

---

# Uso de modelos aditivos generalizados na estimativa da distribuição potencial de espécies

PAULO DE MARCO JÚNIOR

Universidade Federal de Goiás, Goiás, Brasil.  
e-mail: pdemarco@icb.ufg.br

## RESUMO

Os modelos de distribuição potencial têm sido cada vez mais utilizados para prever a distribuição atual das espécies, discutir padrões de riqueza e modelar sua distribuição em relação às mudanças climáticas globais. Nesse trabalho, apresento algumas bases teóricas relevantes à compreensão desses modelos e à discussão sobre os mecanismos que determinam a distribuição das espécies. A partir disso, discuto a estrutura estatística dos modelos aditivos generalizados e seu potencial como métodos preditivos eficientes de modelagem. Um exemplo, usando espécies de planta do gênero *Inga*, é apresentado, mostrando a facilidade de execução dessa estratégia utilizando a metodologia GRASP (predição espacial por regressão generalizada) e algumas limitações e perspectivas para melhorar as técnicas de definição de pseudoausências nesses modelos.

## ABSTRACT

*Species distribution models are increasingly used to predict present day distribution of species, discuss their richness patterns and model its distribution in relation to global climatic change. Here, I present some theoretical basis to understand these models and the discussion about the mechanisms that determine species distribution. Based on this general background, I discuss the statistical structure of the generalized additive models and its potential use as efficient predictive modeling technique. A working example using plant species of *Inga* is presented, showing how easy is the execution of this strategy using GRASP (generalized regression and spatial prediction) methodology and some limitations and perspectives to improve the pseudo-absence use in these models.*

## COMO SE ESPERA DETERMINAR A DISTRIBUIÇÃO POTENCIAL DE UMA ESPÉCIE?

Dos distintos modelos que estão sendo apresentados neste conjunto de trabalhos e em muitos outros, que estão disponíveis na literatura (Peterson, 2003; Guisan & Thuiller, 2005; Phillips *et al.*, 2006), pode-se perceber

que a busca para determinar a distribuição potencial de espécies biológicas mantém um procedimento geral comum. Na maioria destes métodos, se assume que:

1. As espécies respondem a gradientes ambientais de uma forma previsível;
2. A forma da relação entre a presença ou a presença/ausência da espécie relatada para uma parte de sua

área de distribuição ou em uma amostra de sua distribuição reflete a forma desta relação em toda a área;

3. É possível estimar a forma desta distribuição e extrapolar este resultado para outras áreas, determinando áreas onde potencialmente esta espécie está ou deveria estar presente.

O item 1 é básico e implicitamente admitido em todos os modelos. O item 2 é um dos que têm atraído grande atenção por incluir o problema sobre a forma como os dados são relatados. Podem-se distinguir dois tipos de dados: i) dados onde só a presença é relatada, como os dados de museus e herbários; e ii) dados de presença e ausência. Espera-se que dados de presença e ausência sejam tomados de forma padronizada entre locais de coleta, evitando problemas de amostragem (Zaniewski *et al.*, 2002). Modelos como o GARP – “Genetic Algorithm for Rule Prediction” (Stockwell & Noble, 1992; Peterson, 2001) e o ENFA – “Ecologic Niche Factor Analysis” (Hirzel *et al.*, 2001, Brotons *et al.*, 2004) podem trabalhar com dados apenas de presença de espécies e são, nesse sentido, mais robustos.

A maior parte dos métodos se distingue pela forma como estimam a forma da curva (o item 3 no esquema acima), mas existem diferenças maiores em alguns deles. Em especial, os modelos aditivos generalizados vão diferir de uma forma radical neste aspecto, o que pode determinar que seu uso deva estar relacionado a diferentes propostas, como espero demonstrar no decorrer deste texto.

## QUE FATORES DETERMINAM A DISTRIBUIÇÃO DAS ESPÉCIES?

O padrão de distribuição de espécies sempre tomou conta de uma parte considerável da literatura ecológica (Andrewartha & Birch, 1954; Bock & Ricklefs, 1983; Brown, 1984; Arita *et al.*, 1990; Austin, 2007). A forma como determinada espécie se distribui ao longo de um gradiente ambiental tanto é usada como base para teorias gerais, incluindo sucessão ecológica (Reader, 1992) ou a teoria do rio-contínuo (Vannote *et al.*, 1980), como para explicações de sua distribuição em uma escala biogeográfica.

