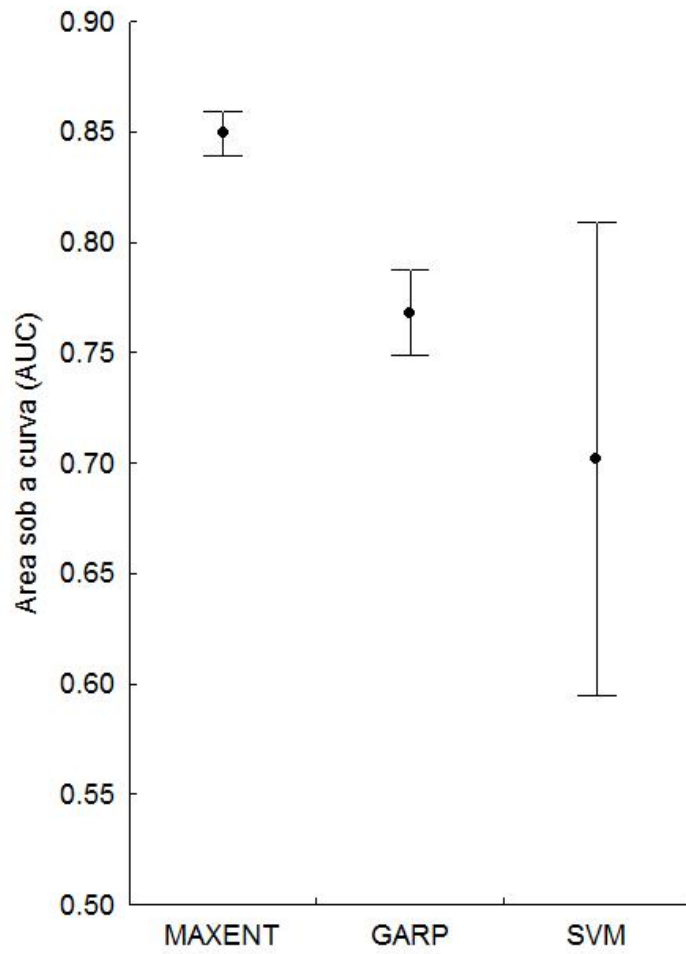


Figura 4. Comparação entre os algoritmos Maxent, Garp e SVM em relação à área sob a curva (AUC). Os pontos são os valores médios de AUC (n=10) e seus desvios-padrão.

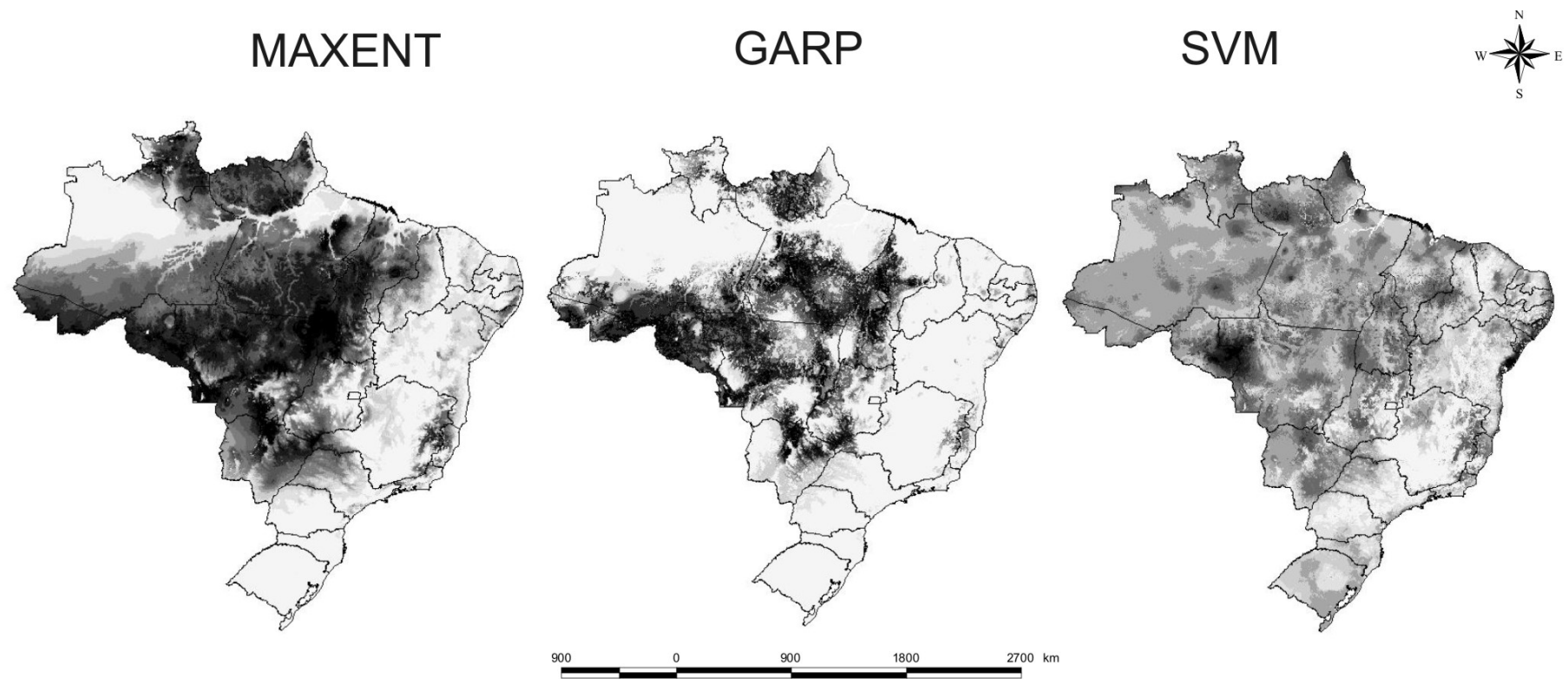


Figura 5. Exemplo de mapas gerados pelos algoritmos MAXENT, GARP e SVM, respectivamente. Para esses modelos foi utilizado um conjunto de 75 pontos da espécie artificial *Cyberconius cupparii*.

CAPÍTULO DOIS

Estratégias para conservação de espécies raras com uso de modelos de distribuição potencial de espécies: um estudo de caso com *Heliconius nattereri* Felder, 1865 (Nymphalidae: Heliconiinae).

RESUMO

BARRETO, Francisco Candido Cardoso, D.Sc., Universidade Federal de Viçosa, Dezembro de 2008. **Estratégias para conservação de espécies raras com uso de modelos de distribuição potencial de espécies: um estudo de caso com *Heliconius nattereri* Felder, 1865 (Nymphalidae: Heliconiinae)**. Orientador: Paulo De Marco Júnior. Co-orientadores: Marinez Ferreira de Siqueira e Lúcio Antônio de Oliveira Campos.

