

UNIOESTE – Universidade Estadual do Oeste do Paraná

CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS

Colegiado de Ciência da Computação

Curso de Bacharelado em Ciência da Computação

**Avaliação de Três Métodos para Geração de Pontos de
Pseudo-Ausência Sobre a Qualidade dos Modelos de
Distribuição de Espécies Ajustados pelo Sistema
SAHGA SDM**

Pedro Patitucci Finamore

CASCABEL

2010

PEDRO PATITUCCI FINAMORE

**AVALIAÇÃO DE TRÊS MÉTODOS PARA GERAÇÃO DE PONTOS DE
PSEUDO-AUSÊNCIA SOBRE A QUALIDADE DOS MODELOS DE
DISTRIBUIÇÃO DE ESPÉCIES AJUSTADOS PELO SISTEMA SAHGA
SDM**

Monografia apresentada como requisito para
obtenção do grau de Bacharel em Ciência da
Computação, do Centro de Ciências Exatas e
Tecnológicas da Universidade Estadual do
Oeste do Paraná - Campus de Cascavel

Orientador: Prof. Dr. Adair Santa Catarina

CASCADEL

2010

PEDRO PATITUCCI FINAMORE

**AVALIAÇÃO DE TRÊS MÉTODOS PARA GERAÇÃO DE PONTOS DE
PSEUDO-AUSÊNCIA SOBRE A QUALIDADE DOS MODELOS DE
DISTRIBUIÇÃO DE ESPÉCIES AJUSTADOS PELO SISTEMA SAHGA
SDM**

Monografia apresentada como requisito para obtenção do Título de *Bacharel em Ciência da Computação*, pela Universidade Estadual do Oeste do Paraná, Campus de Cascavel, aprovada pela Comissão formada pelos professores:

Prof. Dr. Adair Santa Catarina (Orientador)
Colegiado de Ciência da Computação,
UNIOESTE

Prof. M.Eng. André Luiz Brun
Colegiado de Ciência da Computação,
UNIOESTE

Prof. M.Eng. Carlos Jose Maria Olguín
Colegiado de Ciência da Computação,
UNIOESTE

Cascavel, 04 de Novembro de 2010.

A minha família, namorada e amigos dedico.

AGRADECIMENTOS

Aos meus pais, Giane e Nelson, pela educação, incentivo, apoio e confiança incondicionais em todos os anos de minha formação.

Aos meus irmãos, Wanilton e Samuel, pelo companheirismo e amizade.

À minha namorada, Vanessa, pela paciência, confiança, amizade e companheirismo em todos estes anos.

Aos meus amigos da Unioeste: Alessandro, Xmia, Martins, Fred, Gustavo, Marco, Almeida e Grandó pelos momentos agradáveis, pelas risadas e pela amizade proporcionada.

Ao meu grande amigo Ivan pela força nos momentos de fraqueza, pela amizade e pela consideração.

Ao “Bonde”: Ivo, Hugo, Carlos e Darkan, pela contínua contribuição em minha formação pessoal.

Ao professor Adair Santa Catarina, pela paciência, empenho e dedicação nas atividades de orientação.

Lista de Figuras

2.1. Classificação dos modelos matemáticos aplicados em ecologia	8
2.2. Elementos essenciais na modelagem de distribuição de espécies	9
2.3. Estrutura geral de um sistema para geração de SDM	10
2.4. Estrutura Geral do Sistema SAHGA SDM	12
2.5. Regra Empregada na Construção da GPM	13
2.6. Matriz de Confusão	15
2.7. Simulação de um modelo com erros de omissão e comissão	15
2.8. Curvas ROC para três graus de capacidade de discriminação	18
2.9. Ordem Cronológica das GUI.....	19
2.10. Ambiente de Desenvolvimento do Qt Creator	21
2.11. Gerenciador de Camadas do openModeller Desktop	22
2.12. Visualizador de Mapas do openModeller Desktop	23
3.1. Geração aleatória de pontos de pseudo-ausência	24
3.2. Geração de pontos de pseudo-ausência baseado em distância	25
3.3. Geração de pontos de pseudo-ausência segundo o Bioclim	26
3.4. Geração de Pontos de Pseudo-Ausência Pelo Método da Aleatoriedade	27
3.5. Geração de Pontos de Pseudo-Ausência Pelo Método da Distância	29
3.6. Geração de Pontos de Pseudo-Ausência Pelo Método de envelopes bioclimáticos	30
3.7. Exemplo de delimitação de área para a geração de pontos de pseudo-ausência	31
4.1. Curvas ROC para os modelos selecionados para o experimento	35
A.1. Curva ROC correspondente ao experimento TR50.....	40
A.2. Curva ROC correspondente ao experimento TR65.....	40
A.3. Curva ROC correspondente ao experimento RD50	41
A.4. Curva ROC correspondente ao experimento RD65	41
A.5. Curva ROC correspondente ao experimento DS50.....	42
A.6. Curva ROC correspondente ao experimento DS65.....	42
A.7. Curva ROC correspondente ao experimento BC50.....	43
A.8. Curva ROC correspondente ao experimento BC65.....	43
B.1. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo TR50)	44
B.2. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo TR65)	45
B.3. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo RD50).....	45
B.4. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo RD65).....	46
B.7. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo DS50)	46

B.6. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo DS65)	47
B.7. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo BC50)	47
B.8. Distribuição potencial da espécie <i>Thalurania furcata boliviana</i> (modelo BC65)	48

Lista de Tabelas

2.1. Alguns possíveis usos de SDMs na ecologia e biologia conservativa	9
2.2. SDMs publicados, artigo de referência, métodos de modelagem utilizados e link <i>www</i> (quando disponível)	11
2.3. Métricas derivadas da matriz de confusão	16
2.4. Qualidade do Índice Kappa	17
4.1. Conjunto de Parâmetros pré-definidos do algoritmo SAHGA	32
4.2. Métricas para avaliação dos modelos selecionados para o experimento.....	34
4.3. Métricas para avaliação das médias dos modelos selecionados para o experimento	36
4.4. Taxas de presença e ausência segundo os modelos TR50 e TR65	36
4.5. Taxas de presença e ausência segundo os modelos RD50 e RD65	37
4.6. Taxas de presença e ausência segundo os modelos DS50 e DS65	37
4.7. Taxas de presença e ausência segundo os modelos BC50 e BC65.....	37

Lista de Abreviaturas e Siglas

AUC – Area Under the Curve

FN – Falso Negativo

FP – Falso Positivo

GPM – Generalized Proximity Matrix

ROC – Receiver Operating Characteristic

SDM – Species Distribution Models

SAHGA – Spatially Aware Hybrid Genetic Algorithm

VN – Verdadeiro Negativo

VP – Verdadeiro Positivo

SUMÁRIO

<i>Lista de Figuras</i>	<i>vi</i>
<i>Lista de Tabelas</i>	<i>viii</i>
<i>Lista de Abreviaturas e Siglas</i>	<i>ix</i>
<i>Resumo</i>	<i>xi</i>
<i>Introdução</i>	<i>1</i>
1.1. Hipótese Central	4
1.2. Organização do Texto	5
<i>Referencial Teórico</i>	<i>7</i>
2.1. Modelos de Distribuição de Espécies (SDMs)	7
2.2. Graphic User Interface (GUI)	18
2.3. openModeller Desktop	21
<i>Metodologia</i>	<i>24</i>
3.1. Geração dos Pontos de Pseudo-Ausência	26
<i>Estudo de Caso</i>	<i>32</i>
4.1. Espécie <i>Thalurania fucata boliviana</i> Boucard, 1984	33
4.2. Resultados	34
<i>Capítulo 5</i>	<i>38</i>
<i>Conclusões</i>	<i>38</i>
5.1. Trabalhos Futuros	39
<i>Anexo A</i>	<i>40</i>
<i>Curvas ROC</i>	<i>40</i>
<i>Anexo B</i>	<i>44</i>
<i>Referências Bibliográficas</i>	<i>49</i>

Resumo

A utilização de sistemas de modelagem de distribuição de espécies tem se mostrado uma ferramenta eficaz para diversas situações do cotidiano ecológico. O efeito da interferência humana em um nicho realizado, a causa da predação entre espécies e a luta pela preservação de espécies em extinção são apenas alguns exemplos da utilização destes modeladores (SDMs). Entretanto, alguns destes modeladores necessitam de um dado particularmente precioso para realizar a modelagem: pontos de ausência, que são, muitas vezes, difíceis de encontrar pela baixa disponibilidade destes dados. A alternativa para suplantar a ausência destes pontos, é gerá-los automaticamente. Pontos de pseudo-ausência são alternativas plausíveis para os pontos de ausência, pois simulam o efeito destes na modelagem de distribuição de espécies. Para tanto, este trabalho propõe a avaliação de diferentes estratégias para a geração automática de pontos de pseudo-ausência no sistema SAHGA SDM. O sistema supracitado encaixa-se no grupo de modeladores que necessitam de pontos de ausência para realizar a modelagem, porém não implementa qualquer técnica de geração de pontos de pseudo-ausência quando não há pontos de ausência verdadeiros. Utilizou-se o SAHGA SDM, juntamente com os três métodos de geração automática de pontos de pseudo-ausência – aleatório, distância e envelopes bioclimáticos - para a modelagem da distribuição potencial da espécie *Thalurania furcata boliviana*. Os resultados obtidos mostraram que, para o conjunto de dados analisado, os três métodos desenvolvidos geraram pontos de pseudo-ausência que contribuíram para o ajuste de modelos com bom poder discriminatório.

Palavras-chave: SDM, SAHGA SDM, pseudo-ausência.

Capítulo 1

Introdução

Todos os organismos têm alguma capacidade de se mover do seu local de nascimento para novos locais (Iwashita, 2007), seja por acontecimentos naturais (enchentes, mudanças geográficas), mudanças climáticas globais (Vivo e Carmignotto, 2004) ou, mais recentemente, pela ação humana. Esse deslocamento favorece as espécies porque geralmente locais diferentes são mais benéficos ao desenvolvimento do indivíduo, em parte devido à competição intra-específica, e em parte porque a qualidade do ambiente natal está sempre mudando (Daubenmirre, 1968; Iwashita, 2007). Desta maneira, através deste deslocamento, a espécie procura por um novo habitat, de acordo com suas necessidades básicas de sobrevivência.

