

Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Programa de Pós-Graduação em Botânica

**Modelando ocorrência e abundância de espécies arbóreas no entorno de
uma usina hidrelétrica no sul do Brasil**

Aluno: Ernestino de Souza Gomes Guarino
Orientador: Jorge Luiz Waechter

Porto Alegre, outubro de 2010

Modelando ocorrência e abundância de espécies arbóreas no entorno de uma usina hidroelétrica no sul do Brasil

Ernestino de Souza Gomes Guarino

Tese apresentada ao Programa de Pós-Graduação em Botânica da Universidade Federal do Rio Grande do Sul como um dos pré-requisitos para a obtenção do título de Doutor em Ciências: Botânica.

Orientador: Jorge Luiz Waechter (UFRGS)

Veni, vidi, vici
Júlio César, 47 a.C.

Dedico este trabalho aos meus pais
Domingos Guarino Neto e Sônia
Gomes Guarino. A distância espacial
nos separa, mas o coração nos une.

AGRADECIMENTOS

Final de jornada é sempre esquisito, um misto de alegria pelo dever cumprido e saudosismo por tudo que aconteceu durante aquele período, mas, principalmente, dúvidas em relação ao que vem pela frente. Confesso que essa foi a jornada mais difícil que enfrentei até o momento, mas com certeza a melhor e mais divertida de todas! Tudo começou em Brasília em 2006, veio para Porto Alegre em 2007, passou por Évora (Portugal) em 2009 e voltou para Porto Alegre para terminar no final de setembro de 2010. Quatro intensos anos que marcaram minha vida para sempre. Como esperado, durante esse período muita gente entrou e saiu da minha vida, porém vou aqui tentar lembrar e agradecer todos que de qualquer forma ajudaram a consolidar mais esse sonho.

Começo agradecendo a base de tudo. Meus pais, Domingos Guarino Neto e Sônia Gomes Guarino e minhas irmãs Sandra e Simone Gomes Guarino. Novamente sem o suporte de vocês nada disso teria acontecido! Muito obrigado pela confiança no meu plano mirabolante de sair de Brasília. Acho que deu certo!

A Embrapa Recursos Genéticos e Biotecnologia pelo apoio logístico e financeiro para a realização do levantamento de campo. Reservo um agradecimento especial aos amigos do antigo Laboratório de Ecologia e Conservação, Aldicir Scariot, Bruno Walter, Daniel Viera, os irmãos Bonesso Sampaio (Alexandre e Maurício), as Béis (Isabel Figueiredo e Isabel Schimdt) e ao Anderson Sevilha. Agradeço aos “guerreiros” Aécio Amaral, Juarez Amaral e Maurício Bonesso Sampaio pelo auxílio constante nas expedições de campo. Enfrentar quatro dias dentro de uma Toyota Bandeirantes apertada (dá-lhe azulona!), frio fora do nosso padrão *cerradense* e sobe e desce morro não é para qualquer mortal! Um grande pedaço disso aqui é de vocês.

Aos amigos de Brasília Daniel Vieira (Butão), Luís Henrique Mantovani (Gugu) e Hélder Consolaro (Japa)! Obrigado pelas conversas quando eu ainda morava em Brasília. A coragem para sair da minha “zona de conforto” veio com certeza da convivência diária com vocês! Aos amigos de Porto Alegre Rogério Enderle e Tiago Azevedo pelo primeiro ano inesquecível! A saudade de casa ficou menor tendo companheiros como vocês! Aos amigos do PPG Botânica e principalmente do Laboratório de Fitoecologia e Fitogeografia da UFRGS,

Fernando Rocha (Caixa d'Água), Rodrigo Leonel (Caixinha), Pedro Rates Vieira, Eduardo Giehl, Guilherme Ceolin, Jair Kray (KRAY!), Jaqueline Durigon, Martin Molz, Stephanie Weege (Susa), Talita Camargo, Mateus Reck e Bianca Spalding.

Aos mestres Jorge Luiz Waechter e João André Jarenkow, que, do jeito deles, formam a cada ano mais e melhores profissionais. Agradecimento especial devo ao meu orientador Jorge Luiz Waechter, que mesmo sem me conhecer aceitou o desafio de me orientar. Obrigado pela confiança!

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de doutorado (processo nº 140712/2007-0) e a CAPES pela concessão da bolsa de doutorado sanduíche (processo nº 2298/09-0). Agradeço também aos doutores Miguel Bastos Araújo e Ana Márcia Barbosa, por abrirem as portas da Cátedra Rui Nabeiro de Biodiversidade, onde passei 6 proveitosos meses aperfeiçoando meus poucos conhecimentos sobre “modelagem de distribuição de espécies”.

Por último agradeço a Ana Simeonidis (Aninha). Aninha mais do que amiga, virou namorada, companheira e porto seguro, que ao acalmar minha mente me deu forças para seguir em frente e terminar essa empreitada.

MUITO OBRIGADO!

SUMÁRIO

1. Introdução Geral	1
RESUMO GERAL	2
GENERAL ABSTRACT	3
Organização geral	4
Introdução	4
A bacia do rio Pelotas	5
Objetivos e Perguntas	6
Breve descrição da área de estudo	7
Espécies estudadas	8
Levantamento de campo	8
Variáveis ambientais, SIG e análise dos dados	8
Referências	9
2. Modelling occurrence and abundance of threatened plant species around a hydroelectric power dam reservoir	13
ABSTRACT	15
Introduction	16
Material & Methods	18
Study area	18
Target species	19
Sampling	19
Environmental variables	20
<i>Topographic variables</i>	20
<i>Soil Variables</i>	21
<i>Current vegetation cover</i>	22
Data analysis	22
<i>Presence-absence models</i>	22
<i>Abundance models</i>	23
Results	25
<i>Presence-absence models</i>	25
<i>Abundance models</i>	26
<i>Environmental correlates of occurrence and abundance</i>	27
Discussion	27
Conclusions and future applications	32
Acknowledgments	34
References	34
Web references	41
Figures captions	42
Figure 1	43
Figure 2	44
Tables	45
Appendices	51
3. Efeitos da detectabilidade no desempenho de modelos de distribuição de espécies de plantas	54

RESUMO	56
ABSTRACT.....	57
Introdução	58
Material & Métodos	59
<i>Área de estudo e espécies estudada</i>	60
<i>Levantamento de campo e manipulação dos dados</i>	61
<i>Variáveis ambientais</i>	61
<i>Modelos de ocorrência</i>	62
<i>Avaliação da precisão dos modelos</i>	62
Resultados	63
Discussão	64
Agradecimentos	68
Referências	69
Legenda das Figuras	73
Figura 1	74
Figura 2	75
Tabelas	76
Conclusão Geral	79

INTRODUÇÃO GERAL

Modelando ocorrência e abundância de espécies arbóreas no entorno de uma usina hidroelétrica no sul do Brasil¹

Aluno: *Ernestino de Souza Gomes Guarino*

Orientador: *Jorge Luiz Waechter*

¹ Tese de doutorado apresentada ao Programa de Pós-graduação em Botânica da Universidade Federal do Rio Grande do Sul.

RESUMO GERAL

Modelos de ocorrência e abundância de espécies são importantes ferramentas para a elaboração de estratégias para a conservação da biodiversidade. Nestes artigos aplicamos, pela primeira vez, modelos de distribuição e abundância no entorno de uma Usina Hidroelétrica no Sul do Brasil, com o objetivo de mostrar possíveis aplicações destas técnicas no planejamento de ações de coleta e conservação, *in situ* e *ex situ*, destas espécies. Nossos objetivos secundários foram (i) modelar a ocorrência e abundância de plantas em um trecho da bacia do rio Pelotas e (ii) verificar o efeito da detectabilidade em modelos de ocorrência. Para tanto dividimos a tese em dois capítulos e cinco questões. No primeiro capítulo procuramos responder as seguintes questões: (1) existe relação entre a probabilidade de ocorrência e a abundância observada? (2) Modelos baseados na abundância são melhores para prever a ocorrência de espécies do que modelos baseados apenas na ocorrência das espécies? No segundo capítulo, tentamos responder se (3) amostragens exaustivas, com presenças e ausências bem descritas, produzem modelos de ocorrência mais precisos? (4) Se a relação entre a ocorrência da espécie e as variáveis ambientais que descrevem sua distribuição é alterada de acordo com a detectabilidade? (5) Qual o impacto de falsos zeros na área ocupada estimada para a espécie? O trabalho foi realizado no entorno da UHE Barra Grande, erguida no rio Pelotas, entre os Estados de Santa Catarina e Rio Grande do Sul. Localizada na bacia do rio Pelotas, seu reservatório ocupa uma área de aproximadamente 90 km² e seu entorno perfaz uma área aproximada de 4.600 km², onde amostramos 388 parcelas georreferenciadas (10 x 50 m) distribuídas aleatoriamente. Em cada parcela contamos o número de indivíduos de nove espécies arbórescentes com altura $\geq 1,5$ m. No primeiro artigo, modelamos, utilizando modelos lineares generalizados (GLM), a distribuição e abundância das nove espécies estudadas em relação a 15 variáveis ambientais, e demonstramos que modelos de ocorrência são mais precisos do que modelos de abundância. Demonstramos também que a probabilidade de ocorrência é positivamente correlacionada com a abundância observada, o que abre a possibilidade de utilizarmos a probabilidade de ocorrência como um indicador da abundância das espécies. No segundo capítulo, verificamos o impacto de falhas de detecção em modelos de distribuição de espécies, utilizando como espécie modelo *A. angustifolia*. Simulamos o impacto de falsos zeros na precisão dos modelos, na seleção das variáveis resposta (variáveis ambientais) e nas estimativas de área ocupada pela espécie. A redução da detectabilidade ocasionou alterações tanto na precisão dos modelos quanto na área estimada de ocorrência e nas variáveis ambientais que explicam a ocorrência de *A. angustifolia*. A área ocupada não apresentou padrão bem definido, porém mostrou tendência de elevação de acordo com a diminuição da detectabilidade, elevando com isso o erro de comissão dos modelos. A precisão dos modelos (AUC e correlação entre probabilidade de ocorrência e abundância observada), apresentou queda em relação ao redução da detectabilidade, enquanto apenas uma variável ambiental foi incluída em todos os modelos (pH do solo). Existe um grande possibilidade de uso destas técnicas como informação básica para planos de conservação, monitoramento e manejo da biodiversidade no entorno de empreendimentos hidroelétricos, porém, para que possamos tirar inferências fortes sobre as espécies e ecossistemas em questão, devemos observar sempre a necessidade de levar em conta os impactos da detecção imperfeita das espécies.

Palavras-chave: rio Pelotas, conservação *in situ*, conservação *ex situ*, falsos negativos, precisão, GLM, área ocupada.

GENERAL ABSTRACT

Species occurrence and abundance models are important tools in the development of biodiversity conservation strategies. This study presents, for the first time, the application of distribution and abundance models in the environment surrounding a hydropower dam in southern Brazil. The aim is to demonstrate the possible application of these techniques in planning collection and conservation strategies of the species inhabiting the area, *in situ* and *ex situ*. We also aimed at (i) modeling the occurrence and abundance of plants in a section of the Pelotas River basin and (ii) verifying the effect of detectability in occurrence models. This thesis was therefore divided into two chapters and five questions. The first chapter is an attempt to answer the following questions: (1) Is there a relationship between occurrence probability and abundance? (2) Are models based on abundance better to predict the occurrence of species as compared to models that utilize only species occurrence? The second chapter addresses the questions: (3) Do exhaustive sampling procedures, with well-sampled presences and absences, lead to more accurate occurrence models? (4) Does detectability influence the relationship between species occurrence and environmental variables? (5) What is the impact of false zeros in the estimated area occupied by one given species? This study was conducted in the area surrounding the hydropower plant “Barra Grande”, in the Pelotas River, on the border of the Rio Grande do Sul and Santa Catarina states. Located in the Pelotas River basin, the dam occupies an area of approximately 90 km². The surrounding area accounts for 4,600 km², where 388 plots with 10 x 50 m were randomly distributed and georeferenced. In each plot, individuals of nine selected tree species taller than 1.5 m were surveyed. In the first paper, we modeled the distribution and abundance of the species studied using generalized linear models (GLMs), utilizing 15 environmental variables. Occurrence models were proved to be more accurate than abundance models. The positive correlation between occurrence probability and observed abundance was also demonstrated, which opens the possibility to utilize occurrence probability as an indicator of species abundance. The second chapter reports the impact of failures in detection in species distribution models, using *A. angustifolia* as model species. We simulated the impact of false zeros in model accuracy, in the selection of response variables (environmental variables) and in estimates of the area occupied by *A. angustifolia*. The decrease in detectability caused changes in both model accuracy and in the environmental variables that explain the occurrence of *A. angustifolia*. The area occupied did not have a well-defined pattern, though it showed a trend towards elevation, in accordance with the decrease in detectability, thus raising the commission error in the model. The accuracy of models (AUC and correlation between occurrence probability and observed abundance) decreased with the decrease in detectability, while only one environmental variable was included in all models (soil pH). These techniques offer great potential in the collection of essential information in biodiversity conservation, monitoring and managements strategies in hydropower projects. However, robust inferences about species and ecosystems in question can only be drawn considering the need to evaluate the impact caused by defective species detection.

Keywords: Pelotas River; *in situ* and *ex situ* conservation; false negatives; accuracy; GLM; occupied area.

Organização geral da tese

A presente tese está organizada em uma introdução geral e dois capítulos independentes, redigidos de acordo com as normas dos periódicos *Ecological Modelling* e *Journal of Applied Ecology*, respectivamente capítulos 1 e 2.

No primeiro capítulo modelamos a ocorrência e abundância de sete espécies de árvores e três arborescentes (duas palmeiras e uma samambaia arborescente) no entorno da UHE Barra Grande. Nosso objetivo foi demonstrar a viabilidade do uso desta ferramenta, tendo como objetivo a aplicação destas informações em planos de coleta de germoplasma e a conservação de recursos genéticos *in situ* e *ex situ* em outras situações semelhantes. No capítulo 2 verificamos qual o impacto de falhas na detecção de uma espécie (*Araucaria angustifolia*) em diferentes parâmetros dos modelos de distribuição: precisão, seleção das variáveis independentes e nas estimativas área ocupada. Nosso objetivo foi demonstrar qual o impacto da detectabilidade, fator geralmente negligenciado, na modelagem de distribuição de espécies.

Introdução

Até o início do século XXI, mais de 34.000 km² de terras foram inundados no Brasil para a formação de reservatórios para usinas hidrelétricas (Bermann 2001). Durante vários anos foi dada atenção apenas ao resgate da fauna ameaçada com o enchimento dos reservatórios, negando a importância do resgate da flora ameaçada por estes empreendimentos (Walter & Cavalcanti 2005). Até o início da década de 1990, foram realizadas várias tentativas de resgatar a flora ameaçada por grandes empreendimentos hidroelétricos no Brasil, porém direcionados a algumas espécies (Walter & Cavalcanti 2005). O projeto pioneiro que demandou resgatar amostras representativas da flora

ameaçada por um empreendimento hidroelétrico foi realizado no início da década de 1990, no Aproveitamento Hidroelétrico Serra da Mesa (GO), quando a Embrapa Recursos Genéticos e Biotecnologia executou o resgate e a conservação *ex situ* da flora ameaçada pelo empreendimento. Diferentemente da conservação *ex situ*, em que a conservação dos componentes da diversidade biológica é feita fora de seu habitat natural, a conservação *in situ* preconiza a manutenção e restituição de populações viáveis das espécies em seu ambiente natural, permitindo a continuidade dos processos de evolução nas populações conservadas (Prance 1997).

Apesar da grande importância para o manejo da biodiversidade, dados de distribuição espacial das espécies são geralmente de baixa qualidade e difíceis de obter (Austin & Meyers 1996). Em decorrência dos recentes avanços tecnológicos ocorridos tanto na estatística espacial, quanto nos processos de generalização de dados com a disseminação do uso de Sistemas de Informações Geográficas (SIG), modelos estatísticos que correlacionam a ocorrência de espécies com a distribuição espacial de diferentes características ambientais se tornaram ferramentas fundamentais para o manejo de ecossistemas (Guisan & Zimmermann 2000, Barbosa et al. 2003, 2009), porém atualmente cresceu o interesse no uso potencial de modelos capazes de prever a abundância ou densidade e não apenas a ocorrência para o mapeamento e delineamento de habitats de alta qualidade indicados para a conservação (Pearce & Ferrier 2001), sendo esse um critério meramente intuitivo, onde quanto maior o número de indivíduos de uma espécie em uma reserva, maior a probabilidade de sobrevivência da mesma (Araújo & Williams 2000).

A bacia do rio Pelotas

Divisa natural entre os Estados de Santa Catarina e Rio Grande do Sul, o rio Pelotas nasce dentro dos limites do Parque Nacional de São Joaquim (SC), e é o principal tributário do rio

Uruguai. Os primeiros sinais de colonização nesta região datam do meio do século XVII e início do século XVIII, com a chegada das primeiras bandeiras ao “*Sertão das Lagens*” em Santa Catarina e aos “Campos de Cima da Serra” no Rio Grande do Sul (Ribeiro & Pozenato 2005). Os mais de 300 anos de colonização, com a exploração e o uso do solo para a agricultura, geraram um padrão de fragmentação florestal muito comum em todo o bioma Floresta Atlântica (SOS Mata Atlântica & INPE 2002) e, mais especificamente na região do Planalto Meridional brasileiro, que tem como fisionomia particular à floresta ombrófila mista (Veloso et al. 1991).

A bacia do rio Pelotas é uma das 147 áreas prioritárias para a conservação da flora da Mata Atlântica, sendo uma das 79 áreas de extrema importância biológica (MMA 2000). Nesta região encontram-se dois dos mais ameaçados ecossistemas deste bioma. O primeiro e mais conhecido, a floresta ombrófila mista (floresta com araucária) hoje ocupa menos de 5% de sua área original. O segundo são as florestas estacionais decíduais, concentradas nas margens do rio Pelotas e alvo de grandes empreendimentos hidrelétricos, sendo um dos últimos, a Usina Hidroelétrica (UHE) de Barra Grande. Devido à grande importância biológica da bacia do rio Pelotas, a construção desta hidroelétrica demandou a conservação *ex situ* das espécies de plantas atingidas pela construção do reservatório, a indicação de áreas mais apropriadas para conservação *in situ* no entorno do reservatório e para realocação de amostras das espécies resgatadas.