Neste trabalho utilizei um modelo de distribuição potencial de espécies para determinar áreas adequadas à presença de populações da borboleta ameaçada *Heliconius nattereri*. Adotei o software MAXENT (Maximum Entropy Modelling) para gerar mapas preditivos dessa espécie e de sua planta hospedeira *Tetrastylis ovalis*. Cruzei essas informações com o conjunto de unidades de conservação nos estados do Rio de Janeiro, São Paulo, Minas Gerais, Espírito Santo e Bahia. Mais de setenta unidades de conservação de várias categorias foram então classificadas de acordo com seu grau de adequabilidade à presença de cada espécie. De acordo com o modelo e as análises posteriores, as variáveis ambientais que mais afetaram a distribuição potencial de *H. nattereri* foram as temperaturas média e máxima do trimestre mais seco e do mês mais quente, respectivamente. O desenvolvimento de estratégias para conservação dessas espécies não é trivial e requer o esforço integrado em diversas frentes por pesquisadores na área de biologia da conservação. O Parque Nacional da Serra da Bocaina é apontado como possível local para encontro de novas populações das duas espécies.

ABSTRACT

BARRETO, Francisco Candido Cardoso, D.Sc., Universidade Federal de Viçosa, December, 2008. **Strategies to the conservation of rare species using potential distribution models: a case study with *Heliconius nattereri* Felder, 1865 (Nymphalidae: Heliconiinae).** Advisor: Paulo De Marco Júnior. Co-advisors: Marinez Ferreira de Siqueira and Lúcio Antônio de Oliveira Campos.

In this work I used a species distribution modelling approach to determine adequate areas of occurrence for populations of the endangered species *Heliconius nattereri*. I adopted the software MAXENT (Maximum entropy modelling) to generate models for this species and its host plant *Tetrastylis ovalis*. I crossed these information with the conservation areas' records in Rio de Janeiro, São Paulo, Minas Gerais, Espírito Santo and Bahia. More than seventy conservation areas were classified according with its suitability to the presence of this species. According to model applied and its interpretation, the potential distribution of *H. nattereri* was especially influenced by the average temperature of the driest trimester and the maximum temperature of the hottest month. The development of conservation strategies for these species is not trivial and only the integrated efforts in different fronts of biological conservation can lead to itseffective protection. The National Park of Serra da Bocaina is indicated as a new potential local for occurrence of both species.

INTRODUÇÃO

Heliconius nattereri Felder, 1865 é a espécie mais rara do gênero *Heliconius* (Brown Jr, 1981) e está ameaçada de extinção, tendo registros recentes de ocorrência apenas nos estados de Minas Gerais, Espírito Santo e Bahia (IBAMA, 2003). Essa espécie se diferencia dos demais *Heliconius* por apresentarem monofagia – as fêmeas adultas ovipõem apenas sobre o ápice vegetativo de uma única espécie de maracujá, *Tetrastylis ovalis* (Vell.) Killip e possuem dimorfismo sexual (Brown Jr, 1972). Os ovos de *H. nattereri* tem coloração creme e aproximadamente 1.1 x 0.8 mm (altura x largura). As fêmeas ovipõem normalmente de 1 a 3 ovos em cada meristema apical e ramos da planta. As larvas apresentam coloração branca com pontos pretos e cabeça preta e amarela. São lagartas gregárias, relativamente pequenas, com comprimento de cerca de 1.8 cm (Brown Jr, 1972). *Heliconius nattereri* ocorre desde o nível do mar até aproximadamente 1300 metros de altitude, em florestas úmidas. Os machos apresentam um vôo rápido, normalmente no dossel da floresta, enquanto as fêmeas voam mais lentamente no extrato arbustivo.

A monofagia dessa espécie de borboleta é provavelmente seu maior problema e uma das causas de sua posição na lista de espécies ameaçadas, pois *Tetrastylis ovalis*, membro da família *Passifloraceae*, é raro e também se encontra ameaçado de extinção (IBAMA, 2003).

A bibliografia sobre *T. ovalis* é escassa, com notas e comentários presentes nos trabalhos de Brown (Brown Jr, 1972; Brown Jr, 1981; Brown Jr., 1970) e um único artigo publicado que lida diretamente com a biologia e ecologia da espécie, pertencente à Buzato & Franco (1992).

Nesse trabalho, as autoras descrevem *T. ovalis* como uma trepadeira lenhosa, que tem preferência por habitat mais abertos, crescendo em galhos e ramos que se sobrepõem à altura do dossel. Suas flores são agrupadas em inflorescências racemosas, sem folhas. Seu crescimento vegetativo é lento, assim como sua produtividade anual, que é pequena. A polinização dessa espécie dá-se através da visitação de morcegos da família *Phyllostomidae*, as espécies *Glossophaga soricina* (Pallas) e *Rhinophylla pumilio* (Peters).

. Além da problemática referente à baixa oferta de planta-hospedeira, a espécie *H. nattereri* ainda sofre com competição direta e indireta de espécies mais comuns, abundantes e com maior distribuição geográfica do gênero *Heliconius*, como *Heliconius erato*, *Heliconius ethilla* e *Heliconius sara*. Essas espécies apresentam lagartas agressivas que podem atacar os ovos e pequenas larvas de *H. nattereri* (Brown Jr & Freitas, 2000; Brown Junior *et al.*, 1974; Brown Jr, 1972).

A delimitação da distribuição geográfica das espécies é um importante requisito para a adoção de ações conservacionistas mais eficientes (Austin *et al.*, 1990; Austin, 2002a). Existem diversos métodos para a delimitação da área de ocorrência de uma dada espécie, que vão desde a simples representação pontual em um mapa cartográfico até a aplicação de modelos de distribuição potencial, que visam, através do cruzamento de dados climáticos e geográficos com pontos georeferenciados de ocorrência de indivíduos da espécie de interesse, determinar áreas com graus variados de adequabilidade ambiental (Guisan & Thuiller, 2005; Austin, 2002b; Araujo *et al.*, 2005; Pearson *et al.*, 2004; Allouche *et al.*, 2006; Tsoar *et al.*, 2007). Tais modelos utilizam uma vasta gama de algoritmos distintos, mas podem ser agrupados pelo seu referencial teórico comum: a representação parcial do nicho ecológico da espécie (Santana *et al.*, 2008; Soberon, 2007).

Na literatura ecológica, o termo “nicho” é utilizado sob duas diferentes abordagens (Vandermeer, 1972; Chase & Leibold, 2003): (i) Os requisitos ambientais necessários para subsistência de uma espécie sem migração (Grinnell, 1917) e (ii) suas interações com outras espécies (Elton, 1933). Grinnell (1917) listou os fatores que potencialmente afetam a distribuição das espécies. Essas variáveis, localmente mensuradas, incluíam vegetação, alimentação, clima, solo, reprodução e locais de refúgio, efeitos interespecíficos e preferências individuais/raciais. Soberón e Peterson (2005) discutiram sobre essa separação do conceito de Nicho, com objetivo de deixar mais claras (e menos equivocadas) as interpretações dos modelos de distribuição potencial de espécies.

O desenvolvimento do sensoriamento remoto e sistemas de informação geográfica (SIG) tornaram mais disponíveis um vasta gama de variáveis ambientais, cobrindo grandes áreas que são conhecidas por influenciar a

macro-distribuição das espécies (como clima, topografia, cobertura -ou uso - da terra, e imagens de satélite) (Hirzel & Le Lay, 2008).