Habitat pode ser definido como o ambiente de um organismo. Em outras palavras, é o local onde ele vive (Raven *et. al.*, 2001). No habitat a espécie nele alocada, encontrará, além de abrigo das intempéries do meio físico e de eventuais predadores, comida e espaço para se reproduzir.

O conceito de nicho está intimamente associado com o conceito de habitat, pois designa o conjunto de atividades desempenhado por uma determinada espécie dentro de seu habitat (Raven, 2001). O termo nicho é derivado do francês medieval *nicher* que quer dizer ninho, e foi cunhado pelo naturalista Joseph Grinnel em 1917, porém, o termo popularizou-se através de Hutchinson que formalizou o termo nicho definindo-o como um hipervolume n-dimensional, onde cada dimensão corresponde a um estado do ambiente que permite que determinada espécie exista indefinidamente (Hutchinson, 1957).

A partir desta definição podem-se subdividir os nichos como fundamental, realizado e potencial. O nicho fundamental é aquele em que uma determinada espécie, por causa da ausência de competição com outras espécies, pode existir em determinado habitat; o nicho realizado diz respeito a parte do nicho fundamental onde a espécie realmente ocorre, considerando a competição entre as espécies (Hutchinson, 1957; Araújo e Guisan, 2006). Em

contrapartida, o nicho potencial é um grupo de condições e recursos do meio que permitem que uma população persista quando não está sendo restringida por competição interespecífica ou predação, em uma situação hipotética (Hutchinson, 1957).

Biogeógrafos e ecologistas buscam respostas para saber como as plantas e os animais se distribuíram na Terra no tempo e espaço (Guisan & Thuiller, 2005). Um dos métodos utilizados para tentar encontrar estas respostas é através da modelagem de distribuição de espécies.

Cada vez mais pesquisadores estudam diferentes métodos para modelar a distribuição de espécies, tais como a luta pela preservação de uma espécie em extinção (Godown & Peterson, 2000), o controle de espécies exóticas ou pragas (Sutherst et al., 2000), a predação entre espécies, a interferência humana diminuindo os nichos potenciais e o estabelecimento de áreas apropriadas para translocações/relocações ou cultivo (Cunningham et al., 2002).

A disponibilidade de dados ambientais, juntamente com o uso de computadores, tem permitido a rápida obtenção de modelos de predição de requisitos ambientais e distribuição geográfica das espécies (Phillips et al., 2006).

Neste trabalho foi estudado a modelagem de distribuição de espécies (SDM - *Species Distribution Models*) que, de um modo geral, pode ser utilizada para predizer e avaliar o impacto do uso acelerado da terra e da mudança do meio-ambiente na distribuição de organismos (Guisan & Zimmerman, 2000).

Os SDM são geralmente empíricos, pois baseiam-se em amostras de campo (realidade) e são aplicados especificamente para modelar a ocorrência de uma espécie numa determinada área de estudo, através de métodos estatísticos e/ou computacionais (Guisan e Zimmermann, 2000; Iwashita, 2007). Os SDM se utilizam da modelagem matemática juntamente com as ferramentas computacionais para prever a ocorrência de espécies através da geração de superfícies temáticas, indicando presença ou ausência (Guisan & Thuiller, 2005; Santa Catarina, 2009).

Os estudos envolvendo distribuição espacial de espécies são formados por três componentes básicos: um conjunto de dados descrevendo a presença ou abundância da espécie, juntamente com outro conjunto contendo as variáveis explicativas; um modelo matemático que relaciona a espécie estudada com a variável explicativa; a avaliação da utilidade do modelo através de validação ou por modelos de robustez (Guisan & Zimmerman, 2000).

Para atestar a eficiência destes modelos, geralmente utiliza-se a matriz de confusão de acertos e erros associados à previsão dos modelos. A matriz de confusão é utilizada para quantificar a qualidade do modelo ajustado (Santa Catarina, 2009).

Outro método de avaliação é a curva ROC (Receiver Operating Characteristic). Por definição, uma curva ROC é a representação gráfica dos pares sensibilidade e especificidade, resultantes da variação do valor de corte ao longo de um eixo de decisão; a representação gráfica assim resultante é designada por curva ROC no plano unitário (Braga, 2000).

Um conceito primordial para este trabalho é o conceito de registro zero, ou ausência, que significam locais onde pesquisadores procuraram por indivíduos da espécie estudada, mas não encontraram registro desta espécie no local, ou seja, esta espécie está ausente (Engler *et al.*, 2004). Dados de ausência são difíceis de obter (Santa Catarina, 2009), pois em um dado local pode ser registrada a ausência da espécie por diferentes motivos: a) a espécie não pode ser detectada, embora esteja presente na área; b) por razões históricas e/ou geológicas, a espécie está ausente, porém as condições ambientais permitem que a espécie se adeque; c) as condições ambientais são realmente inadequadas para a espécie (Phillips *et al.*, 2006). Este tipo de dado é de certa forma, precioso, porém escasso. Alguns autores vêm contornando este tipo de problema gerando dados de pseudo-ausência simulados para a modelagem (Engler *et al.*, 2004; Santa Catarina, 2009).

Pontos de ausência por muitas vezes podem ser muito dispendiosos e difíceis de encontrar, por demandarem muitos custos para a pesquisa e, apesar de serem dados considerados extremamente preciosos (Santa Catarina, 2009), não há vantagem em iniciar uma pesquisa para coletá-los (Wisn & Guisan, 2009). Sendo assim, pontos de pseudo-ausência são alternativas plausíveis para a modelagem quando não há pontos de ausência disponíveis, pois estes assumem o papel dos pontos de ausência (Mateo *et al.*, 2010).

Os pontos de pseudo-ausência são geralmente utilizados em modelos que contêm somente pontos de presença (Chefaoui & Lobo, 2008). A escolha destes pontos reflete diretamente o resultado do modelo de distribuição de espécies (VanDerWal *et al.*, 2009).

Assim, a necessidade de gerar pontos de pseudo-ausência torna-se necessária, pois alguns métodos utilizados na modelagem ecológica precisam de dados de ausência ou de dados de *background* para gerar e/ou avaliar SDMs. Embora alguns SDMs consigam realizar a modelagem somente com dados de presença, a falta de dados confiáveis de ausência afeta o desempenho destes métodos (Mateo *et al.*, 2010).

O foco deste trabalho encontra-se na alteração do sistema SAHGA SDM, que emprega o algoritmo SAHGA para modelar a distribuição potencial das espécies. O diferencial deste sistema está na capacidade de construir SDM que considerem os relacionamentos espaciais nos dados de entrada representados numa GPM (Santa Catarina, 2009).

Portanto, a geração de pontos de pseudo-ausência traz consigo algumas questões:

1. É possível gerar pontos de pseudo-ausência, respeitando determinados critérios, de forma automática?
2. De que forma a geração destes pontos impacta no resultado do modelo de distribuição de espécies?
3. O resultado apresentado pelo modelo utilizando pontos de pseudo-ausência difere positivamente do resultado apresentado pelo método gerado aleatoriamente?

A primeira questão refere-se à necessidade de gerar pontos de pseudo-ausência em um sistema de modelagem de distribuição de espécies, respeitando uma série de critérios para a geração destes pontos. A implementação de diferentes estratégias, para geração automática de pontos de pseudo-ausência, torna-se útil e interessante, pois reduz o tempo do agente que realiza a tarefa de modelagem conferindo-lhe ainda a possibilidade de escolher a estratégia a empregar na geração destes pontos.

A segunda questão remete à necessidade de avaliar o impacto da geração destes pontos na modelagem de distribuição de espécies, mais especificamente no sistema SAHGA SDM.

Por fim, há a necessidade de uma comparação dos resultados para responder a terceira questão. Isso será possível através da comparação dos modelos ajustados com os diferentes métodos de geração de pontos de pseudo-ausência, tendo como base de comparação o método aleatório.

1.1. Hipótese Central

A avaliação dos resultados apresentados pelos SDM, através da matriz de confusão e da curva ROC, utiliza dados de ausência das espécies; ou seja, de pontos onde a espécie não é encontrada. Entretanto, pontos com esta característica são raros pois, geralmente, dados coletados em herbários referem-se unicamente à presença da espécie.

Então, para efetuar a avaliação da qualidade dos modelos gerados, devem ser gerados pontos de pseudo-ausência. Estes pontos são, de modo geral, escolhidos pelo agente que está

realizando a modelagem da distribuição da espécie, por um especialista no conhecimento das características da espécie ou através de algoritmos embutidos nos sistemas de modelagem e que funcionam como "caixas-pretas".

A geração de pontos de pseudo-ausência têm se mostrado uma alternativa satisfatória para suplantiar a falta de pontos de ausência. Muitos estudos têm utilizado a geração de pontos de pseudo-ausência para avaliar o desempenho de diversos modelos de distribuição de espécies (Hirzel et al., 2001; Zaniwski et al., 2002; Brotons et al., 2004; Engler et al., 2004; Pearce & Boyce, 2006; Chefaoui & Lobo, 2008).

A hipótese central deste trabalho pode ser assim descrita: "É possível gerar pontos de pseudo-ausência automaticamente, baseando-se em alguns critérios, e incorporá-los a um sistema de modelagem de distribuição de espécies, fazendo com que os resultados apresentados por estes modelos sejam mais precisos."

Para validar esta hipótese, iniciou-se um estudo sobre como seriam gerados os pontos de pseudo-ausência e três métodos foram abordados: pontos gerados randomicamente; pontos gerados baseando-se em um critério de um *buffer* de distância; e pontos gerados utilizando o conceito de envelope bioclimático.

Estes métodos serão incorporados ao sistema SAHGA SDM e serão avaliados, visando identificar aquele que proporcione efeitos positivos sobre os resultados oferecidos por este sistema de modelagem de distribuição de espécies.

Como objetivo secundário pode-se destacar a elaboração da interface gráfica para o sistema SAHGA SDM, já que as ações executadas no sistema são realizadas via console. Como ferramenta para o desenvolvimento da interface, propõe-se a utilização do kit de desenvolvimento de sistemas (SDK) Qt, por ser um SDK multiplataforma e de fácil integração com o sistema já desenvolvido.

Outro objetivo secundário deste trabalho é a avaliação da correlação entre dados ambientais. O foco desta etapa consiste em avaliar o efeito na qualidade do modelo, quando um layer, considerado redundante, for descartado na construção do modelo.