Muitas destas abordagens baseiam suas predições em uma forma específica de relação entre a probabilidade de ocorrência e o gradiente ambiental. Assim, as funções de incidência de Diamond (Diamond, 1975; Gilpin & Diamond, 1981) podem ser consideradas como parte deste sistema geral, distinguindo-se apenas pelo fato do gradiente sob análise ser a área de ilhas oceânicas.

A abordagem das funções de incidência é interessante por diferenciar modelos possíveis de distribuição em função de características bionômicas das espécies, em especial suas habilidades competitivas (Grant, 1966; Grant & Abbott, 1980), mas também suas habilidades de colonização (Cole, 1981; Berggren, 2001). No entanto, outras abordagens também buscam distinguir, em maior ou menor grau, a forma da curva de resposta da espécie em relação ao gradiente ambiental e suas características intrínsecas.

Evidentemente, a forma mais básica que se pode apresentar do que foi exposto acima são as curvas que determinam os limites de tolerância das espécies (Austin, 2007). Em geral, espera-se que a eficiência das espécies em sobreviver, crescer e reproduzir ocorra em função de sua eficiência em resposta ao gradiente ambiental. Os livros-texto de ecologia buscam expressar esta relação com uma função normal ou outra de forma semelhante (*e.g.* Begon *et al.*, 1996).

Se as respostas das espécies aos gradientes ambientais são, em parte, resultado de suas características intrínsecas, espera-se que haja diferentes tipos de formas das curvas de resposta, no mínimo algo mais diverso do que variações de curtose e assimetria da curva normal. Estes diferentes modelos deveriam resultar da forma como a espécie responde ao gradiente, seja ele simplesmente um fator abiótico que interage com o conjunto de adaptações morfológicas e fisiológicas da espécie, ou um recurso que determina complexos modelos de resposta competitiva e de interações intra- e interespecíficas, incluindo um aumento do desvio da simetria típica da curva normal e respostas não uni-modais (Oksanen & Minchin, 2002). Evidentemente, os modelos mais complexos deveriam advir da forma como alguns gradientes abióticos podem determinar a distribuição, o acesso ou a eficiência de uso de recursos no ambiente.

Neste ponto, podemos construir a seguinte cadeia lógica: características bionômicas → resposta a gradientes ambientais → forma da curva de resposta → padrão de distribuição biogeográfica. Partindo desta teoria, modelos estatísticos usados para estimar a forma da curva podem ser também utilizados para testar hipóteses sobre como se processa o efeito dos fatores ambientais, ou como determinadas espécies, que compartilham características ecológicas, devem apresentar o mesmo tipo de resposta ao gradiente. Modelos deste tipo, à semelhança dos modelos baseados na função de incidência, utilizados nos estudos de metapopulação e ecologia da paisagem (Taylor, 1991; Wahlberg *et al.*, 2002; Ovaskainen & Hanski, 2003), podem ser

utilizados não apenas para a estimativa da distribuição potencial de uma espécie, mas também para testar os processos que determinam esta distribuição.

Como qualquer modelo teórico que se presta a distinguir processos, estes métodos devem representar curvas simples. Bons exemplos de modelos com estas características seriam os modelos logísticos (ou Logit) e Probit. Ambos têm certa flexibilidade, ajustando-se a alguns poucos padrões possíveis de distribuição, e são estimados com o uso de poucos parâmetros (Manel *et al.*, 1999; Pearce & Ferrier, 2000). Alguns modelos são, portanto, mais interessantes para refletir essas relações, sendo desejável que nós otimizemos neles sua generalidade, enquanto em outros o que buscamos otimizar é uma predição prática acurada da distribuição potencial da espécie (Guisan & Zimmermann, 2000). Neste último grupo é que os métodos baseados em modelos aditivos generalizados estão agrupados.

## O QUE SÃO MODELOS ADITIVOS GENERALIZADOS?

Os modelos aditivos generalizados foram propostos por Hastie & Tibshirani (1990). Estes modelos são caracterizados pelo fato de assumirem que a média de uma variável resposta depende de um preditor aditivo, através de uma função de ligação. Uma característica importante destes métodos é que a função de distribuição de probabilidade desta variável dependente pode ser qualquer uma dentre aquelas que fizerem parte da família exponencial. Isto permite a construção de modelos aditivos para variáveis com distribuição normal (*e.g.* biomassa de espécies), Poisson (*e.g.* riqueza de espécies) ou binomial (*e.g.* função de incidência das espécies).