Objetivos e perguntas

Nossos objetivos foram (i) modelar a ocorrência e abundância de plantas em um trecho da bacia do rio Pelotas e (ii) verificar o efeito da detectabilidade em modelos de ocorrência. Para tanto dividimos a tese em dois capítulos e cinco questões. No primeiro capítulo tentamos responder as seguintes questões: **(1)** existe relação entre a probabilidade de ocorrência e a abundância? **(2)**

Modelos baseados na abundância são mais eficientes para prever a ocorrência de espécies do que modelos baseados apenas na ocorrência das espécies? No segundo capítulo, tentamos responder se (3) amostragens exaustivas, com presenças e ausências bem descritas, produzem modelos de ocorrência mais precisos? (4) A relação entre a ocorrência da espécie e as variáveis ambientais, que descrevem sua distribuição, é alterada de acordo com a detectabilidade? (5) Qual o impacto de falsos zeros na área ocupada estimada para uma determinada espécie?

Breve descrição da área de estudo

A barragem da Usina Hidroelétrica de Barra Grande (Figura 1) foi erguida entre os Estados de Santa Catarina e Rio Grande do Sul, sul do Brasil. Localizada na bacia do rio Pelotas, seu reservatório ocupa uma área de aproximadamente 90 km² e seu entorno, delimitado pelos contornos da bacia hidrográfica do rio Pelotas, perfaz uma área aproximada de 4.600 km². A altitude varia entre 500 e 1000 m a.n.m. O clima na região é mesotérmico, subtropical úmido (Cfb na classificação de Köppen), com precipitação anual de 1.412 mm e temperatura média anual de 15,2 °C (Maluf, 2000). O relevo varia de suave-ondulado na região do planalto meridional brasileiro, onde predominam os Cambissolos e a Terra Bruna Estruturada, a acidentado nas margens do rio Pelotas, onde as principais classes de solo são os Litólicos, com encaves de Terra Bruna Estruturada (Potter et al. 2004, sensu Sistema Brasileiro de Classificação do Solo - SiBCS). A vegetação é caracterizada pela ocorrência de áreas contínuas de floresta estacional decidual localizadas, predominantemente, nas margens do rio Pelotas. Nas áreas mais altas (≥ 800 m), localizadas no planalto meridional brasileiro, ocorre a floresta ombrófila mista (floresta com araucária), sendo esta fisionomia naturalmente fragmentada pela ocorrência de campos naturais (Klein 1978; Veloso et al. 1991).

Espécies estudadas

Araucaria angustifolia (Bertol.) Kuntze, *Butia eriospatha* (Mart. ex Drude) Becc., *Clethra scabra* Pers., *Dicksonia sellowiana* (Presl.) Hook., *Erythrina falcata* Benth., *Maytenus ilicifolia* (Schrad.) Planch., *Myrocarpus frondosus* Allemão, *Podocarpus lambertii* Klotzsch ex Endl. e *Trithrinax brasiliensis* Mart.. Com exceção de *E. falcata* todas as demais espécies apresentam diferentes riscos de extinção segundo a Lista Vermelha da International Union for Conservation of Nature (IUCN: <http://www.iucnredlist.org>) e/ou a lista de espécies ameaçadas do estado do Rio Grande do Sul (Decreto do Estado do Rio Grande do Sul nº 41.672 de 11 de junho de 2002). Devido à ampla exploração comercial da samambaia arborescente *D. sellowiana*, a mesma tem seu comércio regulado pelo Apêndice II da Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES: <http://www.cites.org>).

Levantamento de campo

Foram amostradas de forma aleatória 388 parcelas georreferenciadas (10 x 50 m) no entorno do reservatório da UHE Barra Grande. Em cada parcela contamos o número de indivíduos das espécies estudadas com altura $\geq 1,5$ m. Para prevenir ou reduzir o impacto da topografia nas amostras, as parcelas foram alocadas ao longo das linhas de contorno do relevo, similar ao descrito por Magnusson et al. (2005) para levantamentos rápidos da biodiversidade.

Variáveis ambientais, SIG e análise dos dados

A fim de conhecer a influência dos fatores ambientais na distribuição espacial e abundância das espécies-alvo na área estudada, amostras aleatórias de solo foram coletadas em todas as parcelas (Lemos & Santos 1996), e analisadas nos laboratórios da Embrapa Cerrados (CPAC, Planaltina/DF). Utilizando o algoritmo Spline com tensão regularizada (RST; Mitasova & Hofierka 1993; Mitasova & Mitas 1993), os dados das variáveis químicas e físicas do solo foram interpolados, gerando superfícies contínuas de dados. Estas variáveis, juntamente com a altitude, o grau de inclinação do relevo e o aspecto do terreno, obtidas do modelo digital de elevação, obtida a partir de imagens do sensor ASTER GDEM e o Índice de Vegetação por Diferença Normalizada, obtida a partir da divisão das bandas 3 (vermelho visível) e 4 (infravermelho próximo) do satélite Landsat 5 TM, foram utilizadas como variáveis preditivas. A junção de todas estas informações foi realizada por meio da criação de camadas contínuas de informação (*layers*) em um Sistema de Informações Geográficas (SIG), implementado no programa gratuito QuantumGIS (QGIS; disponível em <http://www.qgis.org>) e analisadas posteriormente utilizando modelos lineares generalizados (GLM) no programa R (disponível em <http://cran.r-project.org>).

Referências

- Araujo, M.B.; Williams, P.H. (1999). Selecting areas for species persistence using occurrence data. **Biological Conservation** **96**: 331-345.
- Austin, M.P.; Meyers, J.A. (1996). Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. **Forest Ecology and Management** **85**: 95-106.
- Barbosa, A.M.; Real, R.; Olivero, J.; Vargas, J.M. (2003). Otter (*Lutra lutra*) distribution modeling at two resolution scales suited to conservation planning in the Iberian Peninsula. **Biological Conservation** **114**: 377-387.
- Barbosa, A.M.; Real, R.; Vargas, J. (2009). Transferability of environmental favourability models in geographic space: The case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. **Ecological Modelling** **220**: 747-754.

- Bermann, C. (2001). Hidrelétricas: águas para a vida, não para a morte! **Tempo e Presença** 317: 10 –13.
- Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). **Appendix II**. Lista de espécies não necessariamente ameaçadas de extinção atualmente, porém podem se tornar caso o mercado das mesmas não seja controlado. Disponível em: <<http://www.cites.org>>. Acesso em: 15 de janeiro de 2009.
- Guisan, A.; Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. **Ecological Modelling** 135: 147-186.
- International Union for Conservation of Nature and Natural Resource (IUCN). **2006 IUCN Red list of threatened species**. Lista de espécies extintas e ameaçadas de extinção. Disponível em: <<http://www.iucnredlist.org>>. Acesso em: 04 de setembro de 2006.
- Klein, R.M. (1978). **Mapa fitogeográfico do estado de Santa Catarina**. Itajaí, SC: SUDESUL/FATMA/HBR. 24p.
- Lemos, R.C.; Santos, R.D.(1996). **Manual de descrição e coleta de solos no campo**. 3.ed. Campinas, SP: Sociedade Brasileira de Ciência do Solo. 84p.
- Magnusson, W. E., Lima, A. P., Luizão, R., Luizão, F., Costa, F. R. C., Castilho, C. V. d. & Kinupp, V. F. (2005). RAPELD: a modification of the Gentry method for biodiversity surveys in long-term ecological research sites. **Biota Neotropica** 5: 19-24.
- Maluf, J. R. T. (2000) Nova classificação climática do Estado do Rio Grande do Sul. **Revista Brasileira de Agrometeorologia**, 8, 141-150.
- Mitasova, H.; Hofierka, J. (1993) Interpolation by Regularized Spline with Tension:II. Application to Terrain Modeling and Surface Geometry Analysis. **Mathematical Geology** 25: 656-669.
- Mitasova, H.; Mitas, L. (1993) Interpolation by regularized Spline with tension: I . Theory and Implementation. **Mathematical Geology** 25: 641-655.
- MMA (2000). **Avaliação e ações prioritárias para a conservação da biodiversidade da Mata Atlântica e Campos Sulinos**. Brasília: MMA/SBF. 40p.
- Pearce, J.; Ferrier, S. (2001). The practical value of modelling relative abundance of species for regional conservation planning: a case study. **Biological Conservation**: 98: 33-43.
- Potter, R. O.; Carvalho, A.P.; Flores, C.A.; Bognola, I. (2004). **Solos do Estado de Santa Catarina**. Rio de Janeiro: Embrapa Solos. 726p. (Boletim de Desenvolvimento, 46)

- Prance, G.T. (1997). **The conservation of botanical diversity**. In: Maxted, N.; Ford-Lloyd, B.V.; Hawkes, J.G. (Eds.) **Plant genetic conservation: the *in situ* approach**. Londres, Inglaterra: Chapman & Hall. 445p.
- Quantum GIS Development Team (2010). Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. Disponível em: <http://qgis.osgeo.org>
- R Development Core Team (2009). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org>.
- Ribeiro, C.M.P.J.; Pozenat, J.C. (2005). **Fronteiras sem divisas: aspectos históricos e culturais da área da Usina Hidrelétrica Barra Grande**. Caxias do Sul, RS: EDUCS. 375p.
- SOS Mata Atlântica; Instituto Nacional de Pesquisas Espaciais (2002). Atlas dos Remanescentes florestais da Mata Atlântica: período 1995-2000. **Relatório Final**. Disponível em formato PDF em: <http://www.sosmataatlantica.org.br>. Acesso em 08 de setembro de 2006.
- Veloso, H. P.; Rangel Filho, A. L. R.; Lima, J. C. A. (1991). **Classificação da vegetação brasileira adaptada a um sistema universal**. São Paulo, SP: Fundação Instituto Brasileiro de Geografia e Estatística – IBGE. 123 p.
- Walter, B.M.T.; Cavalcanti, T.B. (2005). **Resgate e conservação da flora vascular em aproveitamentos hidrelétricos: exemplos na região do Cerrado**. In: Walter, B.M.T.; Cavalcanti, T.B. (Eds.) **Fundamentos para a coleta de germoplasma vegetal**. Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia. 778p.

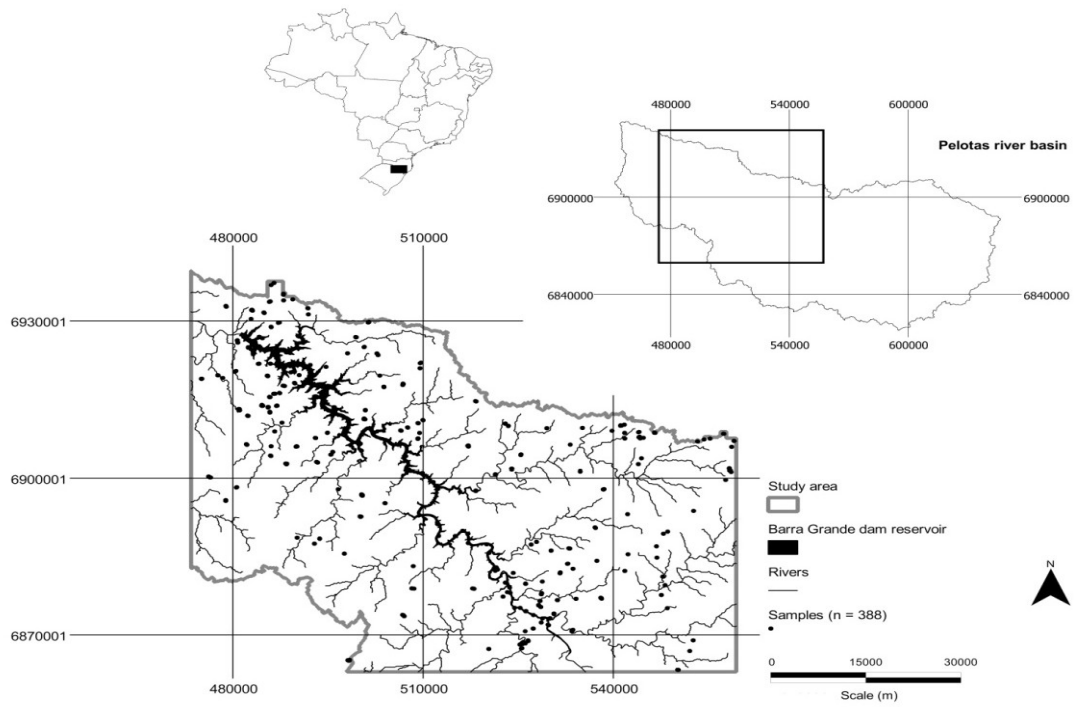


Figura 1. Localização geográfica da área de estudo e das unidades amostrais (sistema de coordenadas: Universal Transverse Mercator, Zone 22J, Hemisfério Sul, Brasil, entre os Estados de Santa Catarina e Rio Grande do Sul).

CAPÍTULO 1

Modelling occurrence and abundance of threatened plant species around a hydroelectric power dam reservoir¹

*Ernestino de Souza Gomes Guarino
Ana Márcia Barbosa
Jorge Luiz Waechter*

¹ Manuscrito escrito de acordo com as normas do periódico *Ecological Modelling*. Fator de impacto em 2009: **1.871** (© Thomson Reuters Journals Reports 2010).

Modelling occurrence and abundance of threatened plant species around a hydroelectric power dam reservoir

E. S.G. Guarino^{a,*}, A.M. Barbosa^b, J.L. Waechter^a

^a *Programa de Pós-Graduação em Botânica, Universidade Federal do Rio Grande do Sul. Avenida Bento Gonçalves, 9500, Prédio 43432.2, Bloco 4, Sala 203, Campus do Vale Bairro Agronomia, CEP 91501-970, Porto Alegre, Brazil*

^b *Cátedra "Rui Nabeiro" Biodiversidade, CIBIO - Universidade de Évora, Largo dos Colegiais, 7004-516, Évora, Portugal and Department of Biological Sciences, Imperial College London, Silwood Park Campus, Ascot (Berkshire) SL5 7PY, United Kingdom*

* Corresponding author. Tel.: +555192273184. *E-mail address*: esguarino@gmail.com (E.S.G. Guarino).

ABSTRACT

Species occurrence and abundance models are important tools that can be used in biodiversity conservation, and can be applied to predict or plan actions needed to mitigate the environmental impacts of hydropower dams. In this study our objectives were to model the occurrence and abundance of threatened plant species, to verify the relationship between predicted occurrence and true abundance, and to assess whether models based on abundance are more effective in predicting species occurrence than those based on presence-absence data. Individual representatives of nine species were counted within 388 randomly georeferenced plots (10x50m) around the Barra Grande hydropower dam reservoir in southern Brazil. We modelled their relationship with 15 environmental variables using both occurrence (GLM) and abundance data (Hurdle and Zero-Inflated Models). Overall, occurrence models were more accurate than abundance models. For all species, observed abundance was correlated with the probability of occurrence, suggesting that future studies could potentially use the probability of occurrence measure instead of the abundance measure. Reconstructing species abundance patterns from occurrence could be an important tool for conservation planning and the management of threatened species, allowing scientists to indicate the best areas for collection and reintroduction of plant germplasm or choose conservation areas most likely to maintain viable populations.

Key-words: species reintroduction, *in situ* and *ex situ* plant conservation, environmental assessment, model validation.

1. Introduction

According to statistics from the Brazilian Electricity Regulatory Agency (ANEEL, 2008), more than 150 hydropower plants were in operation in Brazil as of 2008, amounting to an operating capacity in excess of 74 GW. Hydroelectric energy accounts for more than 70% of the electricity consumed in Brazil, and this percentage is likely to increase because Brazil possesses the world's largest hydroelectric potential, with more than 260 GW available. Some estimates indicate that by the end of the twentieth century, more than 30,000 km² will have been flooded by hydropower dam reservoirs in Brazil. However, few initiatives have been implemented to improve procedures for the mitigation and management of landscapes affected by dam reservoirs.

Due to Brazil's high biodiversity, it is almost impossible to define sampling procedures that take into consideration every organism affected by the development of an artificial reservoir. An insufficient number of field experts, time and money constraints, and large areas covered by hydropower dam reservoirs add to this challenge. Thus, to overcome these obstacles, species distribution models (SDMs) constitute a viable alternative that is effective in predicting the occurrence of different organisms affected by dam construction. The use of these models has greatly increased over the last three decades because of the development and dissemination of Geographic Information Systems (Guisan and Zimmermann, 2000). Today, these models are essential tools within conservation biology (Elith *et al.*, 2006). Using different algorithms, species distribution models correlate species occurrence with environmental data (e.g. climate, soil, topography) to predict species presence on a map (Soberón and Peterson, 2005).

Due to the need to find the highest quality areas for the conservation of biodiversity (Pearce and Ferrier, 2001) and to improve statistical techniques used to predict species abundance, abundance models are today undergoing an expansion similar to that observed in the beginning of

the last decade. A search of papers from indexed journals on the ISI Web of Science (performed on March 5, 2010; [http:// www.isiknowledge.com](http://www.isiknowledge.com)), using the terms *species abundance models* and *species distribution models*, identified over 3,700 papers on this topic published in the last 30 years (1979-2009). More than 50% of these articles were published in the last six years (2004 – 2009), indicating increasing interest in understanding patterns of species abundance.

Information about the abundance of a species provides the following intuitive idea: the greater the number of individuals of a certain species in one area, the greater the probability of maintaining a viable population for that species (Araújo and Williams, 2000).

The relationship between occupancy and abundance is one of the more interesting topics within ecology (Brown, 1984; Gaston *et al.*, 2000; He and Gaston, 2007) and is a topic of constant discussion (Blackburn and Gaston, 2009; Komonen *et al.*, 2009; Kotiaho *et al.*, 2009).

Nevertheless, there are few studies that attempt to recreate abundance patterns using probability of occurrence data generated by species distribution models with different statistical approaches (Jiménez-Valverde *et al.*, 2009; Nielsen *et al.*, 2005; Pearce and Ferrier, 2001; VanDerWal *et al.*, 2009).

Regardless of geographical range or organism type, species distribution models have been successfully used in a variety of conservation biology studies (Cayuela *et al.*, 2009; Rodríguez *et al.*, 2007). Alves and Fontoura (2009) provide an excellent example of the application of species distribution models using fish in a hydrographic basin in southern Brazil. Barbosa *et al.* (2003), Real *et al.* (2009) and Barbosa *et al.* (2010) use these models in studies of otters, lynx and rabbits, and desmans, respectively, in the Iberian Peninsula. Willis *et al.* (2009) uses similar methods in a study on birds in sub-Saharan Africa. Zhu *et al.* (2007) uses these models in a study on invasive plants in China, and Parolo *et al.* (2008) uses them in a study on plant reintroduction in the Alps.