1.2. Organização do Texto

No capítulo 2 tem-se o referencial teórico, que visa apresentar conceitos necessários à compreensão do trabalho desenvolvido. Dividiu-se o capítulo 2 em duas seções. A primeira

seção abordará os conceitos referentes aos SDMs existentes atualmente; a segunda seção apresentará os conceitos básicos sobre a tecnologia de desenvolvimento de interfaces Qt e sua contribuição para a usabilidade do sistema SAHGA SDM.

O capítulo 3 detalha os três métodos desenvolvidos para a geração automática de pontos de pseudo-ausência (aleatório, distância e envelopes bioclimáticos), como estes funcionam e como foram implementados.

O capítulo 4 apresenta a aplicação do algoritmo SAHGA SDM utilizando os três métodos de geração automática de pontos de pseudo-ausência implementados e um estudo de caso, onde se modelou a distribuição potencial da espécie *Thalurania furcata boliviana*.

Finalmente, o quinto capítulo apresenta as conclusões do trabalho.

Capítulo 2

Referencial Teórico

Este capítulo tem a intenção de apresentar os conceitos básicos necessários ao entendimento do trabalho desenvolvido. Os conceitos apresentados originam-se de diversas áreas de conhecimento como a bioecologia, a computação aplicada e a geoinformática.

A primeira seção discutirá os principais métodos e sistemas responsáveis pela modelagem e distribuição de espécies, com enfoque no sistema SAHGA SDM e no BIOCLIM.

A segunda seção tratará da questão da interface gráfica de usuário (GUI), trazendo consigo seu histórico, as vantagens da utilização de uma GUI e uma breve análise do *framework* para desenvolvimento de interfaces Qt, assim como o sistema que inspirou o desenvolvimento da interface do SAHGA SDM.

2.1. Modelos de Distribuição de Espécies (SDMs)

Diversos fatores levaram os pesquisadores a estudar diferentes métodos para distribuir espécies, tais como a luta pela preservação de uma espécie em extinção, o controle de espécies invasoras, a predação entre espécies e a interferência humana diminuindo os nichos potenciais. O método abordado neste trabalho é a modelagem de distribuição de espécies (SDM - *Species Distribution Models*) que, de um modo geral, pode ser utilizada para prever e avaliar o impacto do uso acelerado da terra e da mudança do meio-ambiente na distribuição de organismos (Guisan & Zimmerman, 2000).

Porém, a natureza é muito complexa e heterogênea para ser predita em todos os aspectos de tempo e espaço através de um modelo simples (Guisan & Zimmermann, 2000). Para tanto, a modelagem de distribuição de espécies se ancora em três pilares, a saber: generalidade, realidade e precisão (Guisan & Zimmermann, 2000).

Richard Levins em seu “*The strategy of model building in population ecology*” destacou que apenas duas, das três desejáveis propriedades de modelos (generalidade, realidade e precisão) podem ser implementadas simultaneamente, enquanto uma terceira seria sacrificada.

A consequência disso é a distinção de três diferentes tipos de modelos. O primeiro grupo de modelo diz respeito à generalidade e precisão, estes são chamados de analíticos e são designados a prever uma resposta precisa utilizando uma realidade simplificada e limitada. O segundo grupo de modelos foi feito para ser realista e geral e são chamados de mecanicistas e suas previsões baseiam-se em relacionamentos reais de causa e efeito. O terceiro grupo de modelos especializa-se na precisão e realidade, para sacrificar a generalidade, e é chamado de empírico ou fenomenológico e, como o próprio nome diz, é baseada na observação e amostragem do mundo. A classificação dos modelos matemáticos aplicados na ecologia encontra-se na figura 2.1.

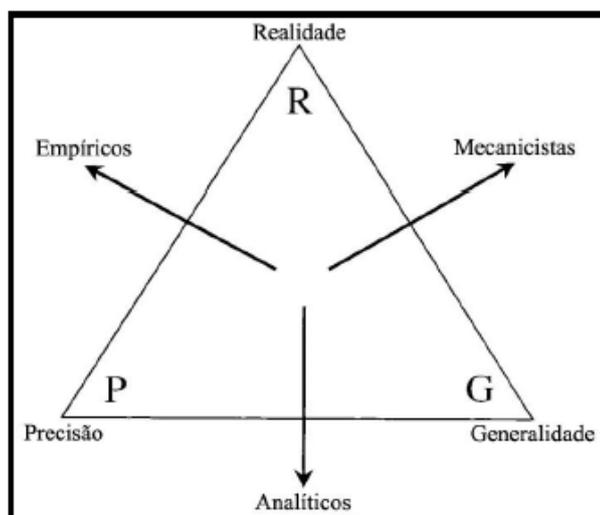


Figura 2.1 - Classificação dos modelos matemáticos aplicados em ecologia
Fonte: Adaptado de Guisan e Zimmerman (2000)

Modelos de distribuição de espécies são modelos empíricos que relacionam observações de ocorrência de uma espécie com variáveis de predição de ambiente, baseadas estatisticamente ou teoricamente (Guisan & Zimmerman, 2000).

Todos os estudos que envolvem SDM possuem três componentes básicos: primeiramente há um conjunto de dados que contém as variáveis explicativas; após, há um modelo matemático que relaciona a espécie com a variável explicativa; finalmente, há a avaliação da utilidade do modelo através de métodos de robustez ou por avaliação (Guisan & Zimmerman, 2000). Isto é representado na figura 2.2.

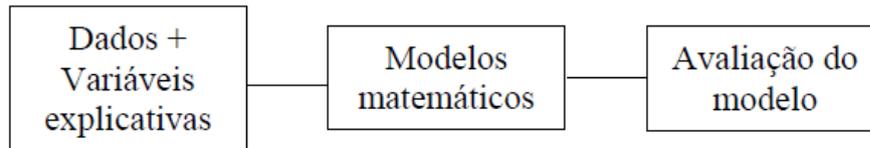


Figura 2.2: Elementos essenciais na modelagem de distribuição de espécies

Fonte: Iwashita, 2007

Ecólogos e biogeógrafos utilizam-se destes modelos para estudar problemas teóricos e práticos tais como a luta pela preservação de uma espécie em extinção (Godown & Peterson, 2000), o controle de espécies exóticas ou pragas (Sutherst *et al.*, 2000), a predação entre espécies, a interferência humana diminuindo os nichos potenciais e o estabelecimento de áreas apropriadas para translocações/relocações ou cultivo (Cunningham *et al.*, 2002), entre outros problemas. Alguns exemplos de utilização de modelos de distribuição de espécies são mostrados na tabela 2.1:

Tabela 2.1: Alguns possíveis usos de SDMs na ecologia e biologia conservativa

Tipos de Uso	Referências
Quantificar o nicho ambiental das espécies.	Austin <i>et al.</i> (1990), Vetaas (2002)
Teste de hipóteses biogeográficas, ecológicas e evolucionárias.	Leathwick (1998), Anderson <i>et al.</i> (2002), Graham <i>et al.</i> (2004b)
Avaliar a invasão e proliferação de espécies.	Beerling <i>et al.</i> (1995), Peterson (2003)
Avaliar o impacto do clima, uso da terra e outras mudanças do ambiente na distribuição de espécies	Thomas <i>et al.</i> (2004), Thuiller (2004)
Sugerir lugares não pesquisados de alto potencial para espécies em extinção	Elith & Burgman (2002), Raxworthy <i>et al.</i> (2003), Engler <i>et al.</i> (2004)
Incentivar planos de gerenciamento apropriados para a recuperação de espécies e o mapeamento de lugares adequados para reintrodução destas espécies	Pearce & Lindemeyer (1998)
Incentivar o planejamento de conservação e seleção de reservas	Ferrier (2002), Araújo <i>et al.</i> (2004)
Modelar a reunião de espécies a partir de predição de espécies individuais	Leathwick <i>et al.</i> (1996), Guisan & Theurillat (2000), Ferrier <i>et al.</i> (2002)
Construção de regiões bio ou ecogeográficas	Sem exemplo publicado
Melhorar o cálculo da distância ecológica entre fragmentos de meta-populações e modelos de fluxo gênico	Sem exemplo publicado

Fonte: Adaptado de Guisan & Thuiller (2005).

Os SDMs estudados neste trabalho se utilizam de modelagem matemática juntamente com ferramentas computacionais para prever a ocorrência de determinada espécie através da geração de superfícies temáticas denominadas *layers*, indicando presença ou ausência desta espécie em questão (Guisan & Thuiller, 2005; Santa Catarina, 2009). A figura 2.3 apresenta a estrutura de um sistema para geração de SDM.

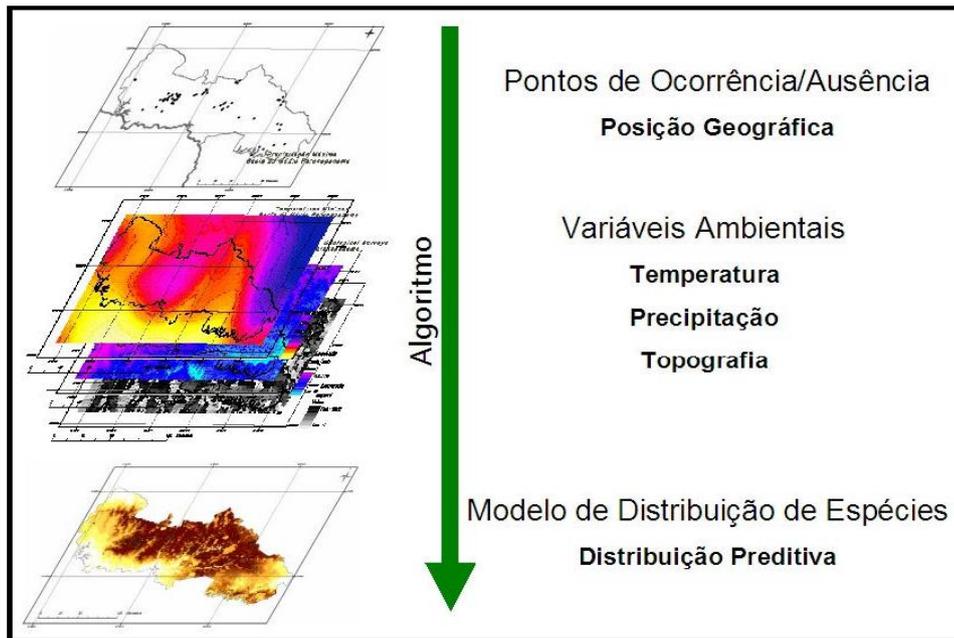


Figura 2.3. Estrutura geral de um sistema para geração de SDM
 Fonte: Siqueira (2005)

O procedimento ideal para a construção de um de um SDM é composto por seis passos, a saber: conceituação do modelo; preparação dos dados; ajuste do modelo; avaliação do modelo; predições de espaço; e a avaliação da aplicabilidade do modelo (Guisan & Thuiller, 2005).