Uma contribuição importante dos modelos de adequabilidade ambiental para a teoria de nicho tem sido sua habilidade de testar a relevância dessas variáveis para a interação espécie-ambiente (Hirzel *et al.*, 2001; Hirzel *et al.*, 2006; Hirzel & Le Lay, 2008). Variáveis climáticas e em particular a temperatura, estão entre os fatores mais importantes que guiam a distribuição das espécies (Guisan & Zimmermann, 2000a; Grinnell, 1917).

Um destes algoritmos de adequabilidade ambiental, está implementado no software homônimo MAXENT (Phillips *et al.*, 2004; Phillips *et al.*, 2006). Esse algoritmo baseia-se no princípio da máxima entropia, que diz que a melhor aproximação para uma distribuição de probabilidades desconhecida é aquela que satisfaça qualquer restrição à distribuição. Trata-se de um método pra realizar previsões ou inferências a partir de informações incompletas. É aplicado em áreas como astronomia, reconstrução de imagens e processamento de sinais. Ele estima a probabilidade de ocorrência da espécie encontrando a distribuição de probabilidade da máxima entropia (que é a distribuição mais próxima da distribuição uniforme), submetidas a um conjunto de restrições que representam a informação incompleta sobre a distribuição alvo (Phillips *et al.*, 2004; Phillips *et al.*, 2006; Dudik *et al.*, 2004).

Meus objetivos neste trabalho foram: gerar um modelo de distribuição potencial da espécie *Heliconius nattereri* e refiná-lo utilizando a distribuição potencial do recurso alimentar de suas larvas, a Passifloracea *Tetrastylis ovalis*, aumentando a acurácia do modelo e aproximando-o mais ainda do nicho ecológico realizado da espécie; e cruzar esse mapa com as unidades de conservação presentes na região Sudeste e Bahia, para indicar áreas propícias para estudos focados na conservação da espécie *in situ*, além de analisar o efeito de diversas variáveis ambientais sobre suas distribuições.

METODOLOGIA

Os pontos de ocorrência das espécies foram obtidos de levantamento da literatura, principalmente por meio dos trabalhos do Dr. Keith Brown Jr. (1972; Brown Jr, 1981; Brown Jr., 1970) para *H. nattereri* e *T. ovalis*, além de uma

pesquisa nas coleções científicas indexadas no *SpeciesLink* (<http://splink.cria.org.br>), sistema distribuído de informações que integra, em tempo real, dados primários de coleções científicas. O sistema foi desenvolvido graças ao apoio das instituições: FAPESP, GBIF, JRS Foundation e CRIA.

Para *Heliconius nattereri* foram obtidos oito pontos distintos de ocorrência (todos derivados da literatura e pertencentes à Coleção de Lepidoptera do Museu de Zoologia da UNICAMP); para *Tetrastylis ovalis* foram 53 pontos, em um total de 36 herbários pesquisados, em todo o país.

Para gerar os modelos de distribuição potencial de *Heliconius nattereri* e de sua planta hospedeira *Tetrastylis ovalis*, adotei o software MAXENT (Maximum Entropy Modelling). Esse software é gratuito, porém de código fechado, e pode ser obtido em "<http://www.cs.princeton.edu/~schapire/MAXENT>".

Esse algoritmo baseia-se na teoria de máxima entropia e usa métodos de máxima verossimilhança para aferir quais variáveis foram importantes para o modelo. Além disso, o software calcula, através do método jackknife, a contribuição de cada variável quando usada isolada e seu peso quando a removemos do modelo final. Com base na diferença entre a máxima entropia (que pode ser, simplificada, pensada como uma distribuição uniforme) e a informação que é dada através do cruzamento de cada ponto de ocorrência da espécie com as camadas climáticas e geográficas escolhidas, o modelo fornece também uma curva de resposta para cada variável e sua contribuição ao modelo.

Esse algoritmo tem mostrado boa performance em vários trabalhos que testaram diversos programas e algoritmos de modelagem de distribuição potencial de espécies (Elith *et al.*, 2006; Guisan *et al.*, 2007; Gibson *et al.*, 2007; Hernandez *et al.*, 2006; Hijmans & Graham, 2006; Pearson *et al.*, 2007)

O conjunto de dados climáticos utilizados foram obtidos do WORLDCLIM 1.3 (Hijmans *et al.*, 2005) com 0,0416° de resolução. Isso corresponde à pixels de aproximadamente 1 km² de aresta abrangendo todo o Brasil.

Além dos modelos digitais de elevação e inclinação de terreno (Derivadas pelo satélite EROS, obtidas em <http://www.usgs.gov/>), utilizei um conjunto de dados que inclui temperatura e precipitação e interpolações dos mesmos. Cada uma dessas variáveis climáticas foi inserida no modelo em

formato de média anual, média mensal, valores extremos máximos e mínimos em cada trimestre do ano, nos períodos mais secos ou mais úmidos. As combinações formam um conjunto de 19 variáveis que somadas à elevação e inclinação resultam em 21 variáveis no modelo inicial. Essas variáveis estão listadas na tabela 1.

Como essa espécie é monofágica, utilizei também o modelo resultante da espécie de Passifloraceae *T. ovalis* para refinar o modelo preditivo final. No ArcView, sobrepos os mapas das duas espécies, produzindo um modelo de ocorrência potencial ecologicamente robusto dos locais onde poderemos encontrar a *H. nattereri*.

De posse de arquivos georeferenciados obtidos no portal do IBAMA (www.ibama.gov.br) cruzei as informações dos modelos finais das duas espécies com mais de 70 unidades de conservação. Cada uma dessas unidades foi classificada de acordo com o grau distinto de adequabilidade ambiental à espécie.

A descrição de cada variável utilizada pode ser encontrada em <http://www.worldclim.org/bioclim.htm> (Detalhes em Hijmans *et al.*, 2005).

Cada modelo gerado pelo algoritmo MAXENT foi avaliado pela área sob a curva (AUC) para seus pontos de teste.

A validação de um modelo é também conhecida como teste externo e consiste do confronto de seus resultados contra dados reais (ou semi-reais) da distribuição da espécie. Isso distingue esse processo do chamado teste interno, também conhecido como verificação, no jargão da modelagem matemática (Thompson, 2003; Perkins *et al.*, 2006).

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

As medidas tradicionais de capacidade discriminatória de um modelo dependem da adoção de um limite de decisão arbitrário (*threshold*) o que acaba inserindo mais um fator de subjetividade na interpretação do modelo (Guisan & Zimmermann, 2000b).

A escolha desse limite de decisão é usualmente baseada no conhecimento da probabilidade – subjetiva – da ocorrência da espécie de interesse (por exemplo, dados sobre ela ser uma espécie rara) assim como do

juízo das conseqüências em tomar decisões erradas (Thompson, 2003; Sargent, 1998; Sing *et al.*, 2004).

Todo modelo apresenta seus erros e acertos, que são avaliados em conjunto para determinarmos a qualidade do mesmo. Uma das melhores fontes de explicação sobre as curvas características de operação pode ser encontrada no relatório de Fawcett (2004).