Suponha que  $y$  é uma variável resposta aleatória e que  $X_1 \dots X_p$  é um conjunto de variáveis preditoras. Em um procedimento de regressão, consideramos como a variável dependente  $Y$  pode depender de  $X_1 \dots X_p$ , e o valor esperado de  $Y$  assume a seguinte forma linear:

$$E(Y) = f(X_1, \dots, X_p) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Equação 1

Dada uma amostra, os parâmetros  $b_1 \dots b_j$  podem ser estimados, normalmente pelo método dos quadrados mínimos. O modelo aditivo generaliza o modelo linear da seguinte maneira:

$$E(Y) = f(X_1, \dots, X_p) = s_0 + s_1(X_1) + \dots + s_p(X_p)$$

Equação 2

Onde  $s_i(\cdot)$  são as funções de ajuste. Já o modelo aditivo generalizado assume também a existência de uma função de ligação entre  $f(X_1, \dots, X_p)$ . É isto que permite que se possam utilizar diferentes funções de distribuições, além da distribuição normal, para modelar os resíduos. É claro que para cada tipo de distribuição haverá uma função de ligação apropriada.

É esta propriedade que expande a utilidade dos modelos generalizados. Usualmente os modelos baseados na função de distribuição normal não são apropriados para uma série de aplicações na área ecológica. Os dados de riqueza de espécies são, na maior parte, considerados discretos (sendo discreta uma variável em que a menor distância entre dois números é grande comparada com a amplitude de variação). Quando estudamos riqueza de espécies de mamíferos ou répteis, o modelo mais apropriado assumiria uma distribuição Poisson para os resíduos. É claro que se você estiver estudando número de espécies de Coleoptera no dossel de florestas Amazônicas, a curva normal poderá ser bastante apropriada.

Voltando ao modelo, a partir da Equação 2:

$$\eta = s_0 + \sum_{i=1}^p s_i(X_i)$$

Equação 3

onde  $\eta$  é considerado o termo aditivo. A relação entre a média da variável resposta (digamos  $\mu$ ) e o fator aditivo é definida pela função de ligação:  $g(E(Y)) = \eta$ .

O ajuste de um modelo generalizado é feito mantendo algumas características dos modelos aditivos que o precederam. Nestes modelos, cada função é ajustada separadamente, garantindo uma maior eficiência local, mas podendo perder em universalidade. Esta estratégia permite enfrentar o que foi chamado de “maldição da dimensionalidade” na qual o aumento de variáveis preditoras eleva de forma drástica as estimativas de variância. Além disto, o ajuste local é feito minimizando uma função de perda (normalmente um erro quadrático) pela escolha de funções, em vez de parâmetros individuais. Este procedimento poderia ser importante para uma maior eficiência quando os parâmetros individuais apresentam alta covariância dentro dos modelos.

Adicionalmente, Hastie & Tibshirani (1990) chamam muita atenção para o algoritmo de ajuste por “backfitting”, considerado essencial nesta aplicação. Neste mecanismo, cada interação do procedimento de ajuste busca minimizar a função de perda em relação a uma das funções (uma das variáveis preditoras)

independentemente, até a convergência. Hastie & Tibshirani (1990) provaram que este algoritmo atinge uma solução única independente de valores iniciais para funções de ajuste simétricas, como, por exemplo, as funções *spline*.

Por fim, o tipo de função de ajuste é uma característica essencial no modelo aditivo generalizado. Nestes modelos, funções não-paramétricas *spline* são utilizadas. Nestas funções, o número de parâmetros pode ser controlado e seu aumento (que leva um aumento do número de graus de liberdade do modelo) leva um aumento da qualidade de ajuste. Este comportamento pode ser comparado ao aumento do número de graus de um ajuste polinomial, que produz uma função mais flexível ao ajuste aos dados coletados com sacrifício de graus de liberdade.

Esse tipo de modelo foi utilizado para predição de distribuição de espécies em uma variedade de contextos nos últimos anos (Guisan *et al.*, 2002; Zaniwski *et al.*, 2002; Leathwick *et al.*, 2006), e faz parte de programas computacionais para análise de distribuição como o BIOMOD (Thuiller, 2003) e o GRASP (Lehmann *et al.*, 2002), o que tende a popularizar ainda mais seu uso.