Na etapa de conceituação são tomadas as decisões mais importantes, tanto que a primeira etapa pode ser dividida em duas sub etapas: a fase de dados e teoria e a fase de modelagem de métodos. Na fase de modelagem de métodos, que pode ser considerada a mais importante, são identificados os métodos mais apropriados para a modelagem de espécies e escolhidos os métodos estatísticos necessários para avaliar a acurácia do modelo (Pearce & Ferrier, 2000). A tabela 2.2 mostra os principais SDMs existentes e que métodos de modelagem eles utilizam.

Tabela 2.2: SDMs publicados, artigo de referência, métodos de modelagem utilizados e link www (quando disponível)

Ferramenta	Referência	Métodos Implementados	URL
BIOCLIM	Busby (1991)	Envelope Climático	http://www.arcscripts.esri.com
ANUCLIM	Veja BIOCLIM	Envelope Climático	http://www.cres.anu.edu.au/outputs/anuclim.php
BAYES	Aspinall (1992)	Aproximação Bayesiana	Disponível a critério do autor
BIOMAPPER	Hirzel <i>et al.</i> (2002)	Análise do fator nicho ecológico	http://www.unil.ch/biomapper
BIOMOD	Thuiller (2003)	GLM, GAM, CART, ANN	Disponível a critério do autor
DIVA	Hijmans <i>et al.</i> (2001)	Envelopes climáticos	http://www.diga-gis.org
DOMAIN	Carpenter <i>et al.</i> (1993)	Envelopes climáticos	http://www.cifor.cgiar.org/docs/_ref/research_tools/domain/index.htm
ECOSPAT	Dados não publicados	GLM, GAM	http://www.ecospat.unil.ch
GARP	Stockwell & Peters (1999)	Algoritmo genéticos	http://www.ligemapper.org/desktopgarp
GDM	Ferrier <i>et al.</i> (2002)	GDM	Disponível a critério do autor
GRASP	Lehmann <i>et al.</i> (2002)	GLM, GAM	http://www.cscf.ch/grasp
MAXENT	Phillips <i>et al.</i> (2005)	ME	Disponível a critério do autor
SPECIES	Pearson <i>et al.</i> (2002)	ANN	Disponível a critério do autor
Disperse	Carey (1996)	Envelopes climáticos	Disponível a critério do autor
Shift	Iverson <i>et al.</i> (1999)	CART	Disponível a critério do autor
ANN = redes neurais artificiais; CART = árvores de classificação e regressão; GAM = modelos aditivos generalizados; GDM = modelagem de dissimilaridade generalizada; ME = máxima entropia.			

Fonte: Adaptado de Guisan & Thuiller (2005).

Como se pode perceber há uma grande variedade de técnicas de modelagem para explorar a relação entre a resposta e as variáveis ambientais preditivas (Iwashita, 2007). Este grupo de modelos pode ser subdividido em duas classes: modelos que utilizam apenas registro de

presença e modelos que utilizam dados de presença e ausência da espécie (Elith & Graham, 2006).

Como dito anteriormente, o dado de ausência é um dado particularmente precioso para a modelagem, pois indica a área onde verdadeiramente a espécie não ocorre. Para alguns SDMs este tipo de dado é fundamental, pois limitam a área de ocorrência da espécie, diminuindo a ocorrência de falsos positivos (Iwashita, 2007).

O foco deste trabalho se encontra em um SDM que se utiliza de dados de presença e ausência da espécie: o SAHGA SDM e o BIOCLIM, que serão apresentados a seguir.

2.1.1. SAHGA SDM

O sistema SAHGA SDM é um sistema que emprega o algoritmo SAHGA para modelar a distribuição potencial de espécies (Santa Catarina, 2009). O que diferencia o SAHGA SDM de outros sistemas é sua capacidade de construir modelos de distribuição de espécies que considerem os relacionamentos espaciais presentes nos dados de entrada, representando-os através de uma matriz de proximidade generalizada (GPM) (Santa Catarina, 2009). A estrutura geral do sistema SAHGA SDM é apresentada na figura 2.4.

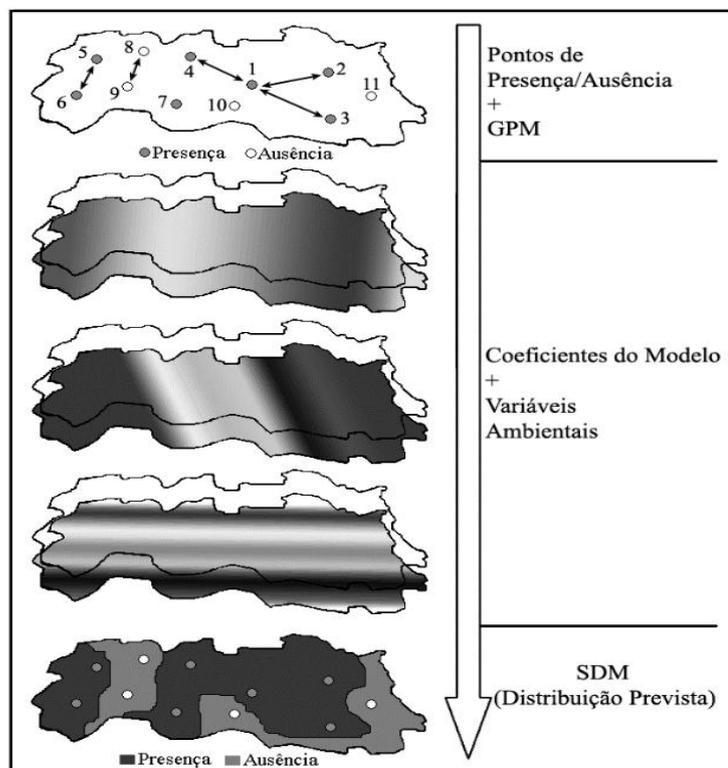


Figura 2.4: Estrutura Geral do Sistema SAHGA SDM

Fonte: Santa Catarina, 2009

Associados às variáveis ambientais encontram-se os coeficientes a elas associados. Estes coeficientes mensuram o efeito das variáveis ambientais sobre a distribuição prevista da espécie. À matriz de proximidade generalizada, associam-se dois objetivos, a saber: incorporar os relacionamentos existentes entre os pontos e a representação do conhecimento pré-existente sobre os elementos naturais e artificiais presentes no espaço, que impactam significativamente na distribuição potencial da espécie modelada (Santa Catarina, 2009). Estes elementos podem ser representados como rios, cadeias de montanhas, lagos, etc.

Estes relacionamentos podem ser vistos na figura 2.4 como as setas que mostram a distância entre um ponto e outro. Os elementos naturais pré-conhecidos podem ser quantificados como pesos dentro da matriz de proximidade generalizada (Santa Catarina, 2009). Quanto maior a distância entre os pontos, menor o peso atribuído e vice-versa, conforme exemplificado na figura 2.5.

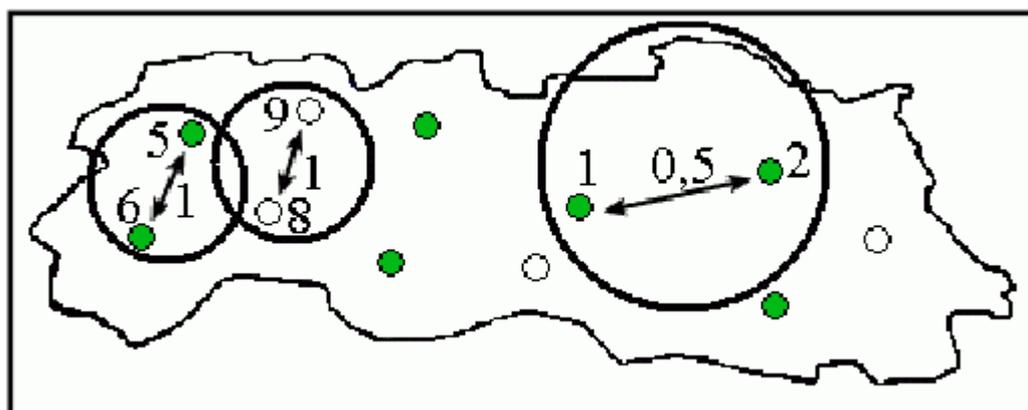


Figura 2.5: Regra Empregada na Construção da GPM
Fonte: Santa Catarina (2009)

Na matriz de proximidade generalizada os pesos são calculados a partir de relações espaciais no espaço absoluto ou no espaço relativo (Aguiar *et al.*, 2003; Pedrosa, 2003). Uma GPM é composta por um conjunto de objetos geoespaciais, um grafo e uma matriz de proximidade (Santa Catarina, 2009). Os objetos geoespaciais são representados por células regulares ou polígonos, dependendo da representação espacial utilizada; O grafo é constituído por um conjunto de vértices e arestas, onde os vértices representam um objeto (célula ou polígono) e as arestas representam o relacionamento de vizinhança existente entre dois vértices. A matriz de proximidade é composta por um conjunto de elementos W_{ij} que serve para indicar o quanto dois objetos estão próximos

Como entrada, o SAHGA SDM utiliza pontos amostrais de presença ou ausência da espécie, com seus relacionamentos espaciais, e o conjunto de layers geográficos que representam as variáveis ambientais que podem delimitar a sobrevivência da espécie (Santa Catarina, 2009).

2.1.2. BIOCLIM

O algoritmo BIOCLIM, como dito anteriormente, implementa o conceito de envelope bioclimático (Nix, 1986; Busby, 1991). Para cada variável ambiental associada aos pontos de presença de uma espécie, o BIOCLIM calcula a média e o desvio padrão. O conceito de envelope surge a partir da definição do intervalo definido para o algoritmo, que pode ser representado por $[m - c*s, m + c*s]$, onde m pode ser caracterizado como a média, s como o desvio padrão e c é uma constante que representa o ponto de corte (Santa Catarina, 2009).