Ao considerarmos um problema de classificação (no jargão da modelagem, o modelo atua como um “classificador”), cada amostra ou é positiva (espécie está presente) ou negativa (espécie ausente). Um classificador determina um valor real a cada exemplo para o qual um limite de decisão pode ser aplicado para prever a que classe ele pertence; para facilitar nós adotamos os rótulos {Y, N} para as predições de classe. A sensibilidade de um classificador para um limite de decisão particular é a taxa de todos os exemplos positivos que são classificados Y, enquanto especificidade é a taxa de todos os exemplos de negativos que são classificados como N. Nesta abordagem binária de um problema, a sensibilidade de um modelo é conhecida como a taxa de verdadeiros positivos, e representa ausência de erro de omissão. A quantidade 1-especificidade também é conhecida como a taxa de falso positivo, e representa o erro de sobreprevisão.

A curva ROC é obtida quando representamos graficamente a sensibilidade no eixo y e 1-especificidade no eixo x para todos os possíveis limites de decisão (Thompson, 2003; Garzóna *et al.*, 2006; Fawcett, 2004). No entanto as críticas quanto à sua utilização discutem que a mesma atribui uma importância semelhante a esses dois tipos de erro, o que nem sempre é verdadeiro. Além disso, sabe-se que a definição do ponto de corte deve ser feito pensando nas implicações para um determinado objetivo. De modo geral, um ponto de corte baixo gera modelos muito abrangentes ou inclusivos, os pontos de corte alto restringem os modelos criando previsões menos abrangentes sob a ótica geográfica.

Neste trabalho adotei como ponto de corte para os modelos gerados para ambas as espécies o valor mínimo atribuído por cada modelo à um ponto de ocorrência verdadeira. Essa abordagem é discutida em Phillips e colaboradores (2006).

A área debaixo da curva (AUC) é normalmente determinada conectando os pontos com linhas diretas; isto é chamado o método de trapezóide (ao invés de métodos paramétricos que ajustam uma curva aos pontos).

Uma área igual a 1 representaria o “modelo perfeito”; uma área de 0,5 indica que o modelo seleciona ao acaso (Fawcett, 2004; Baldi & Brunak, 2001).

A curva ROC é amplamente utilizada na medicina, radiologia, psicologia e outras áreas do conhecimento há décadas e é uma forma direta de avaliar o custo-benefício em uma análise de decisões (Hanley & McNeil B.J., 1982).

Esse método tem sido criticado quando aplicado aos modelos de distribuição potencial de espécies que utilizam apenas pontos de ocorrência (Vaughan & Ormerod, 2005; Manel *et al.*, 2001), e algumas propostas foram feitas para melhorar sua aplicação (Peterson *et al.*, 2008). No entanto, o software MAXENT ainda a utiliza, e as discussões sobre o tema ainda não geraram uma solução ideal (Pearce & Ferrier, 2000; Allouche *et al.*, 2006; Tsoar *et al.*, 2007; Huntley *et al.*, 2004).

Para garantir um grau aceitável de erro nos modelos, mesmo com poucos pontos para *H. nattereri*, dividi os pontos em treino (usados para rodar o modelo) e teste (usados para testar as predições do mesmo). Obtive assim sete partições de seis pontos de treino e seis pontos de teste. Com isso pude calcular uma AUC média e um desvio padrão associado à mesma. O mesmo procedimento foi adotado para *T. ovalis*, obtendo dez partições com 40 pontos de treino e dez partições de 16 pontos de teste ao todo.

RESULTADOS

O modelo resultante da distribuição de *Heliconius nattereri* prevê uma área de ocorrência ampla, bem maior do que o registrado até o momento (Figura 1A). Esse modelo apresenta uma AUC média de 0,899 (o que significa que, quando sorteados dois pontos, uma presença verdadeira e uma ausência verdadeira, o algoritmo tem 89,9% de probabilidade de classificar corretamente cada um deles) com desvio padrão de 0,09. Esse valor foi obtido através da análise da curva ROC. O limite de decisão que balanceia os verdadeiros positivos e verdadeiras ausências é aproximadamente 0,40. É esse o valor que

devemos interpretar com atenção no mapa. Valores abaixo deste têm pouca capacidade de predição.

Ao compararmos o modelo da borboleta com o modelo gerado para sua planta hospedeira (Figura 1B), percebemos que a distribuição desta limita a ocorrência da borboleta. Esse modelo apresenta uma AUC média 0,91 com desvio padrão de 0,03. O limite de decisão desse modelo é de 0,29.

O modelo final para *H. nattereri*, sobreposto ao modelo de distribuição de *T. ovalis*, prevê que elas existam mais fortemente no ES, RJ, SP e litoral da BA.

As variáveis mais importantes para criação dos modelos de distribuição potencial para *H. nattereri* e *T. ovalis* encontram-se listadas na tabela 2.

O modelo MAXENT fornece uma curva de resposta, calculada a partir da contribuição de cada variável para o modelo. Essa curva é muito importante para interpretar os requerimentos ecológicos da espécie, de acordo com as variáveis ambientais utilizadas. As figuras 2 e 3 representam essas curvas, para *H. nattereri* e *T. ovalis*, respectivamente.

A sobreposição das unidades de conservação presentes na área de ocorrência prevista para cada espécie e seus respectivos valores de adequabilidade ambiental estão listados na tabela 3 e 4.

Ao compararmos cada uma dessas unidades de conservação para detectar novas possíveis populações de ambas as espécies, vemos que além das localidades onde sua ocorrência já foi confirmada na literatura, surgiu um novo local onde estudos para preservação das espécies deveriam ser desenvolvidos: o Parque Nacional da Serra da Bocaina, entre os estados do Rio de Janeiro e São Paulo. Esse parque apresenta o maior grau de adequabilidade ambiental para ambas as espécies.

DISCUSSÃO

Existem vários fatores que contribuem para que uma determinada espécie seja rara em ambientes naturais (Gaston, 1994). Podemos pensar em relação à *H. nattereri* que a competição é um dos que exercem maior influência, devido à sua relação de monofagia com *Tetrastylis ovalis*. Essa *Passifloraceae* é utilizada como planta hospedeira pelas outras espécies de

Heliconiinae que são competidoras mais agressivas que *H. nattereri*, tais como *H. ethilla*, *H. sara* e *H. erato* (Brown Jr, 1972). Essa hipótese é discutida em um contexto mais amplo em Benson (1978), que aborda a importância do papel da competição interespecífica no particionamento de recursos dentro do gênero *Heliconius*.

Uma análise criteriosa das curvas de reposta produzidas pelos modelos (Figuras 2 e 3) nos fornecem pistas interessantes sobre as variáveis ambientais que são importantes para as duas espécies. Em relação ao principal recurso das larvas de *H. nattereri*, ou seja, *T. ovalis*, percebemos que deve haver uma pequena variação diurna da temperatura e as temperaturas médias do trimestre mais frio devem ser altas, sendo a espécie de ocorrência em regiões mais quentes. *T. ovalis* também vive em baixas altitudes, com baixa sazonalidade de temperatura. Essa planta ainda suporta uma precipitação baixa no mês mais seco e responde melhor à ambientes que apresentem isothermalidade (variação média diurna/variação média anual).