However, species abundance models are rarely used to solve conservation problems, likely because it is difficult to obtain and analyse abundance or density data.

For this study, our objectives were, (i) to model the occurrence and abundance of rare and threatened plant species and verify the relationship between true occurrence and abundance data predicted and predicted occurrence and abundance values, and (ii) to assess if models based on abundance data are better in predicting species occurrence than those based on occurrence data.

2. Material and methods

2.1. Study area

The field survey was conducted around the Barra Grande dam's reservoir in southern Brazil (Fig. 1). Located in the Pelotas River Basin between the states of Santa Catarina and Rio Grande do Sul, this reservoir encompasses approximately 90 km², and its surroundings cover an area of roughly 4,600 km². Elevation ranges between 500 and 1200 m a.s.l., and the climate types range from Cfa (humid subtropical) to Cfb (oceanic or marine temperate), depending on the elevation quota (Köppen climate classification). Annual precipitation is 1.412 mm, and the mean temperature is 15.2°C (Maluf, 2000). Topography varies from rolling highlands (Southern Brazilian Plateau), where cambisol and alfisol are the most common soil classes, to steep slopes near the Pelotas River, where lithosol is the most frequent (Potter *et al.*, 2004; *sensu* Brazilian soil classification system - SiBCS).

The vegetation is characterised by continuous areas of deciduous forests predominantly located close to the Pelotas River. In the highlands (≥ 800 m), Araucaria forests cover a large area that is naturally fragmented by grasslands (Joly *et al.*, 1999; Klein, 1975).

2.2. Target Species

The plant species studied were *Araucaria angustifolia* (Bertol.) Kuntze (Araucariaceae), *Butia eriospatha* (Mart. ex Drude) Becc. (Arecaceae), *Clethra scabra* Pers. (Clethraceae), *Dicksonia sellowiana* (Presl.) Hook. (Dicksoniaceae), *Erythrina falcata* Benth. (Fabaceae), *Maytenus ilicifolia* (Schrad.) Planch. (Celastraceae), *Myrocarpus frondosus* Allemão (Fabaceae), *Podocarpus lambertii* Klotzsch ex Endl. (Podocarpaceae) and *Trinathrinax brasiliensis* Mart (Arecaceae). Except for *E. falcata*, all of these species possess different threat levels, both globally (IUCN Red List) and locally (Rio Grande do Sul State list of threatened species). Due to the intense commercial exploitation of *D. sellowiana*, this fern had its market regulated by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, Appendix II). All of these species are easily identified in the field, reducing sampling errors (false absences).

2.3. Sampling

Past studies have used standardised abundance estimates obtained from different surveys or estimates of abundance from indirect methods (Figueiredo and Grelle, 2009; He and Gaston, 2007). These methods, according to Austin and Meyers (1996), can create an unwanted bias. In contrast, our work was based on reliable occurrence and abundance data. These data were obtained by conducting a detailed field survey and germplasm collection expeditions for the purpose of *ex situ* conservation of the target species. We randomly sampled 388 georeferenced plots (10 x 50 m) around the Barra Grande hydropower dam reservoir. In each plot we counted the number of individuals of each target species with height greater than 1.5 m. To prevent or reduce the influence

of topography in sampling, plots were allocated along topography contour levels, similar to those described by Magnusson *et al.* (2005) for rapid surveys into biodiversity.

To avoid any effects of spatial autocorrelation, the minimum distance between plots was 50 m. The value of this precaution was later validated by Moran's *I* correlograms. All correlograms were calculated using the software SAM v3.1 (Rangel *et al.*, 2010). For all target species, the variation follows a random pattern with a small oscillation along the zero value, which represents the absence of significant autocorrelation (Fig. 2; Fortin and Dale, 2005; Legendre and Legendre, 1998).

2.4. Environmental variables

We used 15 environmental variables (direct and indirect, *sensu* Austin and Myers, 1996) to predict the spatial distribution of the target species (Table 1). All environmental maps had a spatial resolution of 30 m. All spatial data were stored and analysed using the software Quantum GIS/GRASS v.1.5.0 (Quantum GIS Development Team 2010) and its GRASS plugin.

2.4.1. Topographic variables

Elevation values were obtained from a Digital Elevation Model (DEM) generated by the ASTER sensor (ASTER GDEM). From this DEM, three new topographic variables were generated, as described in the discussion that follows. (i) Northness and eastness: the occurrence of different vegetation physiognomies is intimately connected to the amount of available solar radiation. In the southern hemisphere, places with a north-east solar orientation have greater sun exposure and consequently a higher rate of evapotranspiration, resulting in the occurrence of vegetation with

xerophytic characteristics. Conversely, slopes facing south-southeast, especially during the winter, are exposed to less sunlight (Kirkpatrick et al., 1988; Kirkpatrick and Nunez, 1980). To explore this relationship, aspect-modified maps were created in this study. The maps indicate a trend to the north (northness = \cos [aspect]) and east (eastness = \sin [aspect]) (Roberts, 1986). After this transformation, both variables reached values between 1 and -1, indicating a gradient north to south and west to east. (ii) Topographic wetness index (TWI) describes the spatial pattern of soil wetness and is defined as a function of slope. TWI is obtained using the function $\ln(A/\tan\beta)$, where A is the upslope draining through a determined point \times grid cell size and β is the point slope (Neteler and Mitasova, 2008; Sørensen et al., 2005). (iii) Slope was used to indicate soil depth. A higher degree of slope corresponds to a shallower soil depth (Penížek and Borůvka, 2006; Tsai et al., 2001).

2.4.2. Soil variables

Due to the lack of soil maps for the spatial resolution adopted, we conducted soil samples. We collected 381 soil samples (depth: 0 - 20 cm) and analysed their chemical and physical properties (Table 1). The samples were also interpolated using the regularised spline with tension algorithm (RST; Mitasova and Hofierka, 1993; Mitasova and Mitas, 1993). RST is a robust and flexible method used to select parameters that control the properties of the interpolation (tension and smoothing). In addition, estimates generated by RST have an accuracy similar to traditional geostatistical methods (e.g., ordinary and universal kriging; see Chaplot *et al.*, 2006 for more details). The best interpolation control parameter combination was selected iteratively using cross-validation (*leave-one-out* method; Tomczak, 1998). To ensure accuracy, we examined the decrease in root mean square error (RMSE) and mean differences between the observed and predicted values

(the closer to zero, the better the estimate). In addition, our soil maps were validated using other, coarser-scale soil maps available for this region.

2.4.3. *Current Vegetation Cover*

To estimate current vegetation cover, we used the Normalised Difference Vegetation Index (NDVI), obtained by dividing bands 3 (visible red) and 4 (near infrared) of Landsat 5 TM. NDVI values vary between -1.0 and +1.0, and high pixel values represent plentiful vegetation. To avoid any effects of leaf phenology during the year (forest deciduousness), we used Landsat 5 TM images collected during summer in the southern hemisphere.

2.5. *Data analysis*

Except for the Moran's *I* correlograms described above, all other statistical analyses were conducted with R software (v. 2.10.0; R Development Core Team, 2009). A detailed list of analyses and software packages used is available in Appendix A.

2.5.1. *Presence-absence models*

Generalised Linear Models are the most common regression method used to predict species' spatial distribution. According to McCarthy and Elith (2002), GLMs provide a rigorous and statistically robust method to predict the occurrence (or abundance) of species. We modelled the presence-absence of target species with GLM (binomial distribution, logistic link function) using

the 15 environmental variables described in Table 1 as predictors. The predictors were selected using a stepwise procedure, based on Akaike's Information Criterion (AIC; Akaike, 1974)

We evaluated the model's accuracy using the Area under the Receiver Operating Characteristic (ROC) curve (AUC; Fielding and Bell, 1997). AUC values vary between 0 and 1. Values close to 0.5 indicate that model predictions are no better than random, and AUC values equal to 1 indicate a 100% chance for the model to correctly classify an event. The null hypothesis that the area under ROC curve is ≤ 0.5 was tested using a Mann-Whitney U test (Mason and Graham, 2002; function *roc.area* in R verification package v 1.31). Models were validated using abundance data collected in the same sites. Using Spearman's rank correlation (ρ), we tested the relationship between observed species abundance and the predicted probability of occurrence for each locality. In this study, our premise was that if a model correctly predicts abundance, it has a high discrimination capacity (Jiménez-Valverde *et al.*, 2007). To verify whether the probability of occurrence measure could be used as a substitute for species abundance, we also used Spearman's rank correlation, though now we tested the relationship of locations with observed abundance > 0 with predicted probability of occurrence as suggested by Pearce and Ferrier (2001). We also verified the relationship between mean sample size of occupied sites (abundance > 0) and abundance using a Mann-Whitney-Wilcoxon U test (Zar, 1999).

2.5.2. Abundance models

Regression models based on Poisson distributions are often used to analyse count data (such as abundance). However, the data are usually zero-inflated, i.e., the incidence of zeros is larger than expected by chance (Ridout *et al.*, 1998; Welsh *et al.*, 1996). Zuur *et al.* (2009) described five sources of zeros, two related to species occupancy patterns (i.e., the habitat is not suitable and the

species is not present, and the habitat is suitable but is not used by the species) and three related to sampling errors (i.e., design error, species detectability, and sampling outside species habitat range (“*naughty naughts*” *sensu* Austin and Meyers (1996)). Because of the rigorousness of our field survey and the high detectability of our target species, we believe that our samples do not contain false negatives (sampling errors), and the zeros that do occur are only related to patterns of species occupancy.

To model abundance we tested the following four different algorithms: (i) GLM with Poisson (P) and negative binomial (NB) distributions (GLM_p and GLM_{nb}), (ii) Hurdle Models (P and NB, Zeileis et al., 2008) and (iii) Zero-Inflated Count Data Regression (P and NB, ZIP and ZINB, respectively). GLM_p (standard Poisson) is the simplest method used to model count data and assumes equi-dispersion (variance = mean), but this method is not always true in practice due to zero-inflation (variance > mean). One alternative to this method is to include a dispersal parameter (θ) in the Poisson distribution to deal with overdispersion. This generalised Poisson distribution is called a negative binomial (Gonzales-Barron *et al.*, 2010).

Hurdle models are two-component class models capable of accounting for overdispersion using Poisson or negative binomial distributions. As a first step, Hurdle models use a truncated count component for values > 0. As a second step, a hurdle component models zeros vs. non-zeros values using a binomial GLM (Zeileis *et al.*, 2008). Zero-inflated models are two-component mixture models in which zeros are modelled as originating from two stochastic processes, the binomial process and the count process. Due to this assumption, Zero-inflated models do not need to truncate data on zero. Similar to Hurdle models, Zero-Inflated models use binomial GLM to model the probabilities of measuring zeros, and the count process is modelled by a Poisson (ZIP) or negative binomial (ZINB) GLM (Zuur *et al.*, 2009).

We selected the most parsimonious model for each target species using AIC (Akaike, 1974) and evaluated the selected abundance model for the following five parameters: (i) Pearson correlation coefficient (r), which provides an indication of agreement between observed and predicted values (a perfect adjustment ($r = 1$) does not imply an exact prediction); (ii) Spearman's rank correlation (ρ), which varies from -1 to +1 and provides an indication of similarity in rank between observed and predicted abundance values; (iii) linear regression coefficients, which are obtained by fitting a simple linear regression (observed values = $m(\text{predicted values})+b$) (in a perfect calibrated model, m should equal 0 and b should equal 1); and (iv) root mean square error (RMSE) and average error (AVE_{error}), both of which are dependent on sample size and measure divergences between observed and predicted abundance values (Potts and Elith, 2006). Confidence intervals (95%) for each evaluated parameter were calculated using a bootstrap procedure (1,000 replicates).

3. Results

3.1. Presence-absence models

AUC values ranged from 0.71 to 0.97 and in all cases were significantly different from 0.5 (Mann-Whitney U test, $P < 0.01$; for details about environmental variables included in the GLM models see Appendix B). Overall, the relationship between observed abundance and occurrence probability was positive and statistically significant (Table 2). *M. ilicifolia* was the species with the highest correlation between abundance and probability of occurrence ($\rho = 0.47$; $P < 0.01$), whereas *P. lambertii* had the lowest correlation, although it was still statistically significant. When we truncated the abundance data to eliminate zeros, the results were different. Only *A. angustifolia*

and *C. scabra* retained significant correlations between observed abundance and probability of occurrence values (Table 2). For seven species, the relationship between abundance and probability of occurrence was not significant when unoccupied sites were excluded. However, when considering only occupied sites, the mean sample size for these species was not significantly less (Mann-Whitney-Wilcoxon U test, $p = 0.22$).

3.2. Abundance models

Due to the high prevalence of zero counts (98%), we were not able to model the abundance of *B. eriospatha*. Based on AIC values, Hurdle and Zero-inflated models (Poisson or negative binomial distributions) resulted in a better fit than GLM with Poisson or negative binomial distributions (Table 3).

We could clearly distinguish three groups of species using abundance evaluation parameters. (i) *D. sellowiana* and *P. lambertii* showed the worst-fit models with the smallest values of r (≤ 0.09) and ρ , (≤ 0.25), a poor model calibration ($m = 0.07$ and 0.00 ; $b = 1.75$ and 0.30) and the highest values of RMSE and AVE_{error} . (ii) *E. falcata*, *M. ilicifolia*, *M. frondosus* and *T. brasiliensis* showed accurate abundance estimates, with better model evaluation parameters. These species had relatively high values of r (≥ 0.66) and ρ (≥ 0.32), indicating that both observed and predicted abundance measures were similar in magnitude, but not similarly ordered (Potts and Elith, 2006). Also, these species gave the smallest RMSE and AVE_{error} . (iii) *A. angustifolia* and *C. scabra* showed consistent bias, with similar values of r and ρ (between 0.32 and 0.47), indicating that observed and predicted abundance measures were relatively similar in magnitude and similarly ordered (Potts

and Elith, 2006). *A. angustifolia* presented a poor linear model calibration, whereas *C. scabra* presented a better model calibration (Table 3).

3.3. Environmental correlates of occurrence and abundance

Low soil pH increases the abundance of *A. angustifolia*, *C. scabra*, *D. sellowiana* and *M. frondosus*, while high soil pH values increases the abundance of *T. brasiliensis* (Table 4). Soil pH provides an indirect nutrient gradient in the soil, and its values directly affect the importance of K^+ , S^{2-} and P^+ (low or negative β values, Table 4) for each species. The occurrence of *E. falcata* and *M. ilicifolia* was influenced by soil pH. However, neither abundance nor occurrence of *P. lambertii* was affected by soil pH. The abundance of Fabaceae species, such as *E. falcata* and *M. frondosus*, was negatively affected by low amounts of nitrogen in soil, while the abundance of *P. lambertii* was positively affected by high amounts nitrogen in soil (Table 4). Northness, eastness, slope and TWI each had a different influence on the values of abundance and occurrence, occasionally affecting both simultaneously (Table 4). Low values of NDVI, an indirect indicator of the stage of vegetation succession (current vegetation biomass), affected the abundance of *A. angustifolia*, *C. scabra* and *M. frondosus*. Occurrence of *C. scabra* was also associated with low values of NDVI, while *D. sellowiana*, *E. falcata*, *P. lambertii* and *T. brasiliensis* were positively affected by NDVI (Table 4).

4. Discussion

Because of the need for reliable species distribution models to aid in developing conservation strategies, methods used to assess model accuracy are one of the most important issues in this

growing field of study (see e.g. Elith et al., 2006; Fielding and Bell, 1997; Manel et al., 2001). Lobo *et al.* (2008) conducted a detailed description of the different issues involved with the misuse of the AUC, which is widely used as a measure of accuracy of species' spatial distribution models. One alternative proposed by the authors is the use of abundance data to validate species distribution models. This idea is based on an inductive relationship, where probabilities of occurrence are functionally related to species abundance (Nielsen *et al.*, 2005).

Nielsen *et al.* (2005) recommend an approach based on organism life history to explain the relationship between occurrence and abundance patterns, which is different from the traditional approach based on extensive and exploratory fitting exercises. In that study, *A. angustifolia* and *C. scabra* were the only species that indicated a positive and significant correlation between probabilities of occurrence and observed abundance when zero-abundance plots were excluded from the analysis. Both species are tied to early successional stages in the studied region (Duarte *et al.*, 2006; Sampaio and Guarino, 2007), and several studies indicate that *A. angustifolia* is able to advance forest expansion over natural grasslands and colonise new areas of grasslands in the highlands of southern Brazil (Duarte *et al.*, 2006). Furthermore, the seeds of *A. angustifolia* are dispersed by birds and small mammals (Anjos, 1991; Iob and Vieira, 2008), whereas *C. scabra*'s seeds are wind-dispersed and can thus travel greater distances.

Another interesting case is that of subtropical palms, *B. eriospatha* and *T. brasiliensis*, which are gregarious species with dense but sporadic populations in the study area. Because of the importance of fibre and the high nutritional value of the fruits, both species were influenced by ancient human inhabitants of the region. The population of these species was purposefully dispersed by natives and European settlers, who grazed livestock for centuries in the area we studied (Henderson *et al.* 1995). Past human-mediated dispersion can affect the current patterns of plant distribution, creating artificially clustered populations that are often unnoticed in extensive

exploratory model-fitting exercises. Although this history of plant manipulation is important, rarely has it been included in studies of species distribution (Lutolf *et al.*, 2009).

Pearce and Ferrier (2001) correlated the probability of occurrence with abundance, but with a different purpose. They attempted to use predicted probabilities of occurrence as a surrogate method to predicting species abundance. However, according to these authors, the relationship between the probabilities generated by linear models and observed abundance is weak and restricted to a few species. Similar results have been reported by Nielsen *et al.* (2005) and Jiménez-Valverde *et al.* (2009). In both studies, the probability of occurrence is not correlated with abundance when points with zero abundance are excluded from analysis. Pearce and Ferrier (2001) suggest that the overall correlation between occupied sites (abundance > 0) and probability of occurrence is due mostly to the difference in mean predicted probability between occupied and unoccupied sites. However, the fact that this correlation often disappears when unoccupied sites are excluded may be due to a reduction in sample size and the consequent loss of analytical power. In Pearce and Ferrier's 2001 study, for 59 species, the relationship between abundance and probability of occurrence was no longer significant when unoccupied sites were excluded (mean sample size = 42.98, standard deviation = 38.43). In addition, the mean sample size considering only occupied sites was significantly lower (Mann-Whitney-Wilcoxon U test, $P < 0.001$) than for 13 species for which this correlation remained significant (mean sample size = 111.38, standard deviation = 47.31). These results indicate that sample size may influence the detection of this relationship. However, this pattern is opposite to what Jiménez-Valverde *et al.* (2009) described in their paper, and is also contrary to what we observed in this study.