Além disso, a cada variável ambiental associada aos pontos de presença, há também os limites superiores e inferiores que correspondem aos valores máximos e mínimos associados ao conjunto de pontos de presença da espécie (Santa Catarina, 2009).

Sendo assim, cada ponto pode ser considerado:

- A) Adequado: o ponto pode ser considerado adequado quando todos os valores das variáveis ambientais encontram-se dentro do envelope estabelecido;
- B) Marginal: o ponto pode ser considerado marginal se ao menos um dos valores das variáveis encontra-se fora do envelope estabelecido, porém ainda encontra-se dentro dos valores máximos e mínimos em todos os envelopes;
- C) Inadequado: o ponto pode ser considerado inadequado quando ao menos um valor das variáveis ambientais encontra-se fora do envelope estabelecido. Difere do marginal quando esta variável também se encontra fora dos limites mínimo e máximo em pelo menos um dos envelopes bioclimáticos.

2.1.3. Avaliação de Modelos

Neste tópico serão discutidas algumas técnicas para a avaliação de SDM. Serão apresentadas as duas técnicas utilizadas no SAHGA SDM que são a matriz de confusão e a curva ROC.

2.1.3.1. Matriz de Confusão

O método de avaliação mais empregado nos sistemas de geração de SDM é a matriz de confusão de acertos e erros associados à previsão dos modelos (Santa Catarina, 2009). Esta matriz é utilizada para mensurar a qualidade do modelo ajustado e o seu formato é apresentado na figura 2.6.

Previsão (Modelo)	Amostras	
	Presente	Ausente
Presente	VP	FP
Ausente	FN	VN

Figura 2.6: Matriz de Confusão

Para entender a matriz de confusão, primeiramente devem-se entender as suas representações. Os valores de VP (Verdadeiro Positivo) e VN (Verdadeiro Negativo) são predições corretas, ou seja, quando uma espécie se encontra verdadeiramente naquele local, ou não. Os valores de FP (Falso Positivo) e FN (Falso Negativo) são erros de predição. Erros do tipo FP (Falso Positivo) são considerados erros de comissão, ou de superestimativa. Já os erros do tipo FN (Falso Negativo) são considerados erros de omissão. A figura 2.7 apresenta a ocorrência destes dois tipos de erro.

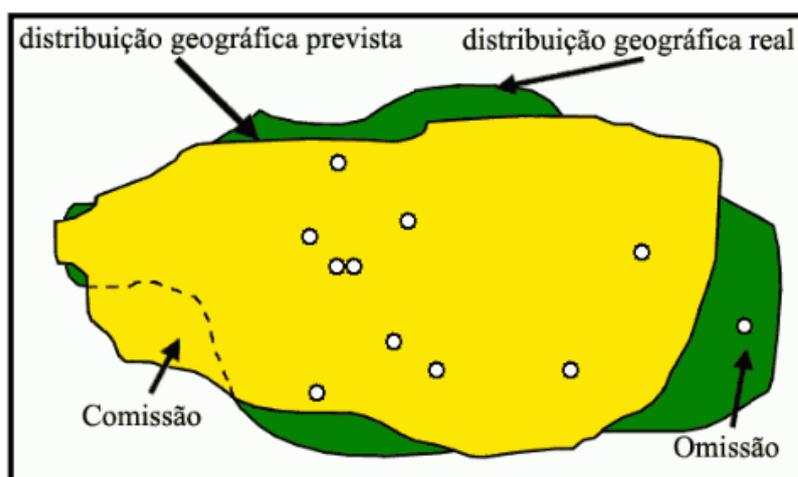


Figura 2.7: Simulação de um modelo com erros de omissão e comissão
Fonte: Adaptado de Siqueira, (2005)

Erros de comissão não são considerados graves, por diversos motivos: as condições ambientais da área são favoráveis, porém não se encontrou registro da espécie na área; as condições da área são favoráveis a espécie, porém fatores topológicos e/ou biológicos impedem que a espécie ocupe a área estudada; a área é mesmo inadequada, situação em que ocorre um erro do tipo falso positivo.

Em contrapartida, os erros de omissão podem ser considerados graves, pois indicam lugares onde a espécie é encontrada, porém está sendo predito como ausente.

Para avaliar um SDM, podem-se obter algumas métricas da matriz de confusão, estas são apresentadas na tabela 2.3.

Tabela 2.3: Métricas derivadas da matriz de confusão

Métrica	Cálculo
Acurácia	$(VP + VN) / (VP + FP + VN + FN)$
Sensibilidade	$VP / (VP + FN)$
Especificidade	$VN / (FP + VN)$
Taxa de Falso Positivo (comissão)	$FP / (FP + VN)$
Taxa de Falso Negativo (omissão)	$FN / (VP + FN)$
Coeficiente de Correlação de Matthews	$\frac{(VP \cdot VN - FP \cdot FN)}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$
Índice Kappa	$\frac{(VP + VN) - \frac{((VP + FN)(VP + FP) + (FP + VN)(FN + VN))}{N}}{N - \frac{((VP + FN)(VP + FP) + (FP + VN)(FN + VN))}{N}}$

Fonte: Adaptado de Fielding & Bell (1997)

A acurácia mede o acerto global do modelo. A sensibilidade mede a probabilidade de uma dada amostra ser classificada corretamente como dado de presença da espécie. Em contrapartida, a especificidade mede a probabilidade de uma dada amostra ser classificada corretamente como ausência da espécie (Fielding & Bell, 1997). O coeficiente de correlação de Matthews considera todas as informações que vêm da matriz de confusão. A partir destas informações, o retorno dado por ele é um valor entre -1 e 1. Um coeficiente de correlação de Matthews com valor igual a 1 indica uma predição perfeita, com o valor igual a 0 tem-se a indicação de uma predição aleatória e quando há o valor igual a -1 a predição foi realizada de

forma inversa. Esta métrica advinda da matriz de confusão é considerada a métrica mais sugerida para a avaliação de SDM (Baldi *et al.*, 2000; Santa Catarina, 2009).

Outro fator a ser considerado na avaliação é o índice Kappa. O índice Kappa é obtido através da matriz de confusão e pode ser considerado uma boa medida para avaliar o modelo, pois faz uso de todas as informações contidas na matriz de confusão (Iwashita, 2007; Fielding & Bell, 1997). Este índice varia de 0 a 1 e pode ser classificado segundo a tabela 2.4:

Tabela 2.4: Qualidade do Índice Kappa

Qualidade	Índice Kappa
Ruim	0,01 – 0,20
Razoável	0,21 – 0,40
Boa	0,41 – 0,60
Muito Boa	0,61 – 0,80
Excelente	0,81 – 1,00

Fonte: Iwashita, 2007

Contudo, deve-se ter cuidado com o resultado oferecido pelo Kappa, pois ele pode levar a conclusões equivocadas (Pontius Jr., 2000). Para garantir medidas apropriadas de desempenho, Pontius e Schneider (2001) indicam o uso do gráfico Receiver Operating Characteristic (Curva ROC) que será explicada a seguir.

2.1.3.2. Curva ROC

Outra medida utilizada para avaliação de SDM é a curva ROC (Receiver Operating Characteristic). A curva ROC é representada num gráfico de sensibilidade x (1 – especificidade). Ou, lembrando os conceitos das matrizes de confusão, Taxa de VP x Taxa de FP. Um exemplo de gráfico utilizando a curva ROC é apresentado na figura 2.8.

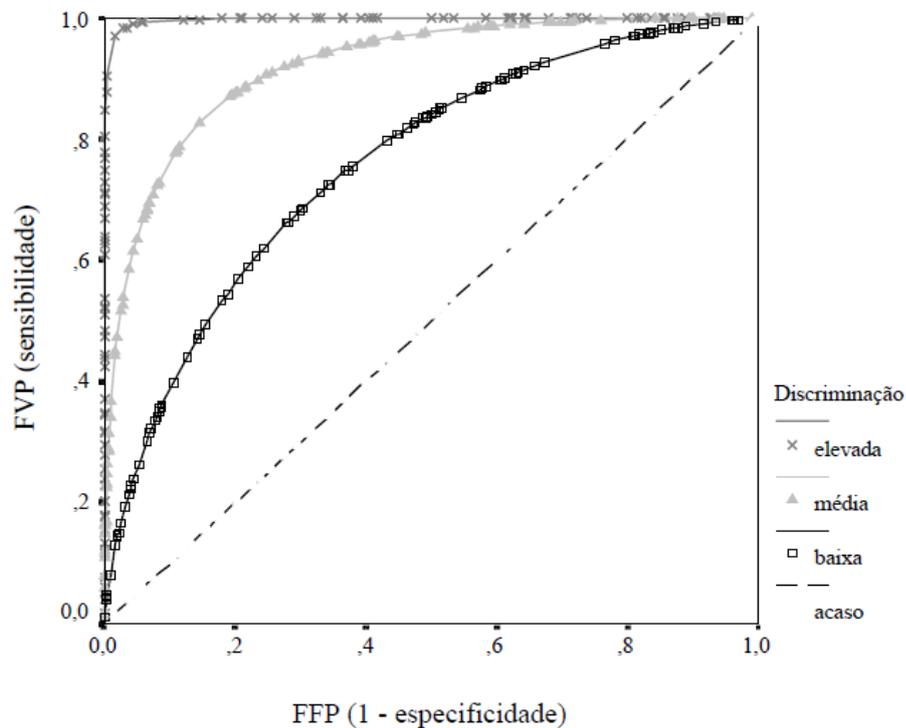


Figura 2.8: Curvas ROC para três graus de capacidade de discriminação
 Fonte: Santa Catarina, 2009 (Adaptado de Braga, 2000)

A área sob a curva (AUC) é a medida utilizada para ratificar a qualidade da curva. Quanto mais esta área se aproximar de 1, melhor o desempenho, sendo que $AUC = 0,5$ equivale a uma predição aleatória (Braga, 2000).

2.2. Graphic User Interface (GUI)

Como acontece com grande parte dos desenvolvimentos da história da computação, as interfaces gráficas de usuário foram pensadas antes de estarem disponíveis para qualquer máquina (Brenlla, 2005).

A história das interfaces gráficas, de certa forma, iniciou-se em 1968, quando Douglas Englebart, inspirado por Vaneevar Bush em seu “As We May Think” (Como nós pensamos), iniciou o trabalho para construir uma máquina em que seria possível mostrar tanto linhas sólidas, quanto texto. Esta máquina ficou conhecida como NLS (oN-Line System), pelo poder de interação com outras máquinas (A History of the GUI, <http://arstechnica.com/old/content/2005/05/gui.ars/2>, acessado em 15 de julho de 2010).