#### **Padrões de distribuição de duas espécies do gênero *Inga***

Em um pequeno exercício sobre o funcionamento deste método, eu usei a distribuição de espécies do gênero *Inga* (Mimosaceae). As razões para isto são: a disponibilidade destes dados (cedidos em formato digital pelo Dr. Mike Hopkins) e a existência, neste conjunto de dados, de informações de espécies restritas e amplamente distribuídas, o que seria desejável para a avaliação do método. Este conjunto de dados refere-se também a uma classe importante de informações para estimar a distribuição potencial de espécies, que são espécies depositadas em herbários e museus. Estes dados não advêm de levantamentos completos de flora, mas do acúmulo de informações na literatura e em herbários. Esta fonte será extremamente utilizada na determinação de áreas de distribuição potencial de espécies de agora a um futuro próximo e é interessante avaliar seus resultados. Por fim, é evidente, mas sempre importante dizer, que estes dados são incompletos e resultam de um sem número de vieses de coletas, de pesquisadores e até mesmo de identificação de espécies. Talvez seja a hora de confiar mais no teorema da estatística que sugere que a soma de  $n$  variáveis aleatórias, independentemente distribuídas, tende a uma distribuição normal.

Informações gerais sugerem que o gênero *Inga* é composto por espécies que se distribuem principalmente na Amazônica e na Mata Atlântica, com espécies que

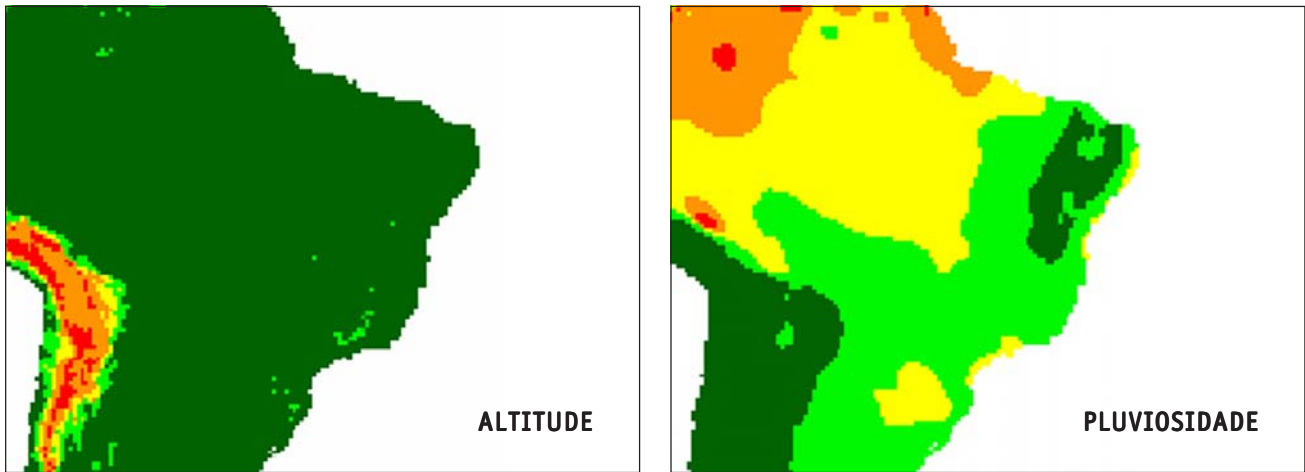
possivelmente estão associadas a regiões de maior pluviosidade e temperatura (M. Hopkins, com. pess.).

#### **AJUSTANDO MODELOS ADITIVOS GENERALIZADOS PELO PROCEDIMENTO GRASP**

Neste exercício usei o procedimento GRASP – “Generalized Regression Analysis for Spatial Prediction” de acordo com a descrição e sugestões presentes em Lehmann *et al.* (2002). O procedimento pode ser implementado através do programa estatístico R, de distribuição gratuita pela rede mundial de computadores. Informações detalhadas e aquisição dos programas necessários podem ser obtidas em <http://www.cscf.ch/grasp>.