Real *et al.* (2009) indicated that predictions of GLMs based on the presence-absence data of the Iberian lynx (*Lynx pardinus*) and wild rabbit (*Oryctolagus cuniculus*) in Spain are significantly correlated with independent abundance data. Even when we repeated their analysis and excluded

the zero-abundance data, the correlations remained significant for the abundance of rabbit (Kendall's $Tau-b = 0.100$, $P = 0.036$, $n = 300$) and lynx in 1950 (there were no zero-abundance data for this year), 1965 ($Tau-b = 0.108$, $P = 0.005$, $n = 355$), and 1975 ($Tau-b = 0.086$, $P = 0.035$, $n = 315$). Only on one occasion (lynx data from 1985) was the correlation no longer significant between predicted favourability and observed abundance, possibly due to the lower sample size ($n = 215$) when zero-abundance cells were excluded for this year. This corroborates the idea that, given sufficient sample size, predicted probability of occurrence may, in general, be a good replacement for actual species abundance.

Reconstructing or inferring abundance from species occurrence data can potentially be an important tool for conservation planning efforts and species management, especially given the difficulty of obtaining and analyzing abundance data. Current techniques are often not reliable because they generate unstable, low-quality results (spurious estimates; Joseph *et al.*, 2009). Like other studies, our analysis showed that the relationship between abundance and predicted probability of occurrence does not exist for all species. Jiménez-Valverde *et al.* (2009) suggest that generalist species and those with a high degree of dispersal show positive and significant relationships between abundance (> 0) and the probability of occurrence. However, unlike Pearce and Ferrier (2001), Nielsen *et al.* (2005), and Jiménez-Valverde *et al.* (2009), we believe that, given sufficient sample sizes, probability of occurrence measures can not only be an indicator of model accuracy, as suggested by Albert & Thuiller (2008) and Lobo *et al.* (2008), but also provide a simple and inexpensive alternative for the species abundance measure. This alternative method also has great potential to be used in environmental assessment (see also Araújo and Williams, 2000; Real *et al.*, 2009).

The relatively low degree of correlation between predicted occurrence probabilities and observed abundance (< 0.5 in all cases) could be due to biotic interactions not included in the

occurrence models. Austin *et al.* (1990) and Guisan and Thuiller (2005), for example, suggested that the response curve of a species along an environmental gradient can be seriously constrained by interaction with biotic factors. This hypothesis was recently tested by Heikkinen *et al.* (2007) and Ritchie *et al.* (2009), using different organisms, and they both found similar results. When data related to interspecific competitors were incorporated into models, species predictions were significantly improved. In addition, this effect likely plays a role in the relationship between occurrence and abundance, obscuring the true relationship between occupied area and abundance. Ritchie *et al.* (2009) confirmed this idea, demonstrating that the predicted abundance of wallaroos and kangaroos was improved with the addition of the occurrence and abundance of interspecific competitors into the models. However, obtaining these data is difficult and model practitioners must continue to use mostly abiotic factors (Barbosa *et al.*, 2009; Elith and Leathwick, 2009).

In general, Zero-Inflated Models performed better than Hurdle Models. This result is contrary to other empirical studies on both real populations (Potts and Elith, 2006) and theoretical studies with pseudo-populations (Miller and Miller, 2008), suggesting that Hurdle Models generally perform better. Because the techniques used in this study to model abundance deal differently with zeros, the results can be interpreted in several ways (Potts and Elith, 2006; Zuur *et al.*, 2009). Two-part modelling techniques, such as GLM Hurdle models, analyse abundance data in the following two steps: (i) zero vs. non-zero values are modelled with a logistic regression (binomial distribution), and (ii) non-zero observations are modelled with a truncated Poisson or negative-binomial regression. Mixed techniques, such as Zero-Inflated models, classify zeros as originating from two different processes, binomial and count processes (Zuur *et al.*, 2009). Interpreting Hurdle models is simpler than interpreting Zero-Inflated models (Potts and Elith, 2006), but, according to Welsh *et al.* (2000), mixed models provide a better tool when there is overdispersion and a large number of true zeros, as was the case with our data.

Our main objective with fitted models was to predict species' distribution and abundance rather than to test the effects of different ecological drivers on species occurrence and abundance. However, we can draw some conclusions about the ecological factors that seem to influence the analysed species' distributions. According to Barrows *et al.* (2005), environmental managers must understand the anthropogenic and environmental factors that influence the occurrence and abundance of species. With this information, managers can employ adaptive management strategies to maintain viable populations of desired species. Overall, occurrence and abundance of studied species are determined, to some extent, by different suites of environmental variables. Regardless of organism type, this pattern has been widely reported (Heinänen *et al.*, 2008; Illan *et al.*, 2010; Truscott *et al.*, 2008), suggesting that biotic and abiotic events associated with plant establishment may be different than those influencing species abundance (Truscott *et al.*, 2008).

Here we highlight the relationship between the abundance of *E. falcata* and total nitrogen availability. Abundance (count model component) of *E. falcata* was negatively affected by low soil nitrogen values. This species, along with others of the same genus, are generally able to produce nitrogen-fixing root nodules (Faria *et al.*, 1984; Faria *et al.*, 1989; Schimann *et al.*, 2008). However, its roots have a unique strategy to acquire soil nitrogen. The roots do not exploit available nitrogen in the soil, and instead use nitrogen in the leaf litter (Chesney and Vasquez, 2007; Payan *et al.*, 2009 to *E. poeppigiana* in agroforestry systems), an environmental variable that was analysed in our study. On the other hand, *M. frondosus*, another Fabaceae species, were not reported to produce nitrogen-fixing nodules (Allen and Allen, 1981; Faria *et al.*, 1984).

5. Conclusions and future applications

Our study is the first to apply a macroecological approach to support *in situ* and *ex situ* plant conservation practices in a region affected by a hydropower dam reservoir. Currently in Brazil, Environmental Impact Assessment (EIA) and Strategic Environmental Assessment (SEA) methods require costly and time-consuming field surveys (floristic and phytosociological studies), often conducted by professionals with an incomplete understanding of the extreme diversity of Brazilian flora. As a result, these surveys often produce low-quality, inapplicable data that are susceptible to environmental frauds. Despite methodological limitations, i.e., low correlation between predicted probability of occurrence and observed abundance, we believe that EIA/SEA projects in Brazil should consider using species distribution models to predict future impacts and plan landscape management strategies and mitigation actions. This approach is quicker and less expensive than the current approach, although this method does require staff training in data analysis and data collection.

An effective way to organise data collection is to combine the *gradsect* method proposed by Austin and Heyligers (1989) and Wessels *et al.* (1998) with the rapid assessment survey method used in this study and described by Magnusson *et al.* (2005). For example, instead of measuring the height and diameter of each tree species within plots, we record only the presence of each species in georeferenced plots. Austin and Heyligers (1989) and Austin and Meyers (1996) support the view that for most environmental goals, including species niche modelling, representative samples are more important than accurate basal area estimations, which is the method currently used in the preparation of EIA/SEA in Brazil. Even in a situation where field surveys are not performed, presence-only data obtained from herbarium or museum collections are viable alternatives. Even with coarse resolution, they can produce pertinent information for the development of fine-scale species conservation planning (Araújo, 2004; Barbosa *et al.*, 2010).

Species distribution models could also potentially guide field expeditions to collect germplasms of rare and endangered species. A concrete example of this application was described by de Siqueira et al. (2009). The authors used a mixture of environmental distance indices, field surveys, and species distributions to find, with great success, new populations of a poorly-understood plant species on the Central Brazil Savanna.

However, the mere occurrence of one species is not sufficient to ensure the persistence of a viable population within a system of protected areas (Barrows *et al.*, 2005). It will contain samples of the largest possible number of local ecosystems, thereby maintaining the ecological processes of these ecosystems (Australia, 1997). An alternative to accomplish this goal would be to integrate information about species occurrence (or abundance) and regional land use in an interactive decision-making system. This method would ensure the conservation of viable samples in different ecosystems occurring in the region of interest (Ferrier *et al.*, 2002).

Acknowledgments

The field work was supported by Embrapa Recursos Genéticos e Biotecnologia. E.S.G.G. was supported by Brazilian National Council for Scientific and Technological Development (CNPq; proc. 140712/2007-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; proc. 2298/09-0) scholarships. A.M.B. was supported by a Post-Doctoral Fellowship (SFRH/BPD/40387/2007) from Fundação para a Ciência e a Tecnologia (Portugal). We are grateful to AA Santos, MB Sampaio, JA Santos and others who helped with field work. MB Araújo, DLM Vieira, AB Sampaio, FS Rocha, PR Vieira and RLL Orihuela helped with important suggestions.

References

- Akaike H. A new look at the statistical model identification. *IEEE T Automat Contr* 1974; 19:716-23, 1974.
- Albert CH, Thuiller W. Favourability functions versus probability of presence: advantages and misuses. *Ecography* 2008; 31:417-422.

- Allen ON, Allen EK. The Leguminosae: A source book of characteristics, uses, and nodulation. 1st ed. Madison: The University of Wisconsin Press; 1981
- Alves TP, Fontoura NF. Statistical distribution models for migratory fish in Jacui basin, South Brazil. *Neotrop Ichthyol* 2009; 7:647-658.
- Anjos L. O ciclo anual de *Cyanocorax caeruleus* em floresta de araucária (Passeriformes: Corvidae). *Ararajuba* 1991; 2:19-23.
- Araújo MB. Matching species with reserves - uncertainties from using data at different resolutions. *Biol Conserv* 2004; 118:533-538.
- Araújo MB, Williams, PH. Selecting areas for species persistence using occurrence data. *Biol Conserv* 2000; 96:331-345.
- Austin MP, Heyligers, PC. Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species. *Biol Conserv* 1989; 50:13-32.
- Austin MP, Meyers JA. Current approaches to modelling the environmental niche of eucalypts: Implication for management of forest biodiversity. *Forest Ecol Manag* 1996; 85:95-106.
- Austin MP, Nicholls AO, Margules CR. Measurement of the Realized Qualitative Niche: Environmental Niches of Five *Eucalyptus* Species. *Ecol Monogr* 1990; 60:161-177.
- Australia C.o. Nationally Agreed Criteria for the Establishment of a Comprehensive, Adequate and Representative Reserve System for Forest in Australia. A Report by the Joint ANZECC/MCFFA National Forest Policy Statement Implementation Sub-committee; 1997.
- Barbosa AM, Real R, Olivero J, Vargas JM. Otter (*Lutra lutra*) distribution modelling at two resolution scales suited to conservation planning in the Iberian Peninsula. *Biol Conserv* 2003; 114:377-387.
- Barbosa AM, Real R, Vargas, JM. Transferability of environmental favourability models in geographic space: The case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. *Ecol Model* 2009; 220:747-754.
- Barbosa AM, Real R, Vargas JM. Use of coarse-resolution models of species' distributions to guide local conservation inferences. *Conserv Biol* 2010; early view (doi 10.1111/j.1523-1739.2010.01517.x).
- Barrows CW, Swartz MB, Hodges WL, Allen MF, Rotenberry JT, Li BL, Scott TA, Chen XW. A framework for monitoring multiple-species conservation plans. *J Wildlife Manage* 2005; 69:1333-1345.

- Blackburn TM, Gaston KJ. Sometimes the obvious answer is the right one: a response to 'Missing the rarest: is the positive interspecific abundance-distribution relationship a truly general macroecological pattern?'. *Biol Lett* 2009; 5:777-778.
- Brown JH. On the relationship between abundance and distribution of species. *Am Nat* 1984; 124:255-279.
- Cayuela L, Golicher DJ, Newton AC, Kolb M, Albuquerque FSd, Arets EJMM, Alkemade JRM, Perez AM. Species distribution in the tropics: problems, potentialities, and the role of biological data for effective species conservation. *Trop Conserv Sc* 2009; 2:319-352.
- Chaplot V, Darboux F, Bourennane H, Leguedois S, Silvera N, Phachomphon K. Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. *Geomorphology* 2006; 77:126-141.
- Chesney P, Vasquez N. Dynamics of non-structural carbohydrate reserves in pruned *Erythrina poeppigiana* and *Gliricidia sepium* trees. *Agroforest Syst* 2007; 69:89-105.
- de Siqueira MF, Durigan G, Junior PM, Peterson AT. Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *J Nat Conserv* 2009; 17:25-32.
- Duarte LDS, Machado RE, Hartz SM, Pillar VD. What saplings can tell us about forest expansion over natural grasslands. *J Veg Sci* 2006; 17:799-808.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 2006; 29:129-151.
- Elith J, Leathwick JR. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu Rev Ecol Evol S* 2009; 40:677-697.
- Faria BSM, Franco AA, Jesus RM, Menandro MS, Baitello JB, Mucci ESF, Döbereiner J, Sprent JI. New nodulating legume trees from south-east Brazil. *New Phytol* 1984; 98:317-328.
- Faria BSM, Lewis GP, Sprent JI, Sutherland JM. Occurrence of nodulation in the Leguminosae. *New Phytol* 1989; 111:607-619.
- Ferrier S, Watson G, Pearce J, Drielsma M. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiver Conserv* 2002; 11:2275-2307.

- Fielding AH, Bell JF. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 1997; 24:38-49.
- Figueiredo MSL, Grelle CEV. Predicting global abundance of a threatened species from its occurrence: implications for conservation planning. *Divers Distrib* 2009; 15:117-121.
- Fortin MJ, Dale M. *Spatial Analysis: a guide for ecologists*. 1st ed. Cambridge: Cambridge University Press; 2005.
- Gaston KJ, Blackburn TM, Greenwood JJD, Gregory RD, Quinn RM, Lawton JH. Abundance-occupancy relationships. *J Appl Ecol* 2000; 37:39-59.
- Gonzales-Barron U, Kerr M, Sheridan JJ, Butter F. Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. *Int J Food Microbiol* 2010; 136:268-277.
- Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 2005; 8:993-1009.
- Guisan A, Zimmermann NE. Predictive habitat distribution models in ecology. *Ecol Model* 2000; 135:147-186.
- He F, Gaston KJ. Estimating abundance from occurrence: An underdetermined problem. *Am Nat* 2007; 170:655-659.
- Heikkinen RK, Luoto M, Virkkala R, Pearson RG, Korber JH. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecol Biogeogr* 2007; 16:754-763.
- Heinänen S, Ronka M, von Numers M. Modelling the occurrence and abundance of a colonial species, the arctic tern *Sterna paradisaea* in the archipelago of SW Finland. *Ecography* 2008; 31:601-611.
- Illán JG, Gutierrez D, Wilson RJ. The contributions of topoclimate and land cover to species distributions and abundance: fine-resolution tests for a mountain butterfly fauna. *Global Ecol Biogeogr* 2010; 19:159-173.
- Iob G, Vieira EM. Seed predation of *Araucaria angustifolia* (Araucariaceae) in the Brazilian Araucaria Forest: influence of deposition site and comparative role of small and 'large' mammals. *Plant Ecol* 2008; 198:185-196.
- Jiménez-Valverde A, Diniz F, de Azevedo EB, Borges PAV. Species distribution models do not account for abundance: the case of arthropods on Terceira Island. *Ann Zool Fenn* 2009; 46:451-464.
- Jiménez-Valverde A, Ortuno VM, Lobo JM. Exploring the distribution of *Sterocorax ortuno*, 1990 (Coleoptera, Carabidae) species in the Iberian peninsula. *J Biogeogr* 2007; 34:1426-1438.

- Joly CA, Aidar MPM, Klink CA, McGrath DG, Moreira AG, Moutinho P, Nepstad DC, Oliveira AA, Pott A, Rodal MJN, Sampaio EVSB. Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. *Cienc Cult* 1999; 51:331-348.
- Joseph LN, Elkin C, Martin TG, Possingham HP. Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. *Ecol Appl* 2009; 19:631-642.
- Kirkpatrick JB, Fensham RJ, Nunez M, Bowman DMJS. Vegetation-radiation relationships in the wet-dry tropics: granite hills in northern Australia. *Vegetatio* 1988; 76:103-112.
- Kirkpatrick JB, Nunez M. Vegetation-radiation relationships in mountainous terrain: eucalypt-dominated vegetation in the Risdon Hills, Tasmania. *J Biogeogr* 1980; 7:197-208.
- Klein RM. Southern Brazilian phytogeographic features and the probable influence of upper quaternary climatic changes in the floristic distribution. *Bol Parana Geocienc* 1975; 33:67-88.
- Komonen A, Paivinen J, Kotiaho JS. Missing the rarest: is the positive interspecific abundance-distribution relationship a truly general macroecological pattern? *Biol Lett* 2009; 5:492-494.
- Kotiaho JS, Komonen A, Paivinen J. On the obvious positive interspecific relationship between abundance and distribution: a reply to Blackburn and Gaston. *Biol Lett* 2009; 5:779-780.
- Legendre P, Legendre L. *Numerical Ecology*. 2nd ed. Amsterdam: Elsevier Science B.V.; 1998.
- Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol Biogeogr* 2008; 17:145-151.
- Lutolf M, Guisan A, Kienast F. History Matters: Relating Land-Use Change to Butterfly Species Occurrence. *Environ Manage* 2009; 43:436-446.
- Magnusson WE, Lima AP, Luizão R, Costa FRC, Castilho CV, Knupp VF. RAPELD, a modification of the Gentry method for biodiversity surveys in long-term ecological research sites. *Biot Neotrop* 2005; 5:19-24.
- Maluf JRT. Nova classificação climática do Estado do Rio Grande do Sul. *Rev Bras Agrometeorol* 2000; 8:141-150.
- Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *J Appl Ecol* 38, 921-931.
- Mason, S.J., Graham, N.E., 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q J Roy Meteor Soc* 128, 2145-2166.
- McCarthy MA, Elith J. Species mapping for conservation. *Gap Analysis Bull.* 2002; 2:50-57.