Heliconius nattereri, por outro lado, apresenta respostas peculiares às variáveis mais importantes produzidas pelo MAXENT. Todas as suas respostas são decrescentes com o aumento dos valores da variável ambiental, de modo que podemos dizer que essa espécie ocupa ambientes onde as temperaturas médias do trimestre mais seco e a temperatura máxima do mês mais quente são baixas. Ela também ocorre em ambientes com baixa precipitação no trimestre mais frio (inverno seco), uma média de 1.000 mm a 1.500 mm de precipitação anual. Os locais com maior adequabilidade ambiental à essa espécie apresentam baixa variação diurna de temperatura e isothermalidade (Figura 2).

O uso de modelos de distribuição potencial de espécies é uma ferramenta interessante para trabalhar com espécies ameaçadas que normalmente possuem poucos registros de coleta e ocorrência. Esses modelos podem atuar como indicadores de prioridade na elaboração e planejamento de coletas e de elaboração de planos de manejo. A validação dos mesmos é extremamente importante para que seu uso seja mais difundido no meio científico e venham a sobrepor metodologias mais antigas cuja eficiência ainda é questionada, mas que continuam em uso devido à falta de alternativas.

A questão abordada nesse trabalho indica que as informações fornecidas com a aplicação dos modelos, como as variáveis climáticas que influenciam na distribuição da espécie e a restrição geográfica imposta pela ocorrência de sua planta hospedeira (Figura 1B), podem ser de extrema importância para determinar os próximos passos a serem tomados nos projetos de conservação de *H. nattereri*, com a possibilidade de resultados mais concretos e a descoberta de métodos mais eficientes, sem que se caia no senso comum.

Alguns fatores que não puderam ser levantados nesse trabalho dizem respeito à competição interespecífica, predação e parasitoidismo. Sabe-se que outras espécies com distribuição mais ampla utilizam uma vasta gama de espécies de *Passiflora*, incluindo *T. ovalis*. Muitas dessas espécies como *Heliconius erato* e *Heliconius ethilla* possuem larvas canibais que podem competir direta e indiretamente pelo recurso, restringindo ainda mais a possibilidade de ocorrência de *H. nattereri* (Brown Jr & Freitas, 2000; Brown Jr, 1972).

Fora os estudos realizados por K.S. Brown Jr., existe no momento apenas um projeto *in situ* realizado por pesquisadores do Educandário Seráfico São Francisco de Assis, no município de Santa Teresa, no Espírito Santo, que tenta determinar as melhores condições para preservação de *H. nattereri*. Se os dados obtidos forem publicados em revistas indexadas – tornando-se disponíveis para a comunidade científica – essas novas informações permitirão um maior refinamento dos modelos preditivos. A parceria entre pesquisadores de campo e aqueles que desenvolvem trabalhos como este apresentado, aumentaria a precisão na tomada de decisões para a criação de áreas prioritárias para conservação e estratégias com objetivo de preservar espécies singulares como essa.

Ainda nesse contexto, uma outra discussão que não pode ser deixada de lado é o principal fator limitante da espécie: sua planta hospedeira (Figura 1C). O desenvolvimento de pesquisas com foco em *T. ovalis* precisa ser incentivado pois, como dito anteriormente, somente o estudo de Buzato e Franco (1992) levantou dados interessantes e indispensáveis sobre sua biologia. A via de polinização dessa Passifloracea não parece ser o problema, visto que as duas espécies de morcego que a polinizam são consideradas de

pouca preocupação junto à IUCN (Sampaio *et al.* 2008; Barquez *et al.* 2008). Além disso, pelo menos uma das espécies polinizadoras, *Glossophaga soricina*, demonstra grande resistência e adaptação mesmo em ambientes urbanos sendo a terceira mais abundante em um estudo de quatro anos de duração na cidade de Belo Horizonte (De Knecht *et al.* 2005).

A conservação de *H. nattereri* e seu recurso larval depende de uma visão integrada dos sistemas frágeis que essas espécies compartilham, e que cada dia sofrem mais com as atividades praticadas pelos seres humanos.

A indicação de uma nova área de possível ocorrência de *H. nattereri* no Parque Nacional da Serra da Bocaina, se confirmada, pode fornecer um novo alento ao desenvolvimento de planos de conservação eficientes para a espécie.

Do ponto de vista biogeográfico, não existem barreiras naturais que impeçam que *H. nattereri* apresente uma distribuição disjunta, e se essas novas populações existirem, os esforços para preservação dessas espécies devem envolver a obtenção de informações básicas tais como: (a) estudos de dinâmica populacional e (b) variabilidade genética dentro e entre populações. Com isso, podemos proceder à criação direta da espécie em borboletários dentro das unidades de conservação, com os dados obtidos em (a) e/ou a translocação de indivíduos entre populações, se essa necessidade for evidenciada no item (b).

Além disso, e de forma independente, podemos proceder ao cultivo e manejo de sua planta hospedeira *Tetrastylis ovalis*, aumentando assim sua abundância e disponibilidade nas áreas de proteção (Neto, 2005).

Acredito que esses esforços podem, mesmo que lentamente, recuperar os níveis populacionais das duas espécies (naturalmente baixos) e retirar *Heliconius nattereri* de sua posição de vulnerável à extinção.

REFERÊNCIAS BIBLIOGRÁFICAS

- Allouche,O., Tsoar,A. & Kadmon,R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223-1232.
- Araujo,M.B., Pearson,R.G., Thuiller,W. & Erhard,M. 2005. Validation of species-climate impact models under climate change. *Global Change Biology*, **11**, 1504-1513.
- Austin,M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.
- Austin,M.P., Nicholls,A.O. & Margules,C.R. 1990. Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. *Ecological Monographs*, **60**, 161-177.
- Baldi,P. & Brunak,S. 2001. *Bioinformatics: The Machine Learning Approach*. Cambridge, MA.
- Benson,W.W. 1978. Resource Partitioning in Passion Vine Butterflies. *Evolution*, **32**, 493-518.
- Brown Jr,K.S. 1972. Heliconians of Brazil (Lepidoptera - Nymphalidae) .3. Ecology and biology of *Heliconius nattereri*, a key primitive species near extinction, and comments on evolutionary development of Heliconius and Eueides. *Zoologica*, 47-51.
- Brown Jr,K.S. 1981. The Biology of Heliconius and Related Genera. *Annual Review of Entomology*, **26**, 427-456.
- Brown Jr,K.S. & Freitas,A.V.L. 2000. Diversidade de Lepidoptera em Santa Teresa, Espírito Santo. *Boletim do Museu de Biologia Prof.Mello Leitão, Série Proteção a Natureza*, **11/12**, 71-118.
- Brown Jr.,K.S. 1970. Rediscovery of *Heliconius nattereri* In Eastern Brazil. *Entomological News*, **81**, 129-140.
- Brown Junior,K.S., Sheppard,P.M. & Turner,J.R.G. 1974. Quaternary refugia in tropical America: evidence from race formation in Heliconius butterflies. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **187**, 369-378.