Esta máquina dispunha de um monitor, um teclado e um *mouse*¹ e era possível editar texto, colaboração online, mensageiro instantâneo, *e-mail* e até videoconferência. A equipe de Englebart continuou trabalhando nestas ideias até que a instituição fechou em 1989 por falta de fundos. Entretanto, as ideias oferecidas por Englebart influenciaram outras empresas a produzirem suas próprias máquinas com interfaces gráficas para os usuários (A History of the GUI, <http://arstechnica.com/old/content/2005/05/gui.ars/2>, acessado em 15 de julho de 2010). A ordem cronológica dos computadores que apresentavam interfaces é exibida na figura 2.9.

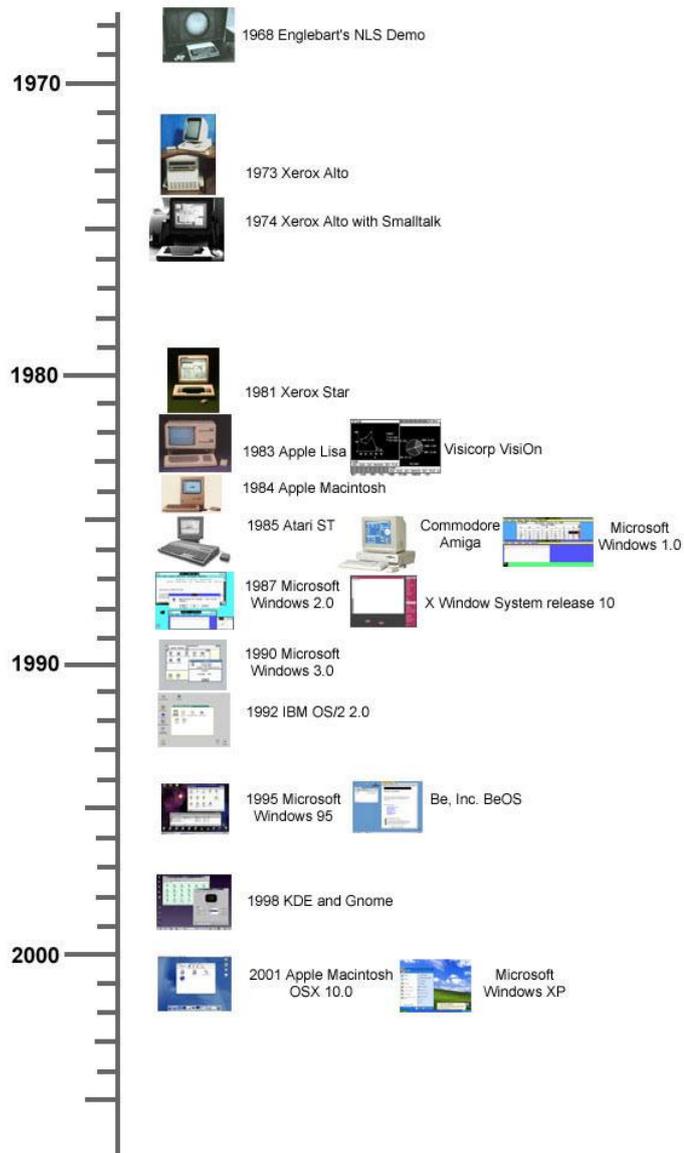


Figura 2.9: Ordem Cronológica das GUI

¹Englebart foi o criador do mouse, porém, na época, o funcionamento interno era diferente, mas o conceito era basicamente o mesmo (Brenlla, 2005).

A utilização de interfaces gráficas tem-se mostrado imprescindível para o manuseio de tecnologias de informação. O uso destas interfaces contribui para a eficácia e eficiência e para a satisfação com as quais usuários específicos atingem objetivos característicos em ambientes particulares (ISO 9241-1, 1997). O objetivo da construção de uma interface sempre se encontra na intuitividade e usabilidade que a mesma pode proporcionar (Oliveira & Santos, 2003). A usabilidade, segundo Nielsen (1993), é composta por cinco atributos, a saber:

1. Aprendizagem: o sistema deve ser de fácil aprendizagem, para que o usuário possa obter rapidamente os resultados;
2. Eficiência: o sistema deve ser eficiente para que, uma vez que o usuário aprendeu a utilizá-lo, este possa usufruir de um alto grau de produtividade;
3. Memorização: a utilização do sistema deve ser de fácil memorização para que, uma vez que o usuário tenha que reutilizar o sistema, a forma de utilizá-lo tenha se fixado em sua memória
4. Tolerância a erros: o sistema deve ter uma alta taxa de tolerância a erros para que, uma vez que usuários inexperientes venham a utilizá-lo, estes cometam um número mínimo de erros possível e que estes erros sejam de fácil recuperação.
5. Satisfação: o sistema deve ser agradável de utilizar de forma que os usuários se sintam subjetivamente satisfeitos ao utilizá-lo.

Para este trabalho procurou-se um sistema de desenvolvimento de interfaces que fosse, preferencialmente, multiplataforma e que fosse de fácil incorporação com o sistema já implementado. Após uma breve pesquisa, optou-se por utilizar o *framework* para desenvolvimento de interfaces Qt, pois este possuía vasto material de referência na *Internet* e seria de fácil integração com o sistema já existente. A seção a seguir apresenta o Qt, assim como suas características.

2.2.1. Qt

O Qt é um *framework* multiplataforma para desenvolvimento de interfaces de usuário (Qt, <http://qt.nokia.com/>, acessado em 16 de julho de 2010). Criado pela empresa norueguesa Trolltech no ano de 1991 e depois comprado pela Nokia em 2008, o Qt é um kit de desenvolvimento de software (SDK) multiplataforma largamente utilizado para o

desenvolvimento de interfaces gráficas de usuário (KDE, Skype, VirtualBox, VLC Media Player, etc.). O ambiente de desenvolvimento *Qt Designer* é mostrado na figura 2.11.

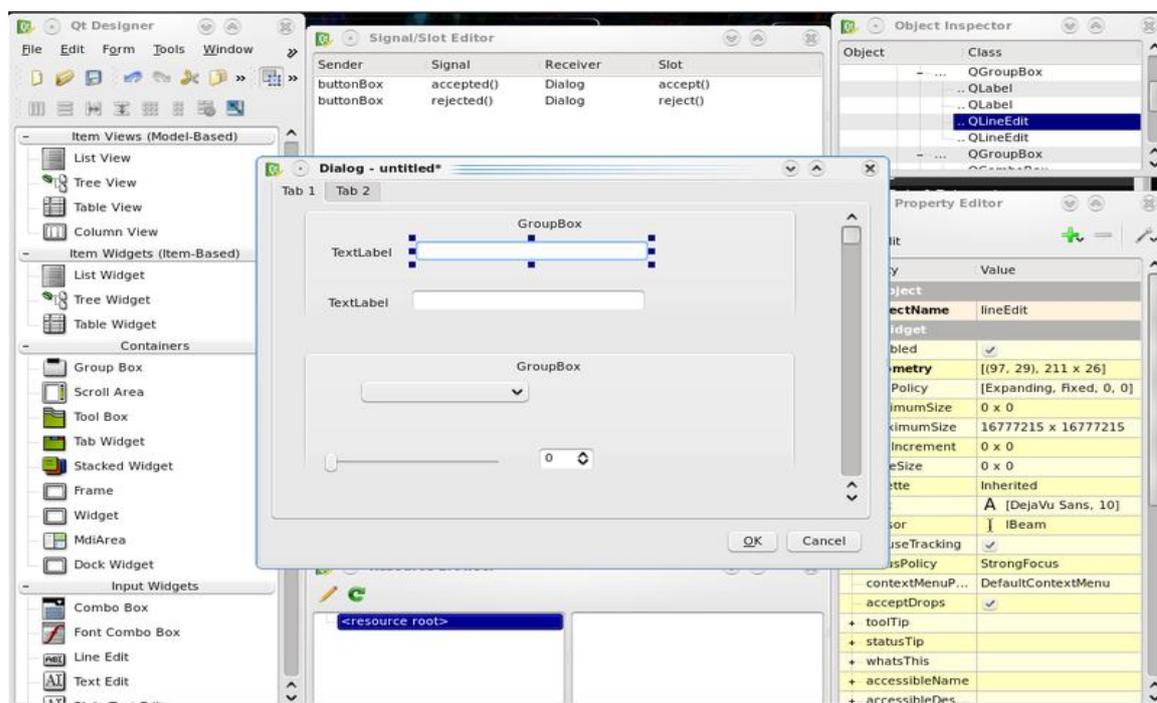


Figura 2.10: Ambiente de Desenvolvimento do Qt Creator

Como dito anteriormente, optou-se por utilizar o SDK *Qt Creator* para o desenvolvimento deste trabalho. Como pode se perceber na figura 2.11, o ambiente de desenvolvimento é bastante intuitivo tornando possível a construção da interface do SAHGA SDM.

Para a implementação da interface procurou-se desenvolver a interface utilizando o sistema openModeller Desktop como exemplo. O openModeller Desktop será explicado na seção a seguir.

2.3. openModeller Desktop

O sistema openModeller Desktop é um sistema de modelagem de espécies financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e desenvolvido pelo Centro de Referência em Informação Ambiental (CRIA), Escola Politécnica da USP (Poli) e o Instituto Nacional de Pesquisas Espaciais (INPE).

A proposta do openModeller Desktop é oferecer um sistema multiplataforma, de código aberto, e que seja de fácil manuseio pelo usuário.

Para a implementação da interface do SAHGA SDM, procurou-se utilizar como referencial operacional o esquema de interfaces proposto pelo openModeller Desktop. Alguns exemplos da funcionalidade da interface de usuário do openModeller Desktop são apresentados nas figuras 2.12 e 2.13.