- Miller JM, Miller MD. No Zero Left Behind: Comparing the Fit for Zero-Inflation Models as a Function of Skew and Proportion of Zeros. *InterStat* 11 2008; 35.
- Mitasova H, Hofierka J. Interpolation by Regularized Spline with Tension:II. Application to Terrain Modeling and SurfaceGeometry Analysis. *Math Geo* 1993; 25:656-669.
- Mitasova H, Mitas, L. Interpolation by regularized Spline with tension: I . Theory and Implementation. *Math Geo* 1993; 25:641-655.
- Neteler M, Mitasova H. *Open Source GIS: A GRASS GIS Approach*. 3rd ed. New York: Springer; 2008.
- Nielsen SE, Johnson CJ, Heard DC, Boyce MS. Can models of presence-absence be used to scale abundance? - Two case studies considering extremes in life history. *Ecography* 2005; 28, 197-208.
- Parolo G, Rossi G, Ferrarini A. Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. *J Appl Ecol* 2008; 45:1410-1418.
- Payan F, Jones DL, Beer J, Harmand JM. Soil characteristics below *Erythrina poeppigiana* in organic and conventional Costa Rican coffee plantations. *Agroforest Syst* 2009; 76:81-93.
- Pearce J, Ferrier S. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biol Conserv* 2001; 98:33-43.
- Penížek V, Borůvka L. Soil depth prediction supported by primary terrain attributes: a comparison of methods. *Plant Soil Environ* 2006; 52:424-430.
- Potter RO, Carvalho AM, Flores CA, Bognola I. Solos do Estado de Santa Catarina. *Bol. Pesquisa*, 46. Embrapa Solos; 2004.
- Potts JM, Elith J, 2006. Comparing species abundance models. *Ecol Model* 1999, 153-163.
- Rangel TF, Diniz JAF, Bini LM. SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography* 2010; 33:46-50.
- Real R, Barbosa AM, Rodriguez A, Garcia FJ, Vargas JM, Palomo LJ, Delibes M. Conservation biogeography of ecologically interacting species: the case of the Iberian lynx and the European rabbit. *Divers Distrib* 2009; 15:390-400.
- Henderson A, Galeano G, Bernal R. *Field Guide to the Palms of the Americas*. 1st ed. New Jersey: Princeton University Press, Princeton; 1995.
- Ridout M, Demétrio CGB, Hinde J. Models for count data with many zeros. In, *International Biometric Conference* 1998, Cape Town; 1998. p. 1-13.

- Ritchie, EG, Martin JK, Johnson CN, Fox BJ. Separating the influences of environment and species interactions on patterns of distribution and abundance: competition between large herbivores. *J Anim Ecol* 2009; 78:724-731.
- Roberts DW. Ordination on the basis of fuzzy set theory. *Plant Ecol* 1986; 66:123-131.
- Rodríguez JP, Brotons L, Bustamante J, Seoane J. The application of predictive modelling of species distribution to biodiversity conservation. *Divers Distrib* 2007; 13:243-251.
- Sampaio MB, Guarino, ESG. Efeitos do pastoreio na estrutura populacional de plantas em fragmentos de Floresta Ombrófila Mista. *Rev Árvore* 2007; 31:1035-1046.
- Schimann H, Ponton S, Hattenschwiler S, Ferry B, Lensi R, Domenach AM, Roggy JC. Differing nitrogen use strategies of two tropical rainforest late successional tree species in French Guiana: Evidence from N-15 natural abundance and microbial activities. *Soil Biology Biochem* 2008; 40:487-494.
- Soberón J, Peterson AT. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Inform* 2005; 2:1-10.
- Sørensen R, Zinko U, Seibert J. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrol Earth Syst Sc* 2005; 2:1807-1834.
- Tomczak M. Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach. *J Geogr Inform Decis Anal* 1998; 2:18-30.
- Truscott AM, Palmer SCF, Soulsby C, Hulme PE. Assessing the vulnerability of riparian vegetation to invasion by *Mimulus guttatus*: relative importance of biotic and abiotic variables in determining species occurrence and abundance. *Divers Distrib* 2008; 14:412-421.
- Tsai C, Chen Z, Duh C, Horng F. Prediction of Soil Depth Using a Soil-Landscape Regression Model: a Case Study on Forest Soils in Southern Taiwan. *Proc Natl Sci Council Repub China [B]* 2001; 25:34-39.
- VanDerWal J, Shoo LP, Johnson CN, Williams SE. Abundance and the Environmental Niche: Environmental Suitability Estimated from Niche Models Predicts the Upper Limit of Local Abundance. *Am Nat* 2009; 174:282-291.
- Welsh AH, Cunningham RB, Chambers RL. Methodology for estimating the abundance of rare animals: Seabird nesting on North East Herald Cay. *Biometrics* 2000; 56:22-30.
- Welsh AH, Cunningham RB, Donnelly CF, Lindenmayer DB. Modelling the abundance of rare species: Statistical models for counts with extra zeros. *Ecol Model* 1996; 88:297-308.

- Wessels KJ, Van Jaarsveld AS, Grimbeek JD, Van der Linde MJ. An evaluation of the gradsect biological survey method. *Biodivers Conserv* 1998; 7:1093-1121.
- Willis SG, Hole DG, Collingham YC, Hilton G, Rahbek C, Huntley B. Assessing the Impacts of Future Climate Change on Protected Area Networks: A Method to Simulate Individual Species' Responses. *Environ Manage* 2009; 43:836-845.
- Zar JH. *Biostatistical Analysis*. 4th ed. Prentice Hall. 1999.
- Zeileis A, Kleiber C, Jackman S. Regression Models for Count Data in R. *J Stat Softw* 2008; 27:1-25.
- Zhu L, Sun OJ, Sang WG, Li ZY, Ma KP. Predicting the spatial distribution of an invasive plant species (*Eupatorium adenophorum*) in China. *Landscape Ecol* 2007; 22:1143-1154.
- Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM. *Mixed Effects Models and Extensions in Ecology with R*. 1st ed. Springer; 2009.

Web references

- ANEEL, 2008. Atlas de energia elétrica do Brasil. Brazilian Electricity Regulatory Agency, Brasília. Available from <http://www.aneel.gov.br/biblioteca/EdicaoLivros2009atlas.cfm> (accessed September 2009).
- ASTER GDEM, ASTER Global Digital Elevation Model. Available from <http://www.gdem.aster.ersdac.or.jp> (accessed September 2009).
- CITES, Convention on International Trade in Endangered Species of Wild Fauna and Flora. Available from <http://www.cites.org> (accessed November 2008).
- IUCN, International Union for Conservation of Nature, International Red List. Available from www.iucnredlist.org (accessed November 2008).
- Quantum GIS Development Team. 2009. Quantum GIS. Geospatial Foundation, Vancouver, British Columbia. Available from <http://qgis.osgeo.org> (accessed September 2009).
- R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org> (accessed November 2009).
- Rio Grande do Sul State list of threatened species, Rio Grande do Sul State law n° 41.672, June 11, 2002. Available from <http://www.al.rs.gov.br> (accessed September 2009).

Figure captions

Fig 1. Location of the study area and sampling sites (Universal Transverse Mercator coordinate system, Zone 22J, south hemisphere, southern Brazil).

Fig 2. Morans' *I* correlograms for (a) *A. angustifolia*, (b) *B. eriospatha*, (c) *C. scabra*, (d) *D. sellowiana*, (e) *E. falcata*, (f) *M. ilicifolia*, (g) *Myrocarpus frondosus*, (h) *P. lambertii* and (i) *T. brasiliensis*.

Figure 1.

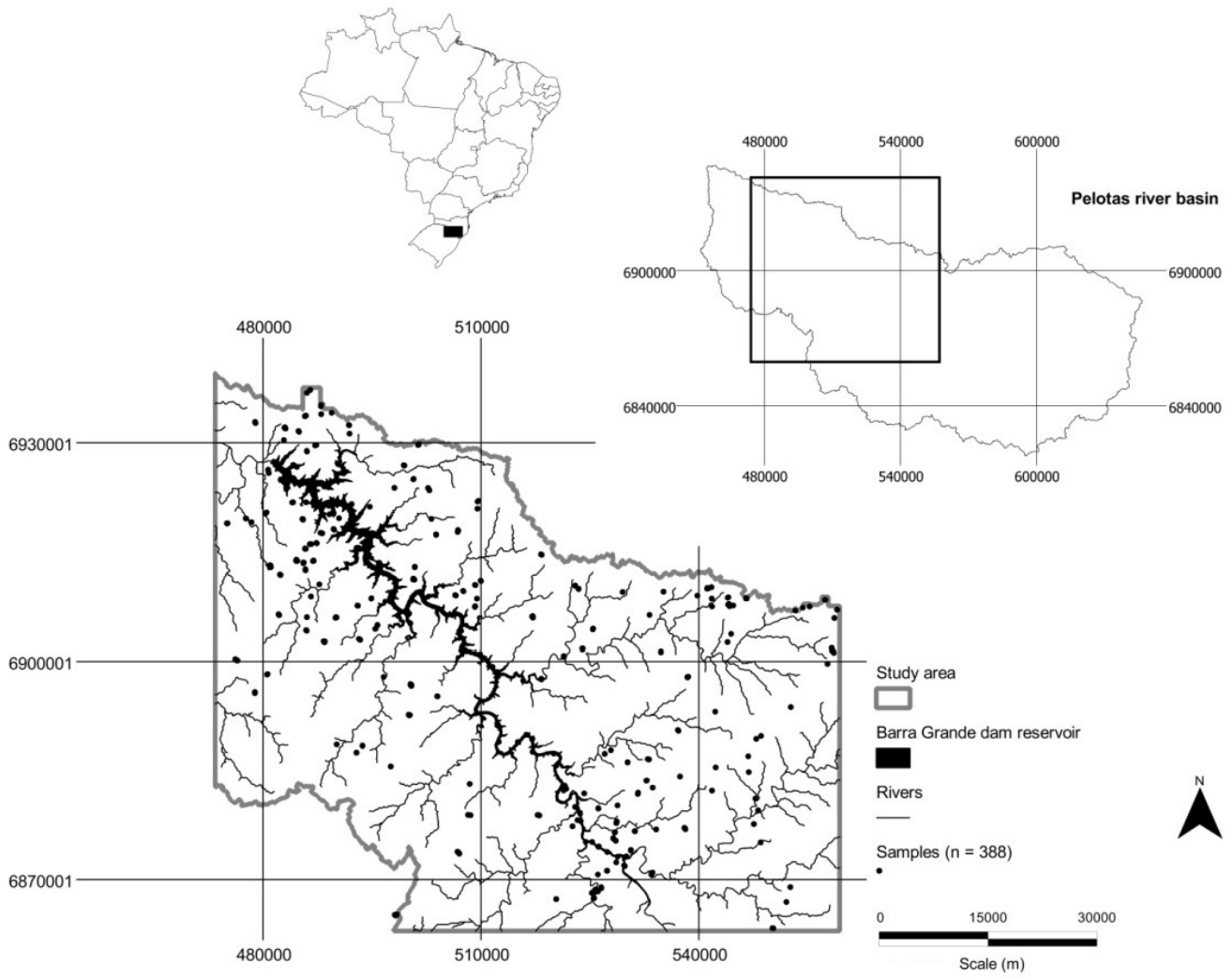


Figure 2.

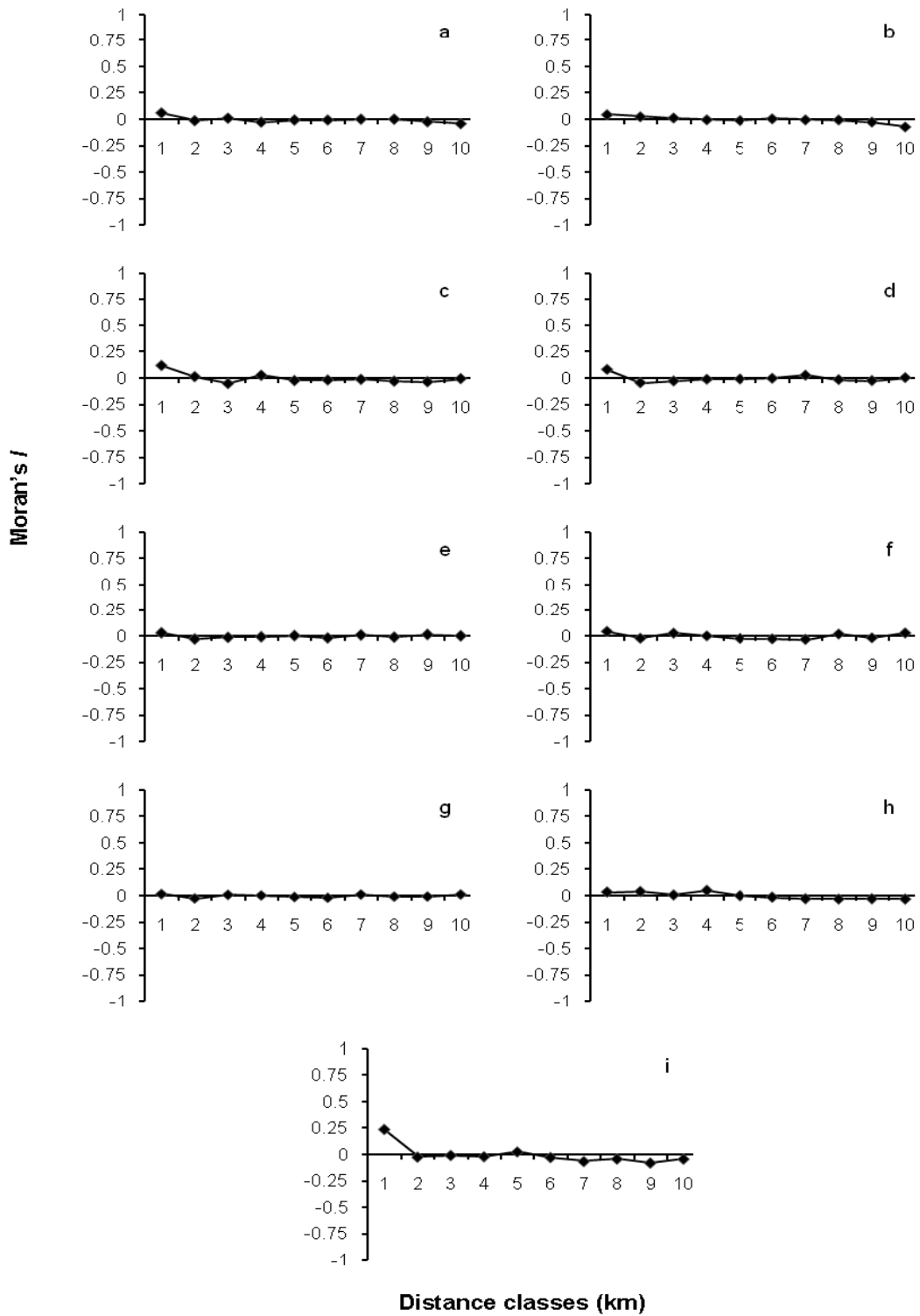


Table 1

Environmental variables used to predict the spatial distribution of the target species. * Variables excluded from analysis due high Pearson correlation coefficient (r) with one or more environmental variable ($r \geq 0.7$; $P < 0.01$).

Group	Variable
Soil	pH H ₂ O
	Ca ⁺² +Mg ⁺² *
	H ⁺ +Al ⁻³ *
	K ⁺
	S ⁻²
	P ⁺
	N total
	Silty:Clay*
	Bulk density
	Topography
Northness	
Eastness	
Topographical wetness index (TWI)	
Slope (degrees)	
Current Vegetation	Normalized Difference Vegetation Index (NDVI)

Table 2

Presence-absence model evaluation. Abundance: median and range, P (%): (Prevalence = [true positives + false negatives]/sample size)*100; *rho*: Spearman's rank correlation (* $P < 0.001$), AUC: area under the receiver operator characteristic curve (Mann-Whitney U test, ** $P < 0.001$).

Target species	Abundance	P (%)	<i>rho</i>	<i>rho</i> (Abd > 0)	AUC
<i>A. angustifolia</i>	4 (0-49)	72	0.42*	0.17*	0.77**
<i>B. eriospatha</i>	0 (0-24)	2	0.21*	0.40	0.95**
<i>C. scabra</i>	0 (0-27)	10	0.30*	0.31*	0.79**
<i>D. sellowiana</i>	0 (0-153)	13	0.25*	0.02	0.71**
<i>E. falcata</i>	0 (0-4)	5	0.23*	0.16	0.83**
<i>M. ilicifolia</i>	0 (0-14)	21	0.47*	0.08	0.84**
<i>M. frondosus</i>	0 (0-29)	5	0.32*	0.11	0.93**
<i>P. lambertii</i>	0 (0-24)	6	0.18*	-0.19	0.73**
<i>T. brasiliensis</i>	0 (0-21)	4	0.32*	0.02	0.97**

Table 3

Abundance selected models based on AIC and evaluation parameters (95% CI). *r*: Pearson correlation coefficient, *rho*: Spearman's rank correlation, *b*: intercept and *m*: gradient of the fitted line (observed = $b + m(\text{predicted})$), RMSE: Root mean square error, AVE_{error}: Average error.