- Buzato, S. & Franco, A.L.M. 1992. *Tetrastylis ovalis*: a second case of bat-pollinated passionflower (Passifloraceae). *Plant Systematics and Evolution*, **181**, 261-267.
- Chase, J.M. & Leibold, M.A. 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. The University of Chicago Press, Chicago, IL.
- Dudik, M., Phillips, S.J. & Schapire, R.E. 2004. Performance guarantees for regularized maximum entropy density estimation. Proceedings of the 17th Annual Conference on Computational Learning Theory, 655-662. New York, ACM Press.
- 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., 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.
- Elton, C. 1933. *The Ecology of Animals*. Methuen, London.
- Fawcett, T. 2003. ROC graphs: notes and practical considerations for researchers. 2004. HP Labs, Palo Alto, CA. Technical Report HPL.
- Garzóna, M.B., Blazek, R., Neteler, M., Diosa, R.S., Olleroa, H.S. & Furlanello, C. 2006. Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestris* L. in the Iberian Peninsula. *Ecological Modelling*, **197**, 383-393.
- Gaston, K.J. 1994. *Rarity*. Chapman & Hall, London.
- Gibson, L., Barrett, B. & Burbidge, A. 2007. Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. *Diversity and Distributions*, **13**, 704-713.
- Grinnell, J. 1917. The niche-relationship of the California Thrasher. *The Auk*, **34**, 427-433.
- Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332-340.
- Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

- Guisan,A. & Zimmermann,N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.
- Hanley,J.A. & McNeil B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29-36.
- Hernandez,P.A., Graham,C.H., Master,L.L. & Albert,D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773-785.
- Hijmans,R.J. & Graham,C.H. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12**, 2272-2281.
- Hijmans, R.J., 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.H., Helfer,V. & Metral,F. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**, 111-121.
- Hirzel,A.H. & Le Lay,G. 2008. Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, **45**, 1372-1381.
- Hirzel,A.H., Le Lay,G., Helfer,V., Randin,C. & Guisan,A. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, **199**, 142-152.
- Huntley,B., Green,R.H., Collingham,Y.C., Hill,J.K., Willis,S.G., Bartlein,P.J., Cramer,W., Hagemeyer W.J.M. & Thomas,C.J. 2004. The performance of models relating species geographical distributions to climate is independent of trophic level. *Ecology Letters*, **7**, 417-426.
- IBAMA. 2003. Lista nacional das espécies da fauna Brasileira ameaçadas de extinção. www.meioambiente.gov.br/port/sbf/fauna/ordem4o.html .
- Manel,S., Williams,H.C. & Ormerod,S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.
- Pearce,J. & Ferrier,S. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127-147.

- Pearson,R.G., Dawson,T.P. & Liu,C. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285-298.
- 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.
- Perkins,N.J., Schisterman,E.F. & Vexler,A. 2006. Receiver Operating Characteristic Curve Inference from a Sample with a Limit of Detection. *American Journal of Epidemiology*, **165**.
- Peterson,A.T., Papes,M. & Soberon,J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213**, 63-72.
- 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., Dudik,M. & Schapire,R.E. 2004. A maximum entropy approach to species distribution modeling. Proceedings of the Twenty-First International Conference on Machine Learning. pp. 655-662.
- Santana,F.S., Siqueira,M.F., Saraiva,A.M. & Correa,P.L.P. 2008. A reference business process for ecological niche modelling. *Ecological Informatics* .
- Sargent,R.G. 1998. Verification and Validation of Simulation Models. D.J.Medeiros, E.F.Watson, , J. S. Carson, and M.S.Manivannan. *Proceedings of 1998 Winter Simulation Conference*.
- Sing,T., Beerenwinkel,N. & Lengauer,T. 2004. Learning mixtures of localized rules by maximizing the area under the ROC curve. Valencia, Spain. Proceedings of the 1st International Workshop on ROC Analysis in Artificial Intelligence. pp. 89-96.
- Soberon,J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115-1123.
- Soberon,J. & Peterson,A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1-10.
- Thompson,M.L. 2003. Assessing the diagnostic accuracy of a sequence of tests. *Biostatistics*, **4**, 341-351.

- Tsoar,A., Allouche,O., Steinitz,O., Rotem,D. & Kadmon,R. 2007. A comparative evaluation of presence-only methods for modelling species distributions. *Diversity and Distributions*.
- Vandermeer,J.H. 1972. Niche theory. *Annual Review of Ecology and Systematics*, **3**, 107-132.
- Vaughan,I.P. & Ormerod,S.J. 2005. The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, **42**, 720-73.

FIGURAS

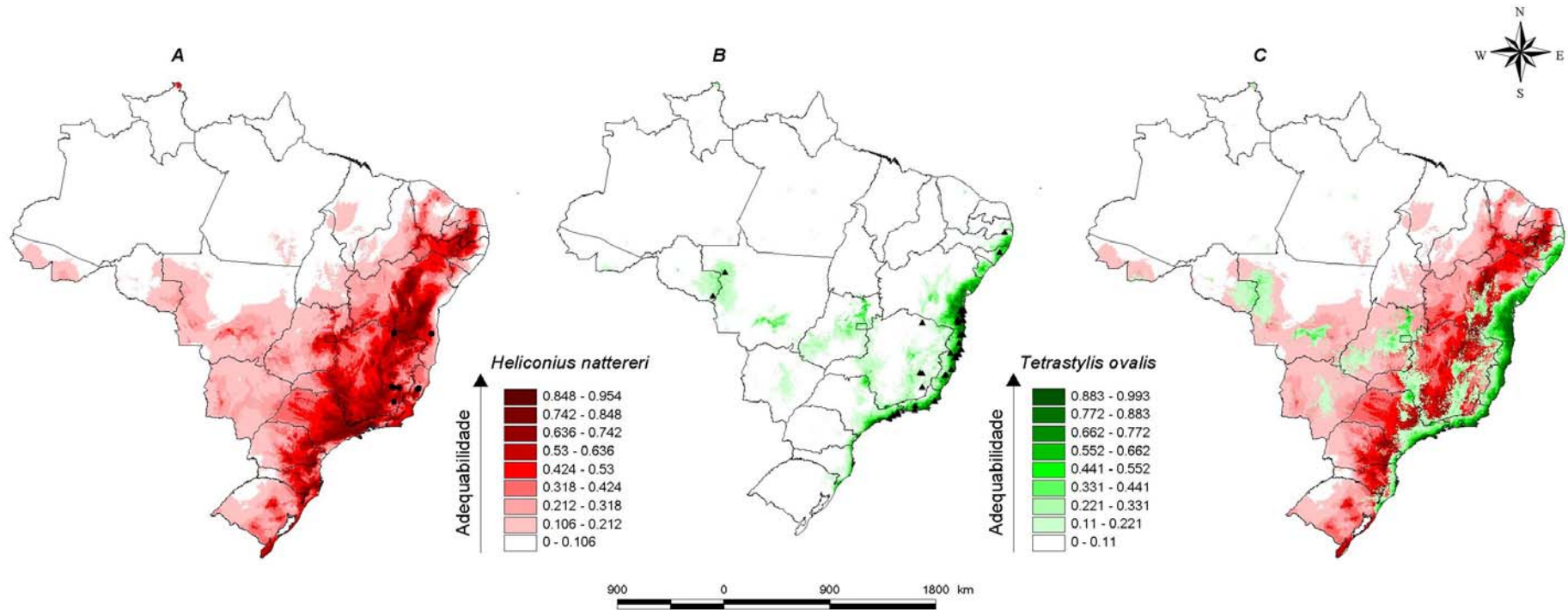


Figura 1. (A) Distribuição potencial de *Heliconius nattereri* (Felder, 1865) e (B) de *Tetrastylis ovalis* (Vell.) Killip. O mapa 'C' representa a sobreposição da distribuição das duas espécies, o que indica que *T. ovalis* atua como limitante sobre a distribuição de *H. nattereri*. Os círculos e os triângulos em preto sobre os mapas 'A' e 'B' são os pontos de ocorrência utilizados na modelagem de *H. nattereri* (8 pontos) e *T. ovalis* (62 pontos), respectivamente.