O processo começa com a aquisição das variáveis preditoras em um formato apropriado para a análise. Usualmente estes dados são informações climáticas, de relevo ou de distribuição de alguma característica considerada importante e que está apresentada no formato de uma matriz, onde cada célula corresponde a um pixel da imagem base do estudo (na terminologia dos usuários de sistemas de informação geográfica: “ASCII RASTER”). Esta imagem representa o mapa da região em que, tanto os dados de distribuição foram obtidos, quanto para onde as predições do modelo serão produzidas. Na Figura 1 apresento dados de altitude e pluviosidade para uma porção da América Latina, os quais serão a base de dados ambiental neste exercício.

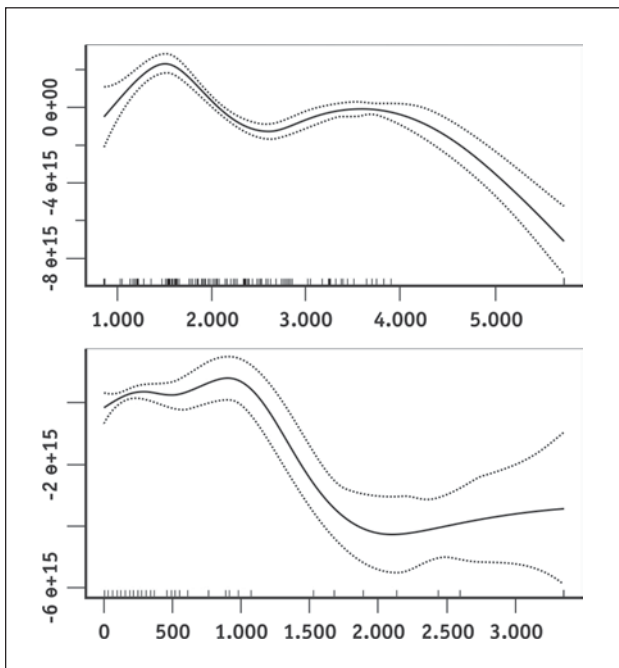
Para este exercício, utilizei os dados de *Inga virescens* Benth. e *I. macrophylla* Willd. No entanto, os dados utilizados provêm de material depositado em herbários onde apenas é relatada a informação da presença de determinada espécie. Zaniwski *et al.* (2002) discutem em detalhe as várias alternativas para se tratar dados de presença apenas, utilizando os modelos aditivos generalizados. O processo passa, necessariamente, por se estabelecer o que aqueles autores chamaram de “pseudoausências”. Ao invés de usar um critério estatístico (que discutirei mais além) para estabelecer as pseudoausências, utilizei um arrazoado prático. Considerei que os coletores que estiveram em campo não tinham vícios sérios de amostragem de uma espécie de *Inga* em relação à outra e que, quando estava presente em uma localidade, ela sempre foi relatada. Assim, associei pseudoausências para cada espécie para os locais onde ela não foi coletada, mas a outra espécie foi. Este método é obviamente viciado, principalmente comparando-se espécies de ampla distribuição com espécies de distribuição restrita, mas foi utilizado aqui apenas como uma abordagem preliminar.



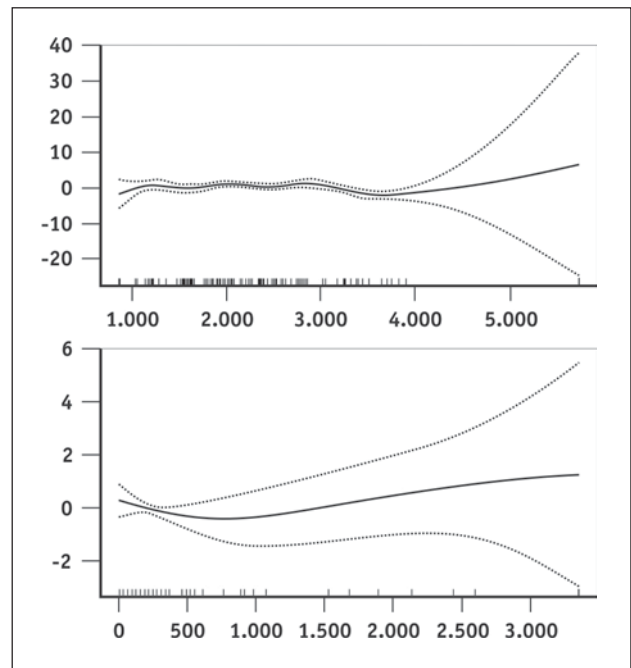
**FIGURA 1** – Distribuição de valores de altitude e pluviosidade em uma porção da América Latina que será a base deste exercício. As cores representam as classes de valores de cada variável em um gradiente, na seguinte ordem: verde escuro, verde claro, amarelo, alaranjado e vermelho escuro. Esta é apenas uma representação geral com os dados em classes para fácil visualização dos grandes padrões, porém os dados utilizados na análise são os dados brutos, sem classificação.