Target species	Model	AIC (d.f.)	<i>r</i>	<i>rho</i>	<i>b</i>	<i>m</i>	RMSE	AVE_{error}
<i>A. angustifolia</i>	ZINB	2149.99 (25)	0.39 (0.31 - 0.47)	0.47 (0.39 - 0.55)	0.95 (-0.24 - 2.16)	0.85 (0.63 - 1.06)	8.00 (6.90-9.12)	0.12 (-0.68 - 0.92)
<i>C. scabra</i>	ZINB	380.72 (25)	0.37 (0.11 - 0.61)	0.32 (0.24 - 0.40)	0.09 (-0.11 - 0.33)	0.53 (0.02 - 0.99)	2.23 (1.41 - 3.12)	0.19 (-0.03 - 0.41)
<i>D. sellowiana</i>	Hurdle _p	170.88 (24)	0.09 (-0.07 - 0.24)	0.12 (0.03 - 0.22)	1.75 (0.90 - 2.76)	0.07 (-0.17 - 0.26)	18.74 (11.85 - 26.43)	1.67 (-0.12 - 3.49)
<i>E. falcata</i>	ZIP	155.70 (24)	0.75 (0.55-1.01)	0.32 (0.24-0.40)	-0.01 (-0.02 - 0.02)	0.97 (0.63-1.31)	0.29 (0.20-0.38)	0.00 (-0.02-0.03)
<i>M. ilicifolia</i>	Hurdle _{nb}	585.20 (24)	0.66 (0.46 - 0.92)	0.48 (0.40 - 0.56)	-0.03 (-0.19 - 0.11)	1.05 (0.74 - 1.42)	1.19 (0.97 - 1.43)	0.00 (-0.12 - 0.12)
<i>M. frondosus</i>	ZIP	179.81 (24)	0.98 (0.90 - 0.99)	0.36 (0.28 - 0.43)	-0.02 (-0.05 - 0.01)	1.08 (0.97 - 1.22)	0.35 (0.27 - 0.45)	0.00 (-0.03 - 0.04)
<i>P. lambertii</i>	ZIP	243.29 (24)	-0.01 (-0.07 - 0.04)	0.25 (0.16 - 0.33)	0.30 (0.11 - 0.50)	0.00 (-0.15 - 0.13)	34.72 (10.44 - 64.07)	2.99 (-0.57 - 6.42)
<i>T. brasiliensis</i>	ZINB	153.18 (25)	0.82 (0.71 - 0.91)	0.34 (0.26 - 0.43)	0.01 (-0.04 - 0.05)	0.88 (0.51 - 1.22)	1.10 (0.67 - 1.57)	0.03 (-0.08 - 0.15)

Table 4

Variables included in abundance models (count and zero components) for studied species. β : coefficient; S.E.: standard error; $P(z)$: z-statistic test.

Target species	Model component	Variable	β	S.E.	$P(z)$
<i>A. angustifolia</i>	Count	pH H ₂ O	-0.76	0.15	> 0.001
		Eastness	1.83	0.40	> 0.001
		slope	0.04	0.01	> 0.001
	Zero	S ⁻²	-0.20	0.05	> 0.001
		N total	-12.03	3.68	0.001
		Northness	11.72	3.39	> 0.001
		Eastness	6.58	2.33	0.004
		NDVI	-4.80	1.71	0.005
		slope	-0.19	0.08	0.01
		Constant	74.95	17.52	> 0.001
<i>C. scabra</i>	Count	PH H ₂ O	-3.23	0.89	> 0.001
		K ⁺	2.27	0.78	0.003
		S ⁻²	-0.07	0.03	0.01
		P ⁺	0.75	0.22	> 0.001
		Bulk Density	-18.46	5.28	> 0.001
		Northness	4.90	2.27	0.03
		NDVI	-15.41	5.03	0.002
	slope	0.19	0.07	0.007	
	Zero	S ⁻²	-0.16	0.06	0.009
		P ⁺	-1.08	0.37	0.003
TWI		-1.12	0.44	0.01	
NDVI		-19.09	8.56	0.02	

Table 4*Continued.*

Target species	Model component	Variable	β	S.E.	$P(z)$	
<i>D. sellowiana</i>	Count	Constant	-8.23	3.39	0.01	
		PH H ₂ O	-0.51	0.09	> 0.001	
		K ⁺	1.04	0.09	> 0.001	
		S ⁻²	-0.08	0.01	> 0.001	
		P ⁺	0.28	0.04	> 0.001	
		N total	-3.86	0.63	> 0.001	
		Bulk Density	5.56	1.24	> 0.001	
		Northness	-1.28	0.56	0.02	
		Eastness	-1.34	0.44	0.002	
		TWI	-0.19	0.06	0.002	
		NDVI	1.18	0.44	0.008	
		slope	0.09	0.01	> 0.001	
		Zero	K ⁺	-0.58	0.28	0.04
			Northness	-2.75	1.27	0.03
<i>E. falcata</i>	Count	Constant	-66.60	16.40	> 0.001	
		K ⁺	2.66	0.64	> 0.001	
		P ⁺	0.45	0.20	0.02	
		N total	-7.44	3.09	0.01	
		Bulk Density	10.77	4.67	0.02	
		TWI	1.76	0.29	> 0.001	
		NDVI	8.64	1.40	> 0.001	
		slope	0.09	0.04	0.04	
		Zero	PH H ₂ O	2.90	0.43	> 0.001
			S ⁻²	0.08	0.01	> 0.001
<i>M. ilicifolia</i>	Count	Northness	1.80	0.93	0.05	
	Zero	Constant	-21.06	7.21	0.003	
		PH H ₂ O	2.90	0.43	> 0.001	
		S ⁻²	0.08	0.01	> 0.001	

Table 4*Continued.*

Target species	Model component	Variable	β	S.E.	<i>P</i> (z)		
<i>M. frondosus</i>	Count	PH H ₂ O	-0.76	0.15	> 0.001		
		K ⁺	1.83	0.40	> 0.001		
		S ⁻²	-12.03	3.68	0.001		
		P ⁺	11.72	3.39	> 0.001		
		N total	-4.80	1.71	0.005		
		Northness	-0.19	0.08	0.01		
		Eastness	8.24	0.77	> 0.001		
		TWI	0.78	0.17	> 0.001		
		NDVI	-4.38	0.49	> 0.001		
		slope	0.17	0.05	> 0.001		
	Zero		<i>n.s.</i>				
<i>P. lambertii</i>	Count	K ⁺	-1.95	0.67	0.003		
		P ⁺	0.78	0.22	> 0.001		
		N total	7.26	2.87	0.01		
		Northness	11.63	3.07	> 0.001		
		Eastness	-8.64	2.44	> 0.001		
		TWI	0.49	0.09	> 0.001		
		NDVI	10.25	2.51	> 0.001		
		slope	-0.65	0.11	> 0.001		
			Zero	P ⁺	0.93	0.43	0.03
				Northness	14.76	5.68	0.009
		slope	-0.63	0.22	0.005		
<i>T. brasiliensis</i>	Count	PH H ₂ O	7.73	2.48	0.001		
		K ⁺	5.72	1.21	> 0.001		
		S ⁻²	-0.25	0.09	> 0.001		
		TWI	0.91	0.19	> 0.001		
		NDVI	6.11	1.60	> 0.001		
		slope	0.52	0.10	> 0.001		
	Zero		<i>n.s.</i>				

APPENDICES

Appendix A. Statistical analysis and R packages.

Statistical analysis	Model
AIC	stats
AUC (Mann-Whitney U test)	verification
GLM	stats
GLM _{nb}	VGAM
Hurdle, ZIP and ZINB	pscl
Pearson correlation (r)	stats
Spearman's rank correlation (ρ)	stats
stepAIC	MASS

Appendix B. Variables included in the GLM models for each target species (presence-absence models). β : coefficient; S.E.: standard error; $P(z)$: z-statistic test.

Target species	Variable	β	S.E.	$P(z)$
<i>A. angustifolia</i>	Constant	5.74	1.42	> 0.001
	pH H ₂ O	-1.41	0.29	> 0.001
	K ⁺	-0.30	0.17	0.08
	S ⁻²	0.05	0.01	> 0.001
	N total	1.30	2.55	0.01
	NDVI	1.28	0.63	0.04
<i>B. eriospatha</i>	Constant	90.59	30.26	0.002
	K ⁺	2.66	0.89	0.002
	P ⁺	-1.26	0.55	0.02
	N total	-18.50	7.39	0.01
	Bulk density	-34.15	11.19	0.002
	Northness	11.31	6.88	0.100
<i>C. scabra</i>	slope	-0.19	-1.46	0.16
	Constant	33.85	7.59	> 0.001
	pH H ₂ O	-1.58	0.43	> 0.001
<i>D. sellowiana</i>	Bulk density	-11.18	2.99	> 0.001
	K ⁺	-0.64	0.23	0.006
	N total	2.83	1.39	0.041
	Bulk density	-3.66	2.22	0.09
<i>E. falcata</i>	Northness	-2.81	1.15	0.015
	Constant	-14.62	3.30	> 0.001
	pH H ₂ O	1.86	0.71	0.009
	N total	-3.14	1.48	0.03
	Northness	0.41	0.18	0.02
	TWI	0.41	0.18	0.02
<i>M. ilicifolia</i>	slope	0.06	0.04	0.12
	Constant	-17.58	2.05	>0.001
	pH H ₂ O	2.67	0.33	>0.001
	S ⁻²	0.08	0.01	> 0.001

Appendix B.

Continued.

Target species	Variable	β	S.E.	P (z)
<i>M. frondosus</i>	Constant	-19.20	4.01	> 0.001
	pH H ₂ O	2.73	0.70	> 0.001
	K ⁺	0.77	0.27	0.003
	Eastness	3.17	1.06	0.002
<i>P. lambertii</i>	Constant	-23.84	8.45	0.005
	N total	5.85	2.43	0.02
	Bulk density	7.70	3.12	0.01
	slope	-0.13	0.05	0.008
<i>T. brasiliensis</i>	Constant	-29.80	8.09	> 0.001
	pH H ₂ O	5.43	1.58	> 0.001
	S ⁻²	-0.39	0.12	0.001
	P ⁺	1.01	0.29	> 0.001
	N total	4.43	3.19	0.16
	Eastness	4.56	2.56	0.07
	slope	-0.19	0.06	0.02

CAPÍTULO 2

Efeitos da detectabilidade no desempenho de modelos de distribuição de espécies de plantas¹

Ernestino de Souza Gomes Guarino
Jorge Luiz Waechter

¹ Manuscrito escrito de acordo com as normas do periódico *Journal of Applied Ecology*. Fator de impacto em 2009: **4.197** (© Thomson Reuters Journals Reports 2010).

Efeitos da detectabilidade no desempenho de modelos de distribuição de espécies de plantas

E. S.G. Guarino^{a,*} & J.L. Waechter^a

^a. Programa de Pós-Graduação em Botânica, Universidade Federal do Rio Grande do Sul. Avenida Bento Gonçalves, 9500, Prédio 43432.2, Bloco 4, Sala 203, Campus do Vale Bairro Agronomia, CEP 91501-970, Porto Alegre, Brazil

* Autor para correspondência. Tel.: +555192273184. *Endereço de e-mail:* esguarino@gmail.com (E.S.G. Guarino).

RESUMO

1. A qualidade dos dados utilizados em modelos de distribuição de espécies (SDMs) é um dos principais gargalos para a produção de modelos confiáveis. Um dos principais fatores que afetam a qualidade do conjunto de dados e, por consequência, a performance destes modelos é a capacidade de detectar corretamente a ocorrência da espécie em um determinado local.

2. Neste artigo tentamos responder três questões: (i) amostragens exaustivas, com presenças/ausências bem descritas, produzem modelos de ocorrência mais precisos? (ii) A relação entre a ocorrência da espécie e as variáveis ambientais que descrevem sua distribuição é alterada de acordo com a detectabilidade? (iii) Qual o impacto de falsas ausências na área ocupada estimada para a espécie?

3. Realizamos um extenso e detalhado levantamento de *Araucaria angustifolia* no entorno da Usina Hidroelétrica de Barra Grande (SC/RS). Amostramos 388 parcelas de 10 x 50 m, onde registramos a abundância de indivíduos de *A. angustifolia* com altura \geq 1,5 m. Para simular a diminuição na detecção de *A. angustifolia*, gradualmente substituímos classes inteiras de abundância por falsos zeros, até as amostras com 30 ou mais indivíduos (classes de abundância). As classes foram posteriormente convertidas em dados de presença/ausência, e modeladas em relação a variáveis ambientais utilizando modelos lineares generalizados (GLM).

4. A redução da detectabilidade ocasionou alterações tanto na precisão dos modelos quanto na área estimada de ocorrência e nas variáveis ambientais que explicam a ocorrência de *A. angustifolia*. A área ocupada não apresentou padrão bem definido, porém mostrou tendência de elevação de acordo com a diminuição da detectabilidade, elevando com isso o erro de comissão dos modelos. A precisão dos modelos (AUC e correlação entre probabilidade de ocorrência e abundância observada), apresentou queda em relação ao redução da detectabilidade, enquanto apenas uma variável ambiental foi incluída em todos os modelos (pH do solo).

5. O monitoramento da biodiversidade cada vez mais se vale do uso de SDMs, porém, estes geralmente negligenciam ou lidam de forma ineficiente com as falhas de detecção. O resultado desta falha é a produção de mapas de distribuição aparente, e não real das espécies, estudadas, inviabilizando a tomada de inferências corretas sobre estas.

Palavras-chave: *Araucaria angustifolia*, abundância, amostragem, falsas ausências, área ocupada, modelo linear generalizado.

ABSTRACT

1. The quality of data used in species distribution models (SDMs) is one of the main bottlenecks in the development of reliable models. One of the main factors that affect the quality of datasets and therefore the performance of such models is the capacity to correctly detect the occurrence of a given species in a given place.
2. This article attempts to answer three questions: (i) Do exhaustive sampling procedures, with well-defined presences/ absences, lead to more accurate occurrence models? (ii) Does detectability influence the relationship between species occurrence and environmental variables that describe this species occurrence? (iii) What is the impact of false zeros in the estimated area occupied by one given species?
3. We conducted an extensive and detailed survey of the occurrence of *A. angustifolia* in the environment surrounding the hydropower dam “Barra Grande” (SC/RS). In total, 388 plots with 10 x 50 m were randomly distributed and georeferenced. In each plot, *A. angustifolia* individuals taller than 1.5 m were sampled. To simulate the decrease in detection of studied species, whole abundance classes were gradually replaced by false zeros, until samples containing 30 or more individuals (abundance classes). Classes were later converted into presence/absence data, and modeled according to the environmental variables using generalized linear models (GLM).
4. The decrease in detectability led to changes both in accuracy of models and area estimated for occurrence, and in the environmental variables that explain the occurrence of *A. angustifolia*. The area occupied did not have a well-defined pattern, though it showed a trend towards elevation, in accordance with the decrease in detectability, thus raising the commission error in the model. The accuracy of models (AUC and correlation between occurrence probability and observed abundance) decreased with the decrease in detectability, while only one environmental variable was included in all models (soil pH).
5. The monitoring of biodiversity increasingly utilizes SDMs, though these in general neglect or inefficiently address failures in detection. The result of this failure is the production of apparent, not real, distribution of the species studied, which renders impossible to correctly infer species distribution.

Keywords: *Araucaria angustifolia*; abundance; sampling; false zeros; area occupied; generalized linear model.

Introdução

A qualidade dos dados utilizados na modelagem da distribuição de espécies (SDMs) é um dos principais gargalos para a produção de modelos confiáveis (Guisan et al., 2007; Cayuela et al., 2009). Um dos principais fatores que afetam a qualidade do conjunto de dados e, por consequência, a performance de SDMs é a capacidade de detectar corretamente a ocorrência de uma espécie em um determinado local (Phillips et al., 2006). Isso significa que ausências registradas podem indicar tanto uma ausência real ou, alternativamente, a não detecção da espécie apesar de sua ocorrência no local (falsas ausências; Chen et al., 2009). Diversos são os fatores que podem afetar a detecção de uma espécie vegetal, como, por exemplo, visibilidade, a arquitetura e forma de vida da planta, comprimento e frequência da floração e frutificação, diferenças entre observadores, esforço de amostragem, abundância da espécie, eficiência da amostragem e tamanho das parcelas (Gu & Swihart, 2004; Kéry et al., 2006; Kéry et al., 2010; Chen, 2009).

Podemos definir detectabilidade como a probabilidade de se encontrar um determinado táxon em uma parcela ou levantamento, isso significa que apenas uma fração (p) de todos os indivíduos ou espécies será detectada após a amostragem de uma determinada área (Tyre et al., 2003; Kéry & Schmidt, 2008). Em geral, assume-se que todas as espécies têm a mesma probabilidade de detecção, porém isso não é a regra (Schmidt & Pellet, 2009). Os efeitos da detecção imperfeita das espécies na modelagem de ocorrência são amplamente discutidos para animais (Kéry et al., 2002; Schmidt & Pellet 2009, para uma série de exemplos), porém tais efeitos ainda são pouco discutidos para organismos sésseis, como as plantas (Chen et al., 2009).

Independente da técnica de regressão adotada para realizar a análise dos dados (*e.g.* GLM, GAM, CART), todos tratam igualmente as ausências corretas e as falsas ausências (Gu & Swihart, 2004; Kéry et al. 2010). Ao assumir isso, a tendência é a subestimação dos coeficientes de regressão associados as variáveis independentes do modelo, podendo, em alguns casos, até mesmo negligenciar a influência de algumas variáveis importantes.

Neste artigo buscamos responder três questões: (i) Amostras exaustivas, com presenças e ausências bem documentadas, produzem modelos de ocorrência mais precisos? (ii) Qual o impacto de falsos zeros na área ocupada estimada para a espécie? Nestes dois casos, nossa hipótese é de que quanto menor a presença de falsos zeros, tanto a precisão dos modelos quanto a área de ocorrência estimada são maiores. (iii) A relação entre a ocorrência da espécie e as variáveis ambientais que descrevem sua distribuição é alterada de acordo com a detectabilidade? Para responder estas perguntas utilizamos um banco de dados criado a partir de um exaustivo levantamento de campo realizado para a espécie *Araucaria angustifolia* (Bertol.) O. Kuntze (Araucariaceae) no entorno do reservatório de uma usina hidroelétrica no sul do Brasil (Capítulo 1). Simulamos mudanças no padrão de detecção da espécie alterando a abundância dentro das unidades amostrais (detecção dependente da abundância). Com isso foi possível acessar o efeito de diferentes níveis de detectabilidade nos modelos de ocorrência.

Material e Métodos

Área de estudo e espécie estudada

O levantamento de campo foi realizado no entorno da Usina Hidroelétrica de Barra Grande, sul do Brasil (Figura 1). Localizado na bacia do rio Pelotas, entre os estados de Santa Catarina e Rio Grande do Sul, seu reservatório cobre uma área de aproximadamente 90 km² e seu entorno, definido pelos contornos da bacia hidrográfica do rio Pelotas, tem aproximadamente 4.600 km². A altitude na região varia entre 500 e 1.200 m a.n.m. e, dependendo da cota altitudinal, o clima na região pode ser classificado como Cfa ou Cfb (classificação de Köppen), com precipitação anual de 1.412 mm e temperatura média de 15,2°C (Maluf, 2000). Nas áreas planas ou levemente onduladas, localizadas no Planalto Meridional Brasileiro, as classes predominantes de solo são os Cambissolos e Argissolos, enquanto que nas encostas íngremes próximas ao rio Pelotas, a classe de solo mais frequente é Litossolos (*sensu* Sistema Brasileiro de Classificação do Solo – SiBCS; Potter, 2004). A vegetação é caracterizada pela ocorrência de áreas contínuas de floresta estacional semidecídua, localizada predominantemente nas margens do rio Pelotas. Nas áreas altas (≥ 800 m) domina a floresta ombrófila mista (floresta com Araucária), a qual é fragmentada naturalmente por campos naturais (Klein, 1978; Veloso, 1991).