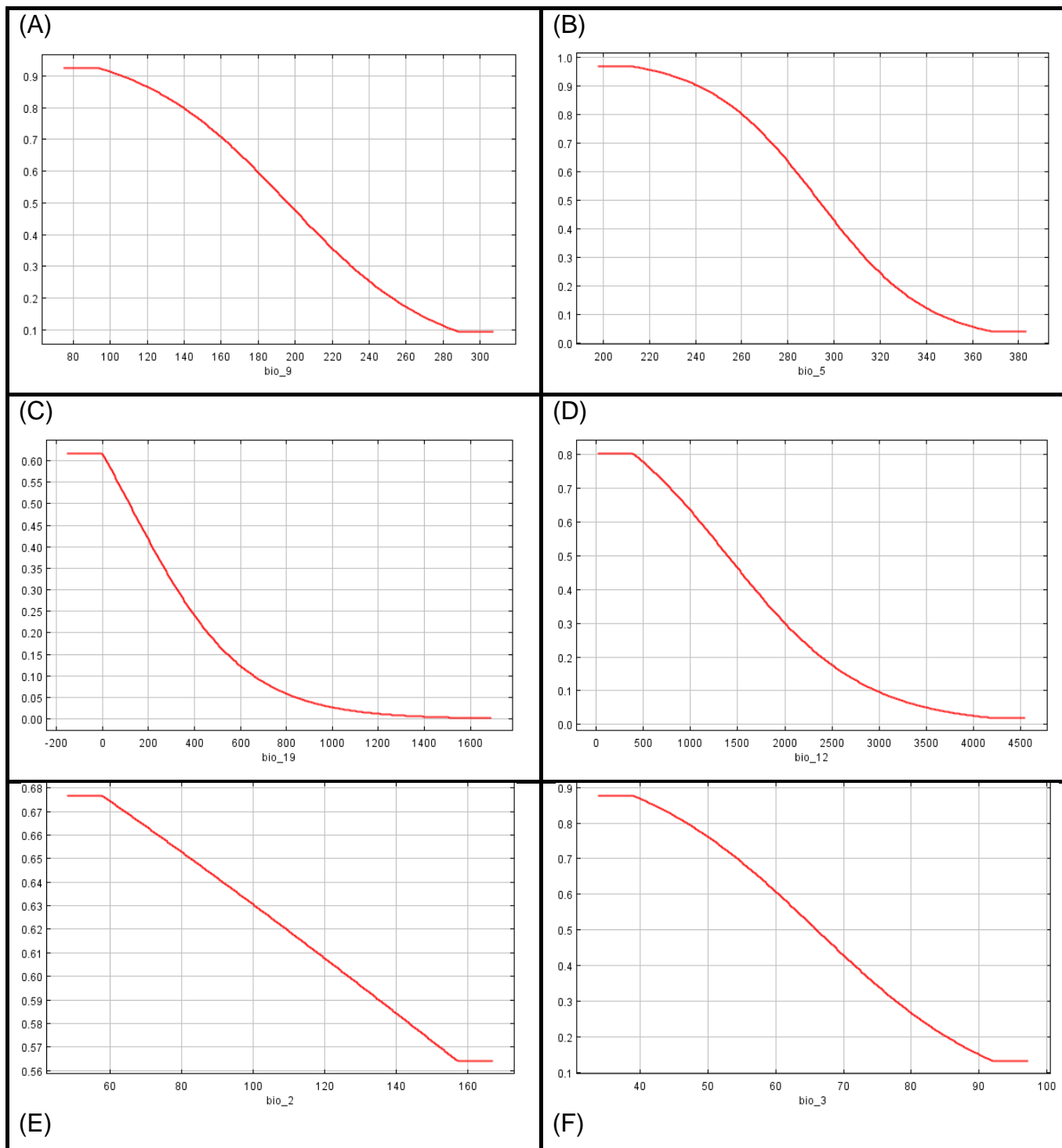


Figura 2. Curvas de resposta à cada variável ambiental utilizada na modelagem com MAXENT para a espécie *Heliconius nattereri*. O Eixo y representa a probabilidade de ocorrência da espécie e o eixo x representa os valores da variável ambiental. (A) é a temperatura média do trimestre mais seco, (B) temperatura máxima do mês mais quente, (C) precipitação no trimestre mais frio, (D) precipitação anual, (E) variação diurna de temperatura e (F) Isotermalidade. A tabela foi elaborada com grau de importância decrescente.