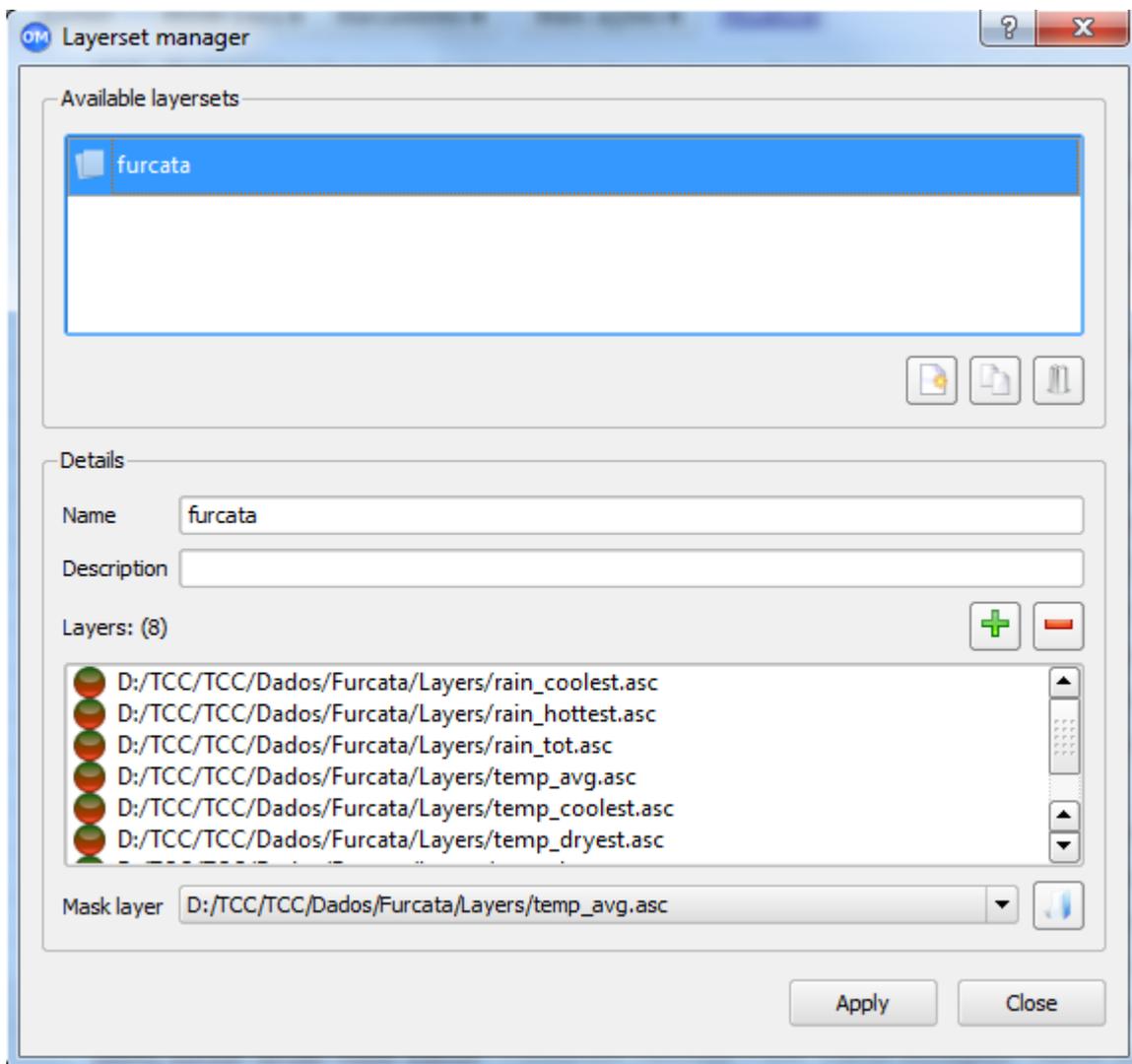


Figura 2.11: Gerenciador de Camadas do openModeller Desktop
Fonte: openModeller, 2010

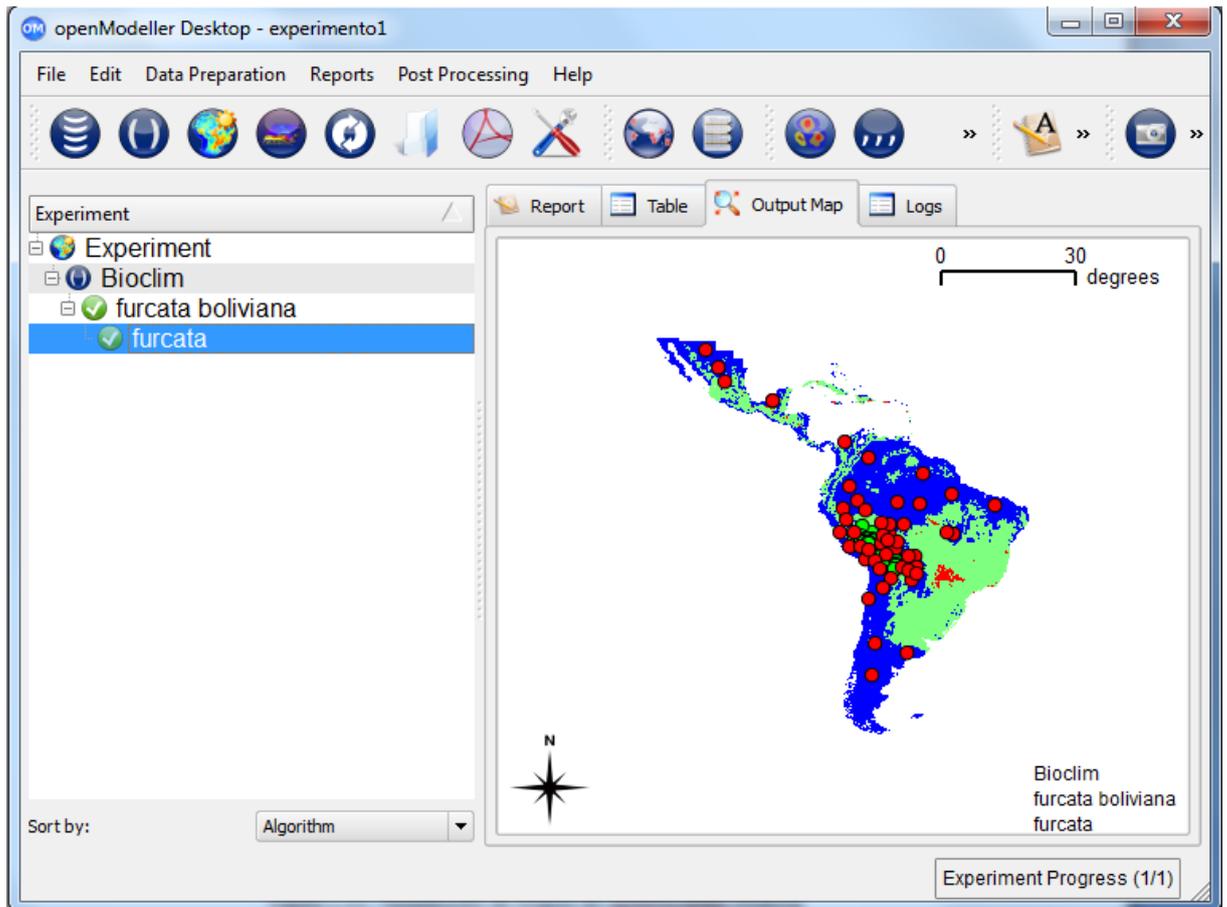


Figura 2.12: Visualizador de Mapas do openModeller Desktop
Fonte: openModeller, 2010

Capítulo 3

Metodologia

O objetivo deste trabalho foi implementar e estudar a influência de três diferentes métodos de geração de pontos de pseudo-ausência sob a qualidade dos modelos de distribuição de espécies ajustados pelo sistema SAHGA SDM.

Os três métodos implementados foram: geração aleatória de pontos (aleatório), geração aleatória de pontos utilizando como critério uma distância mínima aos pontos de presença (distância) e geração de pontos fora dos envelopes bioclimáticos definidos pelo Bioclim.

Como o próprio nome diz, a geração aleatória de pontos consiste em gerar aleatoriamente os pontos de pseudo-ausência; a única restrição será não gerar um ponto de pseudo-ausência sobre um ponto de presença. Isto pode ser demonstrado na figura 3.1, onde os círculos indicam pontos de presença e os pontos de interrogação indicam pontos de pseudo-ausência.

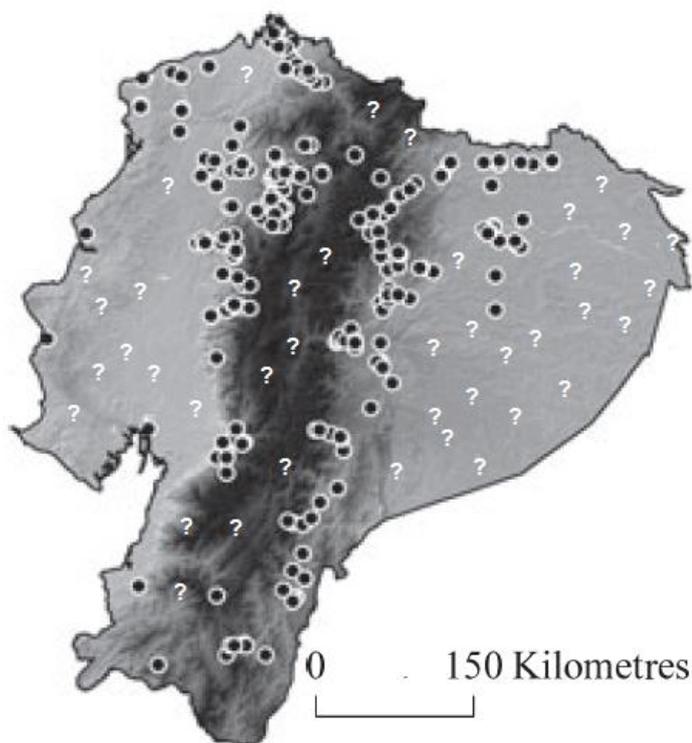


Figura 3.1: Geração aleatória de pontos de pseudo-ausência

Fonte: Adaptado de Mateo et al., 2010

No método que utiliza a distância mínima, define-se uma área circular ao redor dos pontos de presença através da escolha de um raio. Assim, não se podem gerar pontos de pseudo-ausência que estejam contidos na área de influência de qualquer um dos pontos de presença. Como exemplo, pode-se observar a figura 3.2, onde os círculos indicam presença de uma determinada espécie, enquanto que os pontos de interrogação indicam pontos de pseudo-ausência.