O pinheiro-brasileiro (*A. angustifolia*) é uma árvore dióica, perenifólia, com fuste reto, galhos horizontais e alcança alturas entre 25 – 35 m, com alguns exemplares alcançando até 50 m (Mattos, 1994). No Brasil, ocorre de forma descontínua desde o estado de Minas Gerais até o estado do Rio Grande do Sul (Sobral et al., 2006). Devido à elevada abundância, à sua arquitetura única quando adulto (em forma de candelabro) e pela presença de folhas coriáceas com ápice espinhoso, que facilitam a observação e

identificação de indivíduos jovens, *A. angustifolia* é facilmente reconhecida em campo em todos os seus estágios de vida, fato que minimiza possíveis erros de amostragem (falsas ausências).

Levantamento de campo e manipulação dos dados

Amostramos 388 parcelas georreferenciadas (10 x 50 m) distribuídas aleatoriamente no entorno do reservatório da UHE Barra Grande. Em cada parcela contamos o número de indivíduos de *A. angustifolia* com altura $\geq 1,5$ m. Para prevenir ou reduzir o impacto da topografia nas amostras, as parcelas foram alocadas ao longo das linhas de contorno do relevo, similar ao descrito por Magnusson et al. (2005) para levantamentos rápidos da biodiversidade (ver Capítulo 1).

Baseados em um exaustivo levantamento de campo e na facilidade de reconhecimento da espécie estudada assumimos que nosso banco de dados é “completo”, com nenhuma falsa ausência ($p \approx 1$), embora isso não seja um evento frequente em levantamentos de campo (Kéry & Schmidt, 2008). Assumindo que a detectabilidade está associada à abundância local da espécie, gradualmente substituímos todas as classes de abundância por falsos zeros, desde a menor classe de abundância (parcelas com apenas 1 indivíduo; $n = 260$ ocorrências) até as maiores classes de abundância (parcelas com 30 ou mais indivíduos; $n = 12$ ocorrências). Estes dados foram então convertidos em dados de presença ou ausência e posteriormente modelados.

Variáveis ambientais

Para prever a distribuição espacial de *A. angustifolia* utilizamos 11 variáveis ambientais (Tabela 1, ver Capítulo 1). Todas as cartas utilizadas tinham resolução espacial de 30 m e foram construídas e armazenadas em um Sistema de Informações Geográficas (SIG) utilizando o programa Quantum GIS/GRASS v.1.6.0 (Quantum GIS Development Team 2010).

Modelos de ocorrência

Modelamos a presença/ausência (ou o padrão de detecção/não detecção *sensu* Kéry et al., 2010) de *A. angustifolia* utilizando modelos lineares generalizados (GLM; distribuição binomial, função de ligação logística), tendo como preditores as variáveis ambientais descritas na Tabela 1. Os preditores foram selecionados usando um procedimento *stepwise*, com base no Critério de Informação de Akaike (AIC; Akaike, 1974). A estimativa da área ocupada pela espécie, para cada modelo, foi calculada utilizando como limite de corte (*threshold*) para conversão do mapa de probabilidade de ocorrência, em um mapa de presença e ausência, o método de maximização da soma dos parâmetros sensibilidade (capacidade do modelo prever corretamente a ocorrência da espécie) e especificidade (capacidade do modelo prever corretamente a ausência da espécie) proposto por Cantor et al. (1999) e recomendado por Liu et al. (2005).

Avaliação da precisão dos modelos

A precisão dos modelos foi avaliada utilizando a área sob a curva característica do operador-receptor (AUC; Fielding & Bell, 1997). O valor de AUC varia entre 0 e 1,

sendo que valores próximos a 0,5 indicam que a precisão do modelo não é melhor do que ao acaso, enquanto valores próximos a 1 indicam que o modelo tem 100% de chance de corretamente classificar um evento (Fielding & Bell, 1997). Para testar a hipótese nula de que a área sob a curva ROC é ≤ 0.5 , foi utilizado o teste U de Mann-Whitney (função *roc.area* no pacote R *verification* v 1.31; Mason & Graham, 2002).

Também validamos o modelo final para cada classe de abundância utilizando a correlação de Spearman (*rho*) entre o valor predito de ocorrência e a abundância real da espécie coletada nas mesmas parcelas (Albert & Thuiller 2008; Lobo et al. 2008). Neste caso, nossa premissa é que se um modelo corretamente prediz a abundância, ele tem uma forte capacidade de discriminar a ocorrência ou ausência da espécie (Jiménez-Valverde et al., 2007). Todas as análises estatísticas foram realizadas utilizando o programa R (v. 2.10.0; R Development Core Team, 2009).

Resultados

A redução na detectabilidade gerou alterações tanto nas variáveis ambientais que explicam a ocorrência da espécie, quanto na precisão dos modelos de ocorrência (Tabela 2). A única variável ambiental que ocorreu em todos os modelos gerados foi o pH (H₂O) do solo. Com exceção das classes mais extremas de detectabilidade (parcelas com abundância ≥ 20 e 30 indivíduos), essa variável não apresentou nenhum viés, mantendo constantes os valores do coeficiente de regressão (β) e erro padrão (Tabela 2).

Os valores de AUC foram significativos (teste U de Mann-Whitney, $P < 0,05$) variando de 0,84 (parcelas com abundância ≥ 30 indivíduos) a 0,68 (parcelas com abundância > 6 indivíduos), indicando uma redução não linear dos valores de AUC em

relação à diminuição da detectabilidade da espécie estudada. A outra medida adotada para verificar a acurácia dos modelos foi a correlação da probabilidade de ocorrência com os dados de abundância, conforme indicado por Lobo et al. (2008). Neste caso, ocorreu um decréscimo linear entre o modelo com todos os dados (modelo completo) e o modelo com apenas as parcelas com abundância ≥ 30 indivíduos. Alterações na detectabilidade também afetaram a estimativa da área ocupada pela espécie (Tabela 2, Figura 2). A área ocupada foi subestimada em três casos (parcelas com abundância > 1 , ≥ 20 e ≥ 30), nos demais casos a área ocupada foi superestimada em até 34% (parcelas com abundância > 4 indivíduos; Tabela 2).

Discussão

O sucesso de SDMs reside na amostragem de todo o espaço ambiental no qual o organismo em questão ocorre (Vaughan & Ormerod, 2003) e na correta detecção do organismo (Kéry et al., 2010). Falhas nestes itens podem provocar mudanças drásticas na curva de reposta das espécies ao ambiente e na área ocupada pelas mesmas, diminuindo a aplicabilidade dos modelos. Falhas na detecção da ocorrência de espécies são costumeiramente relacionadas a espécies com baixa abundância, porém, também podem ser ligadas a falhas no método de amostragem utilizado (Chen et al., 2009; Kéry et al., 2010). Baseados em um levantamento relativamente extenso, com descrições precisas de ocorrência e ausência, conseguimos demonstrar que a mínima redução na detecção de uma espécie comum afeta a precisão dos modelos (habilidade de prever corretamente ocorrências e ausências), as estimativas de área ocupada e a relação entre as variáveis ambientais que descrevem a distribuição de um organismo.

Muito utilizado como indicador da precisão dos modelos de ocorrência, a área sob a curva ROC (AUC; Fielding & Bell, 1997) tem sido alvo, nos últimos anos, de críticas sobre seu uso indiscriminado (ver Austin, 2007; Lobo et al., 2008). O uso do AUC ganhou espaço principalmente por ser um índice sintético que mede a habilidade dos modelos de discernir a presença ou ausência de forma não subjetiva (*threshold-independent measure*; Fielding & Bell, 1997), diferente de outros índices comumente utilizados (e.g. Cohen's Kappa, taxa de classificação incorreta), que escolhem de forma subjetiva a partir de qual probabilidade assumimos a ocorrência ou ausência de um organismo. Uma das principais críticas à curva AUC tem ligação direta com a capacidade de detecção de um organismo em campo. AUC dá o mesmo peso tanto para os erros de omissão (subestimar a área de ocorrência) quanto de comissão (superestimar da área de ocorrência), porém, o grau de incerteza associado aos dados de ausência são, em geral, maiores do que os associados a dados de presença, principalmente para espécies raras ou com baixos níveis de detectabilidade. Isso significa dizer que o erro de omissão tende a ser mais frequente do que o erro de comissão, tornando essa métrica muitas vezes enganosa (Lobo et al., 2008). Nossos dados indicam, como padrão geral, que a redução na detectabilidade gera um aumento da área estimada de ocorrência, sugerindo um aumento do erro de comissão em detrimento de uma diminuição do erro de omissão, informação omitida pelo AUC. Esse padrão foge do já descrito por outros estudos (Chen et al., 2009; Kéry et al., 2010), os quais indicam que falhas na detecção geram distribuições geográficas subestimadas (maior erro de omissão). Aparentemente, isso é verdade apenas para os casos mais graves de baixa detectabilidade, onde a espécie é detectada apenas em locais onde ela ocorre com elevada abundância. Erros de comissão e omissão têm impactos diferentes de acordo com as possíveis aplicações dos

modelos. Quando o objetivo é a criação e implementação de unidades de conservação ou a identificação de áreas para a translocação ou reintrodução de espécies, o erro de comissão é mais sério do que o erro de omissão, que é indesejável em modelos que têm como aplicação final a identificação de novas áreas de ocorrência de espécies ou populações (Peterson, 2006).

A correlação entre probabilidade de ocorrência e abundância observada se mostrou mais sensível à redução da detectabilidade do que o AUC. Enquanto o AUC sobe nas classes superiores de redução de abundância (parcelas com abundância ≥ 20 e ≥ 30), efeito do aumento da área de calibração do modelo como demonstrado por Hernandez et al. (2006), Wisz et al. (2008), Franklin et al. (2009) e Giovanelli et al. (2010), a correlação entre a probabilidade de ocorrência e abundância observada decresce linearmente conforme diminui a detectabilidade da espécie. Isso ocorre porque a probabilidade de ocorrência é positivamente correlacionada com a abundância (relação Ocupância-abundância; Gaston et al., 2000). Devido a essa ligação intrínseca entre ocorrência e abundância, um dos possíveis usos dos SDMs é em modelos metapopulacionais (Guisan & Thuiller, 2005). Porém, segundo Moilanen et al. (2002) modelos construídos com base em bancos de dados com falsas ausências, podem ocasionar erros, como a superestimação do risco de extinção, da distância de migração e da capacidade de colonização de uma paisagem por determinada espécie.

Devido à facilidade de análise e entendimento, GLM é a técnica mais utilizada para a modelagem de padrões de distribuição espacial de espécies, permitindo a análise de interações complexas entre ambiente e ocorrência, porém, como indicado por Gu & Swihart (2004) e aqui demonstrado, esta técnica é sensível até mesmo a pequenos níveis de alteração na detecção da espécie. Isso ocorre, por GLM tratar igualmente tanto as

ausências corretas quanto as falsas ausências, assumindo, intrinsecamente, que a probabilidade de detecção da espécie estudada é sempre igual a 1 (Gu & Swihart, 2004; Kéry et al., 2010). Ao aceitar essa simples suposição, a não detecção da espécie, causa alterações na área ocupada, como também na relação entre a ocorrência e seus descritores ambientais. Segundo Tyre et al. (2001, 2003), quando não ocorrem falhas na detecção, todos os parâmetros da regressão logística (intercepto [α] e inclinação da curva de resposta [β]) são precisamente estimados, porém na presença de falhas de detecção, estes dois parâmetros são subestimados. Segundo os mesmos autores, o efeito no intercepto é facilmente entendido, pois quanto menor a proporção de presenças no conjunto de dados, menor será o valor do intercepto, porém o efeito na inclinação da curva de resposta da ocorrência em relação às variáveis preditoras é tênue e espacialmente dependente. Uma das principais proposições da relação abundância é a de que a maior abundância encontrada para uma espécie, localiza-se no centro de sua distribuição, decaindo gradativamente em direção às bordas da área de ocorrência da espécie (Brown, 1984). Curiosamente, ao assumirmos isso, observamos que a distribuição espacial das variáveis ambientais não condiciona apenas a distribuição das espécies, mas também a sua detectabilidade (Kéry et al., 2010).

Cada vez mais o monitoramento da biodiversidade se vale do uso de modelos de distribuição e de riqueza de espécies. No entanto, a maioria dos programas de monitoramento ignora ou lida de forma ineficiente com este assunto (Yoccoz et al., 2001). Não contabilizar os efeitos da detectabilidade nestes casos é preocupante. Como a riqueza de espécies é a soma do número de espécies em um determinado local, falhas na detecção das espécies geram predições aparentes de riqueza e não a riqueza específica real de espécies (Royle & Dorazio, 2006; Kéry et al., 2010), fato que

inviabiliza fortes inferências sobre o ecossistema em monitoramento (Yoccoz et al., 2001).

Os resultados aqui descritos devem ser observados com cautela, pois analisamos apenas um dos fatores que afetam a detectabilidade (*i.e.* abundância). Outros fatores, como tamanho da amostra e tamanho e forma da parcela, também devem ser levados em conta e simulados futuramente. Como passo futuro, devemos simular também qual o impacto de falhas na detecção em modelos que utilizam apenas dados de presença (*e.g.* GARP, ENFA, MAXENT). Por usarem apenas a ocorrência da espécie, criando pseudo-ausências (ausências aleatoriamente escolhidas), acreditamos que a influência da detectabilidade é maior do que nos modelos tradicionais (presença/ausência), pois a incerteza associada a estas (pseudo) ausências é maior do que a incerteza associada a dados reais de ausência.

Agradecimentos

Os autores são gratos a AM Barbosa (Universidade de Évora, Cátedra Rui Nabeiro) pelo apoio inicial para a realização do trabalho. FS Rocha (PPG Botânica, Universidade Federal do Rio Grande do Sul) pelas úteis conversas e leitura inicial do manuscrito. O primeiro autor também é grato ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; proc. 140712/2007-0) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; proc. 2298/09-0) pela concessão das bolsas de estudo. O levantamento de campo teve auxílio logístico e financeiro da Embrapa Recursos Genéticos e Biotecnologia.

Referências

- Austin, M. (2007) *Species distribution models and ecological theory: A critical assessment and some possible new approaches*. *Ecological Modelling*, **200**, 1-19.
- Albert, C. H. & Thuiller, W. (2008) *Favourability functions versus probability of presence: advantages and misuses*. *Ecography*, **31**, 417-422.
- Brown, J. H. (1984) *On the relationship between abundance and distribution of species*. *American Naturalist*, **124**, 255-279.
- Cantor, S. B., Sun, C. C., Tortolero-Luna, G., Richards-Kortum, R. & Follen, M. (1999) *A Comparison of C/B Ratios from Studies Using Receiver Operating Characteristic Curve Analysis*. *Journal of clinical epidemiology*, **52**, 885-892.
- Cayuela, L., Golicher, D. J., Newton, A. C., Kolb, M., Albuquerque, F. S. d., Arets, E. J. M. M., Alkemade, J. R. M. & Perez, A. M. (2009) *Species distribution modeling in the tropics: problems, potentialities, and the role of biological data foreffective species conservation*. *Tropical Conservation Science*, **2**, 319-352.
- Chen, G. K., Kéry, M., Zhang, J. L. & Ma, K. P. (2009) *Factors affecting detection probability in plant distribution studies*. *Journal of Ecology*, **97**, 1383-1389.
- Fielding, A. H. & Bell, J. F. (1997) *A review of methods for the assessment of prediction errors in conservation presence/absence models*. *Environmental Conservation*, **24**, 38-49.
- Franklin, J., Wejnert, K. E., Hathaway, S. A., Rochester, C. J. & Fisher, R. N. (2009) *Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California*. *Diversity and Distributions*, **15**, 167-177.
- Giovanelli, J. G. R., de Siqueira, M. F., Haddad, C. F. B. & Alexandrino, J. (2010) *Modeling a spatially restricted distribution in the Neotropics: How the size of calibration area affects the performance of five presence-only methods*. *Ecological Modelling*, **221**, 215-224.
- Gu, W. & Swihart, R. K. (2004) *Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models*. *Biological Conservation*, **116**, 195-203.
- Guisan, A. & Thuiller, W. (2005) *Predicting species distribution: offering more than simple habitat models*. *Ecology Letters*, **8**, 993-1009.

- 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.
- 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.
- Jiménez-Valverde, A., Ortuno, V. M. & Lobo, J. M. (2007) *Exploring the distribution of *Sterocorax Ortuno, 1990* (Coleoptera, Carabidae) species in the Iberian peninsula.* *Journal of Biogeography*, **34**, 1426-1438.
- Kéry, M., Gardner, B. & Monnerat, C. (2010) *Predicting species distributions from checklist data using site-occupancy models.* *Journal of Biogeography*, no-no.
- Kéry, M., Madsen, J. & Lebreton, J. D. (2006) *Survival of Svalbard pink-footed geese *Anser brachyrhynchus* in relation to winter climate, density and land-use.* *Journal of Animal Ecology*, **75**, 1172-1181.
- Kéry, M. & Schmidt, B. R. (2008) *Imperfect detection and its consequences for monitoring for conservation.* *Community Ecology*, **9**, 207-216.
- Klein, R. M. (1978) *Mapa fitogeográfico do Estado de Santa Catarina.* Herbário Barbosa Rodrigues, Itajaí, Brazil.
- Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. (2005) *Selecting thresholds of occurrence in the prediction of species distributions.* *Ecography*, **28**, 385-393.
- Lobo, J. M., Jiménez-Valverde, A. & Real, R. (2008) *AUC: a misleading measure of the performance of predictive distribution models.* *Global Ecology and Biogeography*, **17**, 145-151.
- Magnusson, W. E., Lima, A. P., Luizão, R., Luizão, F., Costa, F. R. C., Castilho, C. V. d. & Kinupp, V. F. (2005) *RAPELD: a modification of the Gentry method for biodiversity surveys in long-term ecological research sites.* *Biota Neotropica*, **5**, 19-24.
- Maluf, J. R. T. (2000) *Nova classificação climática do Estado do Rio Grande do Sul.* *Revista Brasileira de Agrometeorologia*, **8**, 141-150.
- Mattos, J. R. (1994) *O pinheiro brasileiro.* Lages, Brazil.
- McCarthy, M.A & Elith, J. (2002). *Species mapping for conservation.* *Gap Analysis Bulletin*. **2**, 50-57.