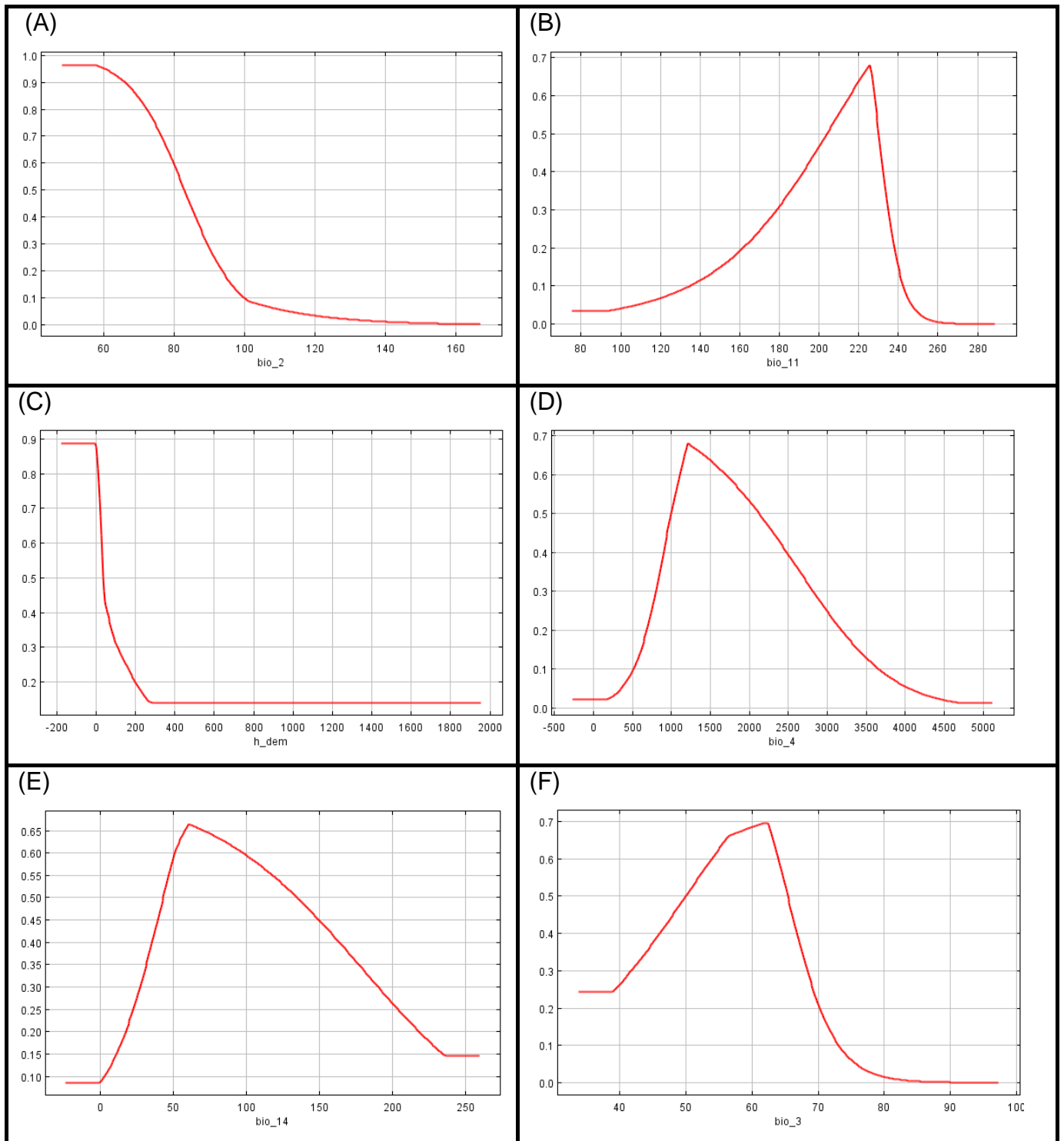


Figura 3. Curvas de resposta à cada variável ambiental utilizada na modelagem com MAXENT para a espécie *Tetrastylis ovalis*. O eixo y representa a probabilidade de ocorrência da espécie e o eixo x representa os valores da variável ambiental. (A) variação diurna de temperatura, (B) temperatura média do trimestre mais frio, (C) altitude (modelo digital de elevação), (D) Sazonalidade da temperatura, (E) Precipitação no mês mais seco e (F) Isotermalidade. A tabela foi elaborada com grau de importância decrescente.

TABELAS

Tabela 1. Variáveis ambientais e topográficas utilizadas na modelagem de distribuição geográfica potencial de *Heliconius nattereri* e *Tetrastylis ovalis*.

Código	Variável
BIO1	Temperatura anual média
BIO2	Variação diurna média de temperatura (média mensal (temperatura máxima-temperatura mínima))
BIO3	Isotermalidade (P2/P7) (* 100)
BIO4	Sazonalidade da temperatura (Desvio padrão *100)
BIO5	Temperatura máxima do mês mais quente
BIO6	Temperatura mínima do mês mais frio
BIO7	Variação anual da temperatura (P5-P6)
BIO8	Temperatura média do trimestre mais úmido
BIO9	Temperatura média do trimestre mais seco
BIO10	Temperatura média do trimestre mais quente
BIO11	Temperatura média do trimestre mais frio
BIO12	Precipitação anual
BIO13	Precipitação do mês mais úmido
BIO14	Precipitação do mês mais seco
BIO15	Sazonalidade da precipitação (Coeficiente de variação)
BIO16	Precipitação do trimestre mais úmido
BIO17	Precipitação do trimestre mais seco
BIO18	Precipitação do trimestre mais quente
BIO19	Precipitação do trimestre mais frio
h_dem	Modelo digital de elevação
h_slope	Modelo de inclinação do terreno

Tabela 2. Variáveis mais importantes na criação do modelo MAXENT para as espécies *Heliconius nattereri* e *Tetrastylis ovalis*.

<i>Heliconius nattereri</i>		<i>Tetrastylis ovalis</i>	
Variável	Contribuição (%)	Variável	Contribuição (%)
bio_9	33.9	bio_2	33
bio_5	26.9	bio_11	25.4
bio_19	23.6	h_dem	22.3
bio_12	14.1	bio_4	7.6
bio_2	1.4	bio_14	3.7
bio_3	0.2	bio_3	2.5

Tabela 3. Adequabilidade ambiental (A.A.) estimada pelo MAXENT para algumas unidades de conservação dentro da distribuição potencial de populações de *Heliconius nattereri*.

<i>Nome da Unidade de Conservação</i>	<i>A.A.</i>
P.E. dos Mananciais Campos do Jordão	0,940
P.E. Campos do Jordão	0,918
E.E. do Bananal	0,898
A.P.A. Campos do Jordão	0,892
A.P.A. do Frade	0,887
R.P.P.N. Ecoworld	0,886
P.Ec. Nascentes do Tietê	0,871
E.B. Mata do Sossego	0,869
P.E. da Cantareira	0,855
A.P.A. Fernão Dias	0,851
P.N. da Serra da Bocaina*	0,831
Estação Biológica de Santa Lúcia	0,760
Reserva Biológica Augusto Ruschi	0,700
Estação Biológica de Caratinga	0,689

Tabela 4. Adequabilidade ambiental estimada pelo MAXENT (A.A.) para algumas unidades de conservação dentro da distribuição potencial de populações de *Tetrastylis ovalis*.

<i>Nome da Unidade de Conservação</i>	<i>A.A.</i>
P.E. Xixová-Japuí	0,970
R.P.P.N. Mãe Mata	0,957
R.P.P.N. Fazenda Paraíso	0,954
R.P.P.N. Mãe da Mata	0,954
R.P.P.N. Ecoparque De Una	0,949
P.E. da Ilha Anchieta	0,947
R.P.P.N. Fazenda São João	0,947
R.P.P.N. Reserva Salto Apepique	0,947
R.P.P.N. Toque Toque Pequeno	0,944
R.P.P.N. Fazenda Arte Verde	0,943
R.P.P.N. Fazenda Meandros	0,942
R. Ex. Arraial do Cabo	0,941
A.P.A. da Costa de Itacaré - Serra Grande	0,937
R.P.P.N. Carbochloro	0,935
R.B. de Una	0,932
R.Ec. de Massambaba	0,928
R.P.P.N. Fazenda Bom Sossego	0,922
R.P.P.N. Fazenda Ararauna	0,922
A.P.A. da Lagoa Encantada	0,918
A.P.A. de Santo Antônio	0,914
A.P.A. de Massambaba	0,913
A.P.A. de Maricá	0,912
R.P.P.N. Fazenda Avaí	0,912
R.P.P.N. Lagoa Do Peixe	0,909
R. Ex. de Canavieiras	0,909
P.E. de Itaúnas	0,905
P.N. da Serra da Bocaina*	0,774
Estação Biológica de Santa Lúcia	0,750
Reserva Biológica Augusto Ruschi	0,700
Estação Biológica de Caratinga	0,642