No ajuste pelo procedimento GRASP, o próximo passo é a escolha do modelo que será utilizado. Como os dados agora representam presença/ausência, escolhi uma ligação por uma função logística e, como função de distribuição dos resíduos, a distribuição Binomial. Outras escolhas podem ser admitidas ou testadas, mas essa é provavelmente a mais apropriada.

A partir disso, o programa exibe uma série de resultados que descrevem o ajuste do modelo e seu resultado geral. Dentre eles, é apresentado um gráfico que apresenta as previsões para cada variável e seus respectivos intervalos de confiança. Estes resultados são apresentados na Figura 2, para *I. virescens*, e na Figura 3, para *I. macrophylla*.



**FIGURA 2** – Apresentação da saída do programa GRASP para o ajuste das presença/ausência de *I. virescens* para pluviosidade (gráfico superior) e altitude (gráfico inferior). As linhas tracejadas representam o intervalo de confiança de 95%.



**FIGURA 3** – Apresentação da saída do programa GRASP para o ajuste das presença/ausência de *I. macrophylla* para pluviosidade (gráfico superior) e altitude (gráfico inferior). As linhas tracejadas representam o intervalo de confiança de 95%.

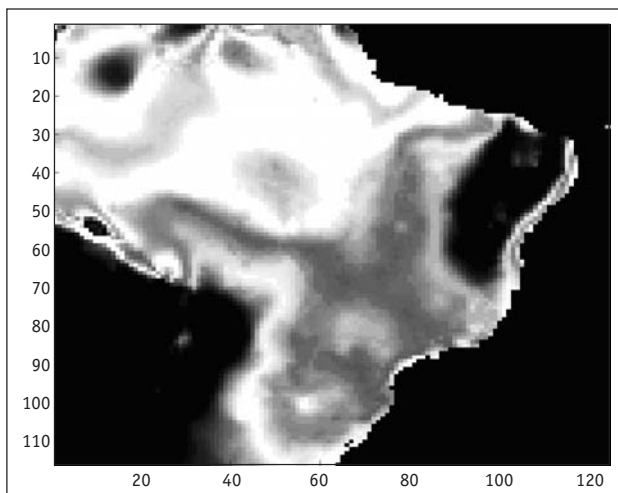
Os resultados sugerem um ajuste muito mais claro para *I. virescens* do que para *I. macrophylla*. Em especial, os maiores intervalos de confiança podem ser observados para o ajuste à altitude em *I. macrophylla*. Isto pode sugerir que a precipitação é mais importante que a altitude na determinação da ocorrência desta espécie. A interpretação do modelo é feita principalmente através da inspeção dos gráficos para “interpretar e entender esses modelos de regressão” (Lehmann *et al.*, 2002).

Na realidade, o ajuste de *I. macrophylla* é pobre ( $R^2=0,199$ ), mas o ajuste da função de precipitação [S(precipitação)] foi significativo ( $\chi^2=23,279$ , graus de liberdade estimado = 8,  $p=0,002$ ), enquanto o ajuste à altitude não foi ( $\chi^2=2,763$ , graus de liberdade estimado = 2,  $p=0,283$ ).

Ao final do processo, o programa produz arquivos de predição da ocorrência da espécie para a área da imagem de entrada das variáveis ambientais, que podem ser lidos e tratados nos sistemas de informação geográfica mais comuns (Figuras 4 e 5). Há uma diferenciação maior entre as áreas de ocorrência para *I. virescens*, com distribuição mais claramente associada a áreas de clima mais úmido, enquanto *I. macrophylla* apresenta uma distribuição que inclui as áreas de cerrado do Brasil central. A ocorrência de ambas não é predita para áreas muito secas do nordeste brasileiro.