- Moilanen, A. (2002) *Implications of empirical data quality to metapopulation model parameter estimation and application*. *Oikos*, **96**, 516-530.
- Pauchard, A., Kueffer, C., Dietz, H., Daehler, C. C., Alexander, J., Edwards, P. J., Arevalo, J. R., Cavieres, L. A., Guisan, A., Haider, S., Jakobs, G., McDougall, K., Millar, C. I., Naylor, B. J., Parks, C. G., Rew, L. J. & Seipel, T. (2009) *Ain't no mountain high enough: plant invasions reaching new elevations*. *Frontiers in Ecology and the Environment*, **7**, 479-486.
- Peterson, A. T. (2006) *Uses and Requirements of Ecological Niche Models and Related Distributional Models*. *Biodiversity Informatics*, **3**, 14.
- Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006) *Maximum entropy modeling of species geographic distributions*. *Ecological Modelling*, **190**, 231-259.
- Potter, R. O., Carvalho, A. M., Flores, C. A. & Bognola, I. (2004) Solos do Estado de Santa Catarina. Boletim de Pesquisa, 46. Embrapa Solos.
- Quantum GIS Development Team (2010). *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project. Disponível em: <http://qgis.osgeo.org>
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org>.
- Royle, A., Nichols, J. & Kéry, M. (2005) *Modelling occurrence and abundance of species when detection is imperfect*. *Oikos*, **110**, 353-359.
- Royle, J. A. & Dorazio, R. M. (2006) *Hierarchical models of animal abundance and occurrence*. *Journal of Agricultural Biological and Environmental Statistics*, **11**, 249-263.
- Schmidt, B. R. & Pellet, J. (2009) Quantifying abundance: counts, detection probabilities, and estimates. *Amphibian Ecology and Conservation: A Handbook of Techniques* (eds C. K. Dodd Jr), pp. 465-480. Oxford University Press.
- Sobral, M., Jarenkow, J. A., Brack, P., Irgang, B., Larocca, J. & Rodrigues, R. S. (2006) *Flora Arborescente do Rio Grande do Sul, Brasil*. RIMA:Novo Ambiente, São Carlos, Brazil.

- Tyre, A. J., Possingham, H. P. & Lindenmayer, D. B. (2001) *Infering process from pattern: can territory occupancy provide information about life history parameters? Ecological Applications*, **11**, 1722-1737.
- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K. M. & Possingham, H. P. (2003) *Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecological Applications*, **13**, 1790-1801.
- Vaughan, I. P. & Ormerod, S. J. (2003) *Improving the Quality of Distribution Models for Conservation by Addressing Shortcomings in the Field Collection of Training Data. Conservation Biology*, **17**, 1601-1611.
- Veloso, H. P., Rangel-Filho, A. L. & Lima, J. C. A. (1991) *Classificação da vegetação brasileira, adaptada a um sistema universal*. IBGE, Rio de Janeiro, Brazil.
- Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A. & Group, N. P. S. D. W. (2008) *Effects of sample size on the performance of species distribution models. Diversity and Distributions*, **14**, 763-773.
- Yoccoz, N. G., Nichols, J. D. & Boulinier, T. (2001) *Monitoring of biological diversity in space and time. Trends in ecology & evolution (Personal edition)*, **16**, 446-453.

Legenda das Figuras

Figura 1. Localização geográfica da área de estudo e das unidades amostrais (sistema de coordenadas: Universal Transverse Mercator, Zone 22J, Hemisfério Sul, Brasil, entre os Estados de Santa Catarina e Rio Grande do Sul).

Figura 2. Exemplos de mapas preditivos produzidos com o uso de modelos lineares generalizados (GLM) com diferentes padrões de detecção de *A. angustifolia*. (a) Todas as parcelas, (b) parcelas com > 1 indivíduo, (c) parcelas com ≥ 30 indivíduos. Valores de AUC, Threshold (T) e ρ são apresentados para cada mapa. Pixels brancos representam a área ocupada pela espécie.

Figura 1

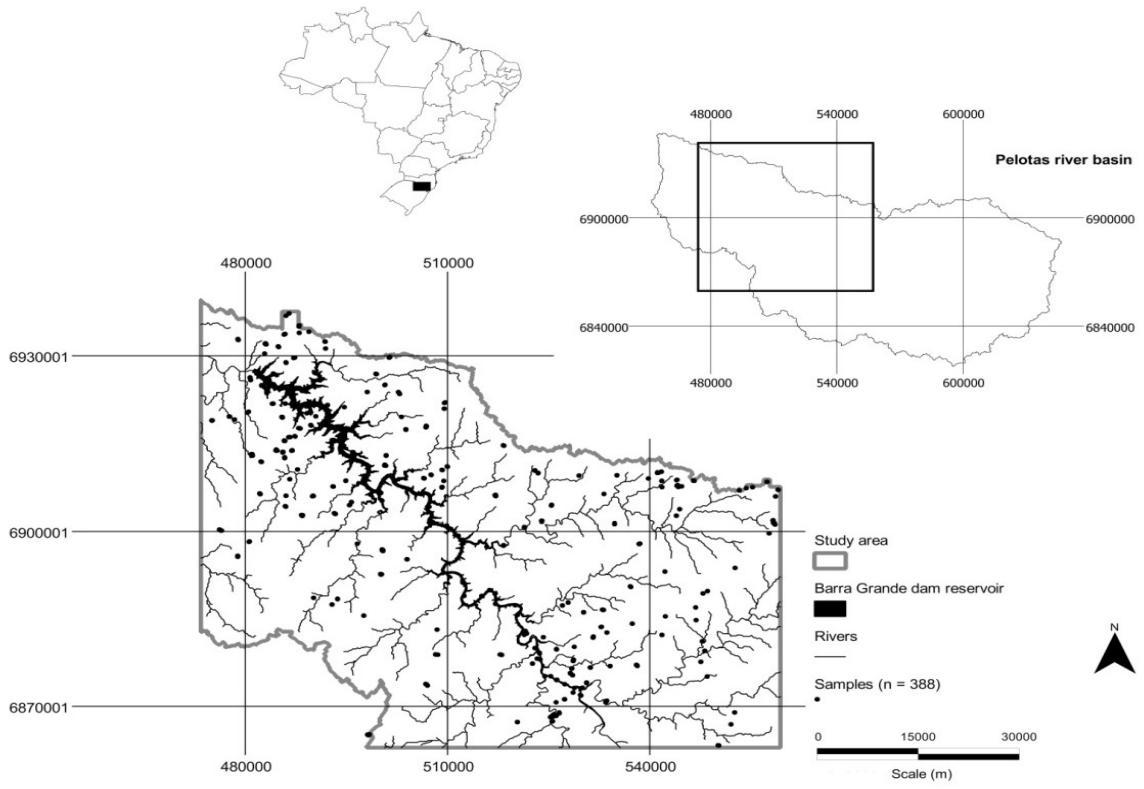


Figura 2

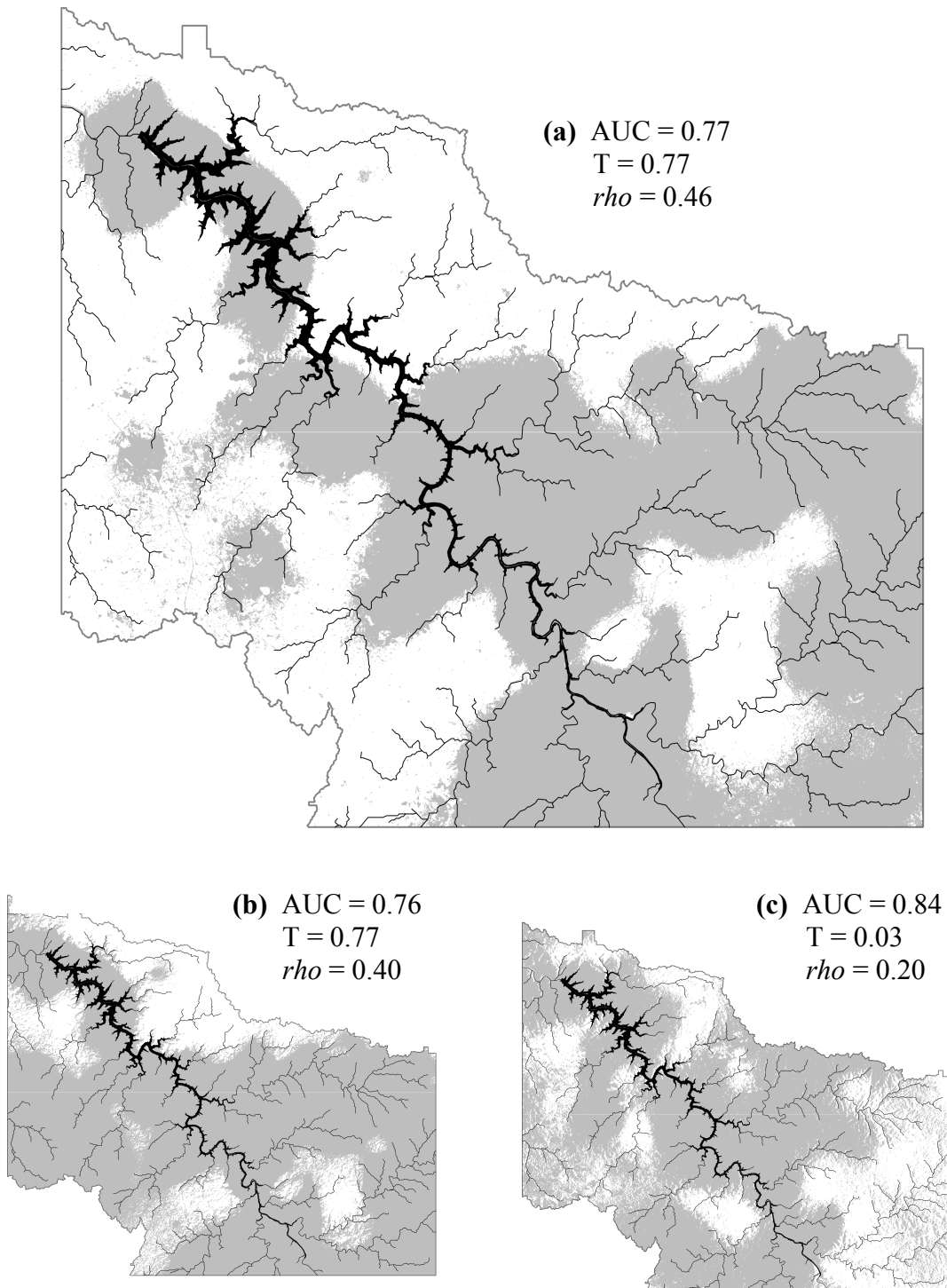


Tabela 1. Variáveis ambientais utilizadas para predizer a distribuição de *A. angustifolia* na área de estudo. * Variáveis excluídas da análise devido ao alto coeficiente de correlação de Pearson (r) com uma ou mais variáveis ($r \geq 0.7$; $P < 0.01$, para maiores informações ver Capítulo 1).

Grupo	Variável
Solo	pH H ₂ O
	Ca ⁺² +Mg ⁺² *
	H ⁺ +Al ⁻³ *
	K ⁺
	S ⁻²
	P ⁺
	N total
	Relação Silte:Argila*
	Densidade do solo
	Topografia
Exposição norte-sul do relevo (<i>northness</i>)	
Exposição leste-oeste do relevo (<i>eastness</i>)	
Índice topográfico de umidade (TWI)	
Declividade (graus)	
Vegetação	Índice de vegetação da diferença normalizada (NDVI)

Tabela 2. Variáveis incluídas nos modelos lineares generalizados (GLM) para as diferentes classes de abundância de *A. angustifolia*. Valores entre parênteses indicam o número de parcelas onde a espécie está presente e ausente. β : coeficiente de regressão; **E.P.:** erro padrão, **AUC:** área sob a curva ROC (* $P \leq 0.01$), **rho:** correlação de Spearman (** $P \leq 0.01$), **T** (threshold): maximização do valor da soma da sensibilidade-especificidade e **AOc:** Área Ocupada (km²). † *não significativa*.

Abundância	Variáveis	β	E.P.	AUC	rho	T	AOc
Modelo completo (281/107)	Intercepto	5,74	1,41	0,77*	0,42**	0,77	2.250,32
	pH H ₂ O	- 1,41	0,29				
	S	0,05	0,01				
	N	3,33	1,30				
	NDVI	1,28	0,63				
	K [†]	-0,30	0,17				
]0,1] (260/128)	Intercepto	6,83	1,42	0,76*	0,40**	0,77	1.019,56
	pH H ₂ O	- 1,55	0,30				
	S	0,03	0,01				
	N	3,08	1,29				
	Eastness	1,56	0,76				
	Declividade	-0,04	0,20				
	Northness [†]	-1,44	0,83				
]0,2] (237/151)	Intercepto	5,75	1,23	0,71*	0,38**	0,65	2.944,53
	pH H ₂ O	- 1,07	0,23				
	K	- 0,32	0,32				
	S [†]	0,02	0,01				
	Northness [†]	-1,33	0,78				
]0,3] (214/174)	Intercepto	5,71	1,22	0,70*	0,36**	0,56	2.691,47
	pH H ₂ O	- 1,10	0,23				
	K	- 0,33	0,16				
	S [†]	0,01	0,01				
	Northness [†]	-1,28	0,79				
]0,4] (185/203)	Intercepto	4,43	1,31	0,69*	0,35**	0,45	3.014,05
	pH H ₂ O	- 1,08	0,23				
	K [†]	-0,27	0,17				
	S [†]	0,01	0,01				
	TWI [†]	0,11	0,08				
]0,5] (165/223)	Intercepto	4,25	1,33	0,70*	0,34**	0,39	2.962,40
	pH H ₂ O	- 1,10	0,24				
	K [†]	-0,32	0,18				
	S [†]	0,01	0,01				
	TWI [†]	0,12	0,07				

Tabela 1. Continuação.

Abundância	Variáveis	β	E.P.	AUC	ρ	T	AOc
]0,6] (144/244)	Intercepto	4,05	1,34				
	pH H ₂ O	- 1,18	0,23				
	TWI	0,17	0,08				
	S [†]	0,01	0,01	0,68*	0,33**	0,36	2.429,61
	P [†]	-0,10	0,06				
	Eastness [†]	1,14	0,78				
]0,7] (125/263)	Intercepto [†]	-5,16	4,23				
	pH H ₂ O	- 1,27	0,25				
	S	0,02	0,01				
	Densidade do solo	3,45	1,62	0,70*	0,35**	0,32	2.353,55
	Eastness [†]	1,27	0,85				
	TWI [†]	0,15	0,08				
]0,8] (112/276)	Intercepto [†]	- 3,67	4,33				
	pH H ₂ O	- 1,36	0,26				
	S [†]	0,01	0,01				
	Densidade do solo [†]	3,12	1,66	0,69*	0,32**	0,25	2.639,25
	Eastness [†]	1,43	0,89				
	TWI [†]	0,12	0,08				
]0,9] (101/287)	Intercepto [†]	-2,57	4,41				
	pH H ₂ O	- 1,45	0,28				
	S [†]	0,02	0,01	0,70*	0,32**	0,26	2.989,57
	Densidade do solo [†]	3,15	1,71				
]0,10] (86/302)	Intercepto [†]	- 2,69	4,69				
	pH H ₂ O	- 1,62	0,32				
	S [†]	0,03	0,01	0,73*	0,33**	0,23	2.299,27
	Densidade do solo [†]	3,28	1,83				
]0,20] (29/359)	Intercepto	- 21,92	7,19				
	pH H ₂ O	-2,13	0,75				
	S	0,04	0,01				
	Densidade do solo	11,12	3,06	0,82*	0,29**	0,08	1.546,44
	Eastness	5,85	2,77				
	Declividade [†]	-0,10	0,06				
]0,30] (12/376)	Intercepto	15,89	5,93				
	pH H ₂ O	- 3,51	1,18				
	P	- 0,60	0,23	0,84*	0,20**	0,03	1.678,84
	Eastness	5,90	2,60				
	NDVI	- 3,32	1,93				

CONCLUSÃO GERAL

Nosso estudo é pioneiro no uso de análises macroecológicas para dar suporte a práticas de conservação *in situ* e *ex situ* em regiões afetadas por empreendimentos hidroelétricos. Atualmente, no Brasil os métodos aplicados no licenciamento ambiental dependem de lentos e caros levantamentos de campo (florísticos e fitossociológicos), geralmente conduzidos por profissionais sem formação para entender de forma ampla a elevada diversidade da flora brasileira. Como resultado dessa combinação, estes levantamentos geralmente produzem dados de baixa qualidade e com pouca aplicabilidade. Mesmo com limitações metodológicas, nós acreditamos que o licenciamento ambiental no Brasil deve considerar o uso de modelos de distribuição de ocorrência e abundância de espécies para prever impactos futuros e planejar estratégias de manejo e mitigação das áreas afetadas pelos empreendimentos a serem licenciados. Estes métodos são mais baratos e rápidos do que o utilizado atualmente. Cabe aqui ressaltar que, igualmente ao procedimento anterior, o aqui apresentado também necessita do treinamento dos técnicos envolvidos, tanto para a análise quanto para a coleta adequada de dados.

No primeiro capítulo demonstramos que modelos de ocorrência são mais precisos do que modelos de abundância e que podemos utilizar os dados de probabilidade de ocorrência como um indicador da abundância local das espécies. Reconstruir padrões de abundância a partir de dados de ocorrência pode ser uma ferramenta importante para o planejamento de ações de conservação e manejo de espécies ameaçadas de extinção,

permitindo a seleção de áreas para a conservação de populações viáveis destas espécies ou ações para a sua coleta e relocação.

Por último, demonstramos no segundo capítulo os impactos de erros de detecção na geração de modelos de distribuição de espécies (redução na precisão dos modelos, alteração da área ocupada pela espécie e alteração nas variáveis independentes do modelo). Não contabilizar os efeitos da detectabilidade nestes casos é preocupante, pois a grande maioria dos programas de monitoramento da biodiversidade se vale do uso da riqueza de espécies como indicador da integridade dos ecossistemas monitorados. Como a riqueza é a soma do número de espécies em um determinado local, falhas na detecção das espécies geram predições aparentes de riqueza e não a riqueza real de espécies, fato que inviabiliza inferências sobre o ecossistema em monitoramento.