#### CONSIDERAÇÕES E PERSPECTIVAS SOBRE O USO DE MODELOS ADITIVOS GENERALIZADOS

O ajuste apresentado nas Figuras 2 e 3 é muito bom para a natureza e quantidade de dados deste exercício. Em

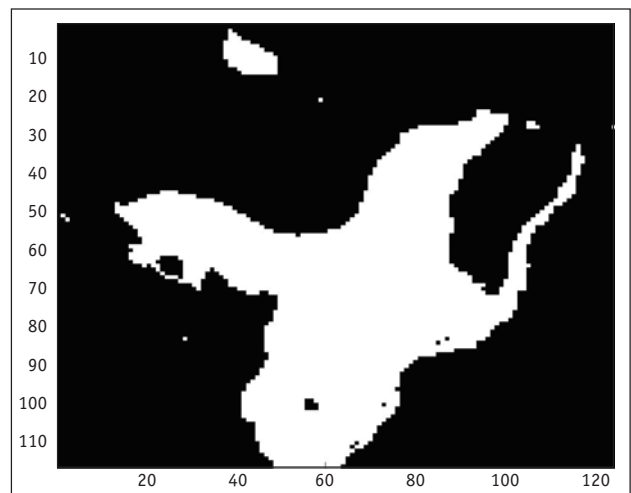


**FIGURA 4** – Predição de ocorrência para *I. virescens*, segundo o modelo GRASP. Áreas claras representam áreas de presença potencial da espécie.

grande parte isto é resultado da grande flexibilidade dos modelos aditivos generalizados, gerada pelo uso de funções não-paramétricas de ordem alta. Isto torna a interpretação dos ajustes como processos causais muito difíceis, podendo preferivelmente ser considerada uma análise exploratória das possíveis relações entre essas variáveis. Em muitas áreas de pesquisa os modelos aditivos generalizados são principalmente considerados como métodos de análise exploratória ou “*data mining*”.

A facilidade de acesso a programas de distribuição livre deve ser considerada uma vantagem dessa abordagem. Espera-se um desenvolvimento de novos aplicativos ou a adição de algumas novas propriedades em modelos como o GRASP, à medida que este ramo de aplicações se desenvolva (Lehmann *et al.*, 2003).

Como quaisquer outros modelos, estes também são sensíveis ao tamanho das amostras. Stockwell & Peterson (2002) demonstraram este efeito sobre vários métodos, incluindo a regressão logística e o GARP, e concluíram que este efeito é grande e que a escolha de quais variáveis ambientais utilizar tem efeito imprevisível sobre a acurácia das estimativas. Nada há que não sugira que o mesmo não ocorra nos modelos aditivos generalizados. Quanto menor a amostragem do padrão de distribuição de uma determinada espécie, maiores podem ser os vícios dos modelos gerados. Espera-se que estes vícios sejam no sentido de prever áreas de distribuição mais restritas, mas a escolha de uma variável incorreta pode superestimar a distribuição de uma espécie. Isto ocorrerá principalmente se outra variável capaz de restringir a distribuição da espécie em questão não for incluída no modelo.



**FIGURA 5** – Predição de ocorrência para *I. macrophylla*, segundo o modelo GRASP. Áreas claras representam áreas de presença potencial da espécie.

Uma das questões centrais é o uso de pseudoausências. Zaniewsky *et al.* (2002) utilizam duas formas distintas de produzir pseudoausências a partir de métodos estatísticos. Seus resultados são bastante razoáveis se comparados com o modelo baseado em dados reais de ausência (produzidos com coletas de campo com esforço padronizado). No entanto, o número de amostras daquele estudo está muito além da maioria dos bancos de dados atuais sobre sistemas amazônicos e, possivelmente, da maioria absoluta dos bancos de dados de distribuição de espécies brasileiras. Este permanece como um ponto sério na utilização destes métodos e sugere que a coleta de informações de campo com esforço padronizado ainda é uma prioridade para a utilização dessa técnica.

Por fim, é preciso salientar que a validação é um ponto crucial na avaliação do sucesso do uso de modelo. A coleta de dados de campo nas áreas onde o modelo prevê e em áreas onde ele não prevê a ocorrência da espécie é vital, tanto para o uso apropriado dos resultados do modelo no futuro, garantindo sua acurácia, quanto para encontrar falhas e, talvez, novas possibilidades de ajuste.

## AGRADECIMENTOS

Os dados utilizados neste trabalho foram gentilmente cedidos pelo Dr. Mike Hopkins, incluindo parte de uma de suas apresentações interativas em um curso de campo. Esse trabalho foi desenvolvido como parte do projeto “Escalas biogeográficas e conservação de espécies: Modelos preditivos para distribuição de espécies e de diversidade alfa e beta”, financiado pelo CNPq, e do projeto “Hotspots de Biodiversidade de Odonata”, financiado pela Fundação “O Boticário” e pelo Instituto de Pesquisas da Mata Atlântica (IPEMA).

## REFERÊNCIAS BIBLIOGRÁFICAS

- Andrewartha, H.G. & L.C. Birch. 1954. The distribution and abundance of animals. University of Chicago Press, Chicago.
- Arita, H.T., J.G. Robinson & K.H. Redford. 1990. Rarity in Neotropical forest mammals and its ecological correlates. *Conservation Biology* 4: 181-192.
- Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200: 1-19.
- Begon, M., J.L. Harper & C.R. Townsend. 1996. *Ecology: individuals, populations and communities*, 3rd edition. Blackwell Science, Oxford.
- Berggren, A. 2001. Colonization success in Roesel's bush-cricket *Metroptera roeseli*: The effects of propagule size. *Ecology* 82: 274-280.
- Bock, C.E. & R.E. Ricklefs. 1983. Range size and local abundance of some North American songbirds: a positive correlation. *The American Naturalist* 122: 295-299.
- Brottons, L., W. Thuiller, M.B. Araujo & A.H. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27: 437-448.
- Brown, J.H. 1984. On the relationship between abundance and distribution of species. *The American Naturalist* 124: 255-279.
- Cole, B.J. 1981. Colonizing abilities, island size, and the number of species on archipelagoes. *The American Naturalist* 117: 629-638.
- Diamond, J.M. 1975. Assembly of species communities. In: M.L. Cody & J.M.C.P. Diamond (eds). *Ecology and evolution in communities*. pp. 342-444. Harvard University Press, Cambridge.
- Gilpin, M.E. & J.M. Diamond. 1981. Immigration and extinction probabilities for individual species: relation to incidence functions and species colonization curves. *Proceedings of National Academy of Sciences of USA* 78: 392-396.
- Grant, P.R. 1966. Ecological compatibility of bird species on islands. *The American Naturalist* 100: 451-462.
- Grant, P.R. & I. Abbott. 1980. Interspecific competition, island biogeography and null hypothesis. *Evolution* 34: 332-341.
- Guisan, A., T.C. Edwards & T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157: 89-100.
- Guisan, A. & W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009.
- Guisan, A. & N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.
- Hastie T. & R. Tibshirani. 1990. *Generalized additive models*. Chapman & Hall, London.
- Hirzel, A.H., V. Helfer & F. Metral. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145: 111-121.
- Leathwick, J.R., J. Elith & T. Hastie. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling* 199: 188-196.
- Lehmann, A., J.M. Overton & J.R. Leathwick. 2002. GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* 157: 189-207.
- Manel, S., J.M. Dias & S.J. Ormerod. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120: 337-347.
- Oksanen, J. & P.R. Minchin. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling* 157: 119-129.
- Ovaskainen, O. & I. Hanski. 2003. The species-area relationship derived from species-specific incidence functions. *Ecology Letters* 6: 903-909.
- Pearce, J. & S. Ferrier. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* 128: 127-147.

- Peterson, A.T. 2001. Predicting species' geographic distributions based on ecological niche modeling. *Condor* 103: 599-605.
- Peterson, A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78: 419-433.
- Phillips, S.J., R.P. Anderson & R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.
- Reader, R.J. 1992. Herbivory, competition, plant mortality and reproduction on a topographic gradient in an abandoned pasture. *Oikos* 65: 414-418.
- Stockwell, D.R.B. & I.R. Noble. 1992. Induction of sets of rules from animal distribution data - A robust and informative method of data-analysis. *Mathematics and Computers in Simulation* 33: 385-390.
- Taylor, B. 1991. Investigating species incidence over habitat fragments of different areas: A look at error estimation. *Biological Journal of the Linnean Society* 42: 177-192.
- Thuiller, W. 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9: 1353-1362.
- Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell & C.E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37: 130-137.
- Wahlberg, N., T. Klemetti & I. Hanski. 2002. Dynamic populations in a dynamic landscape: the metapopulation structure of the marsh fritillary butterfly. *Ecography* 25: 224-232.
- Zaniewski, A.E., A. Lehmann & J.M.C. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157: 261-280.