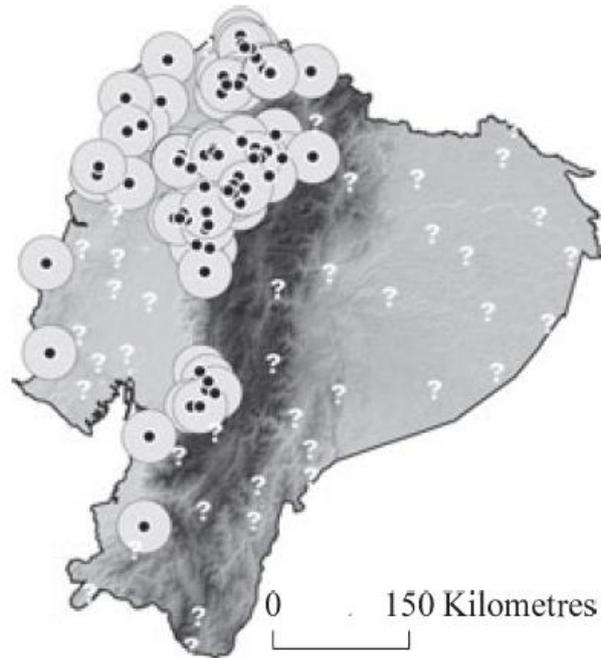


Figura 3.2: Geração de pontos de pseudo-ausência baseado em distância.

Fonte: Mateo et al., 2010

Finalmente, também é possível gerar pontos de pseudo-ausência utilizando os envelopes bioclimáticos, criados pelo modelo Bioclim. Somente poderão ser gerados pontos de pseudo-ausência nas áreas preditas como ausência segundo o modelo. A representação gráfica da geração de pontos de pseudo-ausência pode ser conferida na figura 3.3.

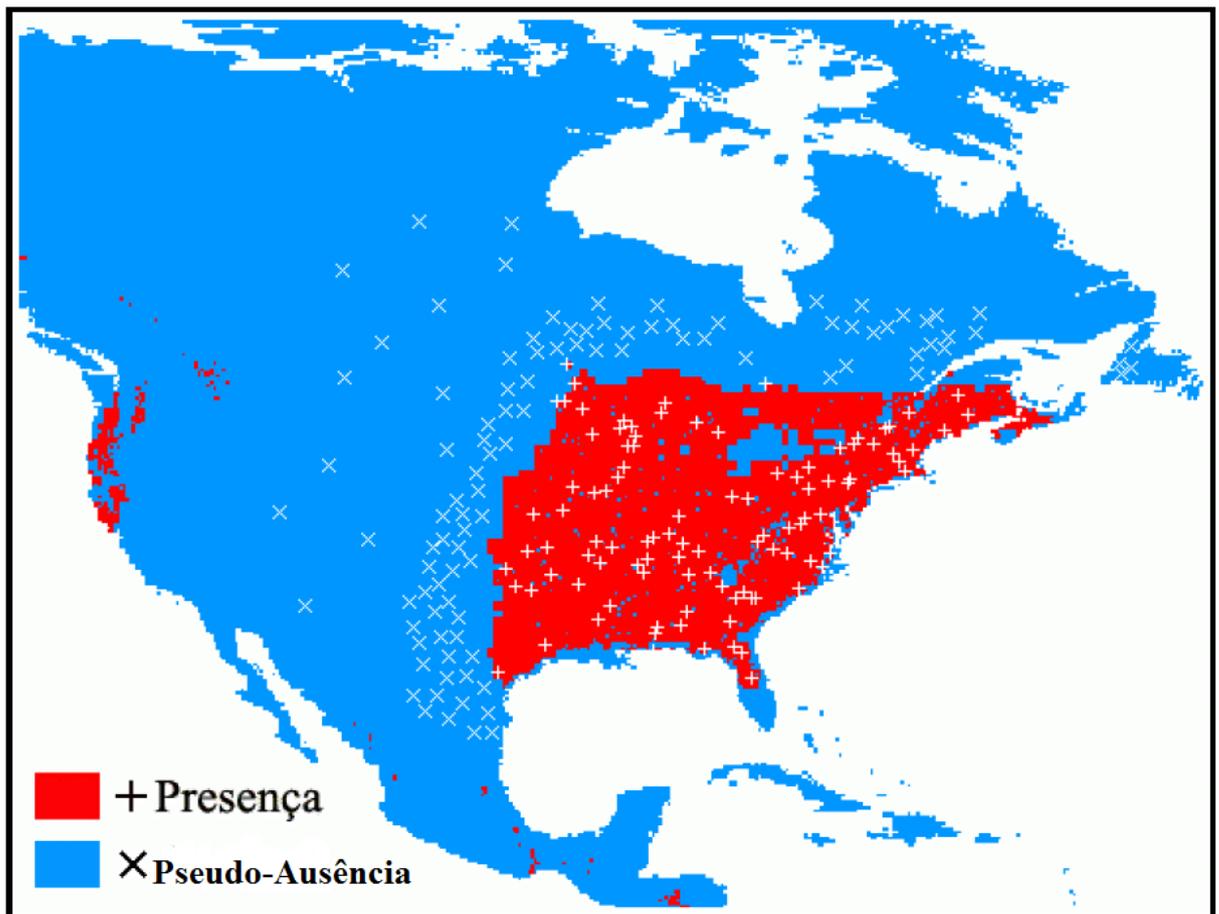


Figura 3.3: Geração de pontos de pseudo-ausência segundo o Bioclim
 Fonte: Adaptado de Santa Catarina (2009)

3.1. Geração dos Pontos de Pseudo-Ausência

Para a etapa de geração dos pontos de pseudo-ausência, utilizados no ajuste e na avaliação de modelos, foram selecionados três métodos: aleatório, distância e envelopes bioclimáticos. Os modelos ajustados pelo sistema serão comparados, tendo como base de comparação os modelos que utilizam pontos de pseudo-ausência gerados através do método aleatório. Os instrumentos utilizados para avaliar a qualidade dos modelos ajustados serão a matriz de confusão e a curva ROC.

Nos próximos tópicos serão abordados, com mais detalhes, as técnicas utilizadas para a geração destes pontos.

3.1.1. Pontos de Pseudo-Ausência gerados pelo Método Aleatório

O primeiro método implementado para a geração dos pontos de pseudo-ausência foi o

aleatório. Este método tem como característica a criação de pontos de pseudo-ausência de forma totalmente randômica, tomando-se apenas duas precauções: gerá-los somente em regiões onde há dados disponíveis e a não sobreposição dos pontos de pseudo-ausência gerados com os pontos de presença.

A figura 3.4 ilustra o funcionamento deste algoritmo.

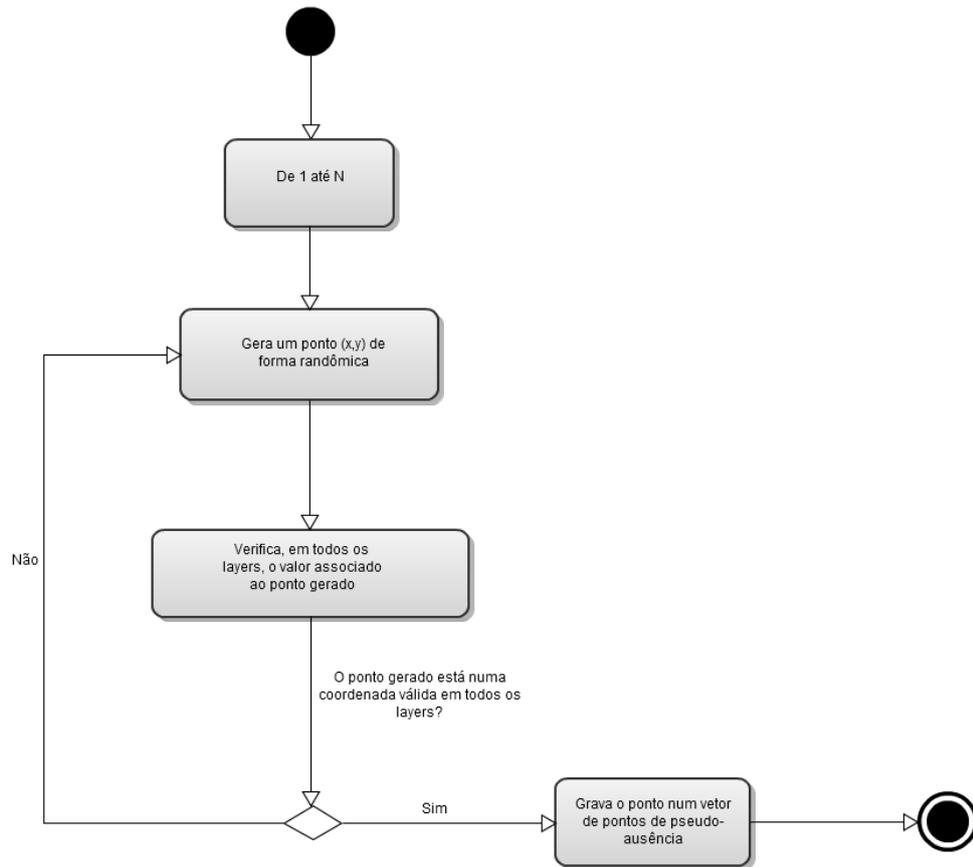


Figura 3.4: Geração de Pontos de Pseudo-Ausência Pelo Método da Aleatoriedade

O funcionamento do algoritmo é bastante simples. O ponto é gerado aleatoriamente e, depois, verificam-se os valores associados à posição deste ponto em todos os *layers*. Caso o ponto tenha sido gerado em região com ausência de dados ele é descartado e um novo ponto é gerado; caso contrário o ponto é aceito como ponto de pseudo-ausência válido.

3.1.2. Pontos de Pseudo-Ausência Gerados pelo Método da Distância

Outro método selecionado para a geração dos pontos de pseudo-ausência foi o da distância. A principal característica deste método, que o difere do método aleatório, é a possibilidade de criar pontos utilizando um *buffer* de distância dos pontos de presença. A expectativa é que os modelos ajustados, com pontos de pseudo-ausência gerados por este método, sejam melhores por garantir o isolamento dos pontos de presença e, por conseqüência, a sua área de influência.

O funcionamento do algoritmo para a geração destes pontos é muito similar ao método aleatório: gera-se um ponto de forma randômica e verifica-se, para todos os *layers* correspondentes às variáveis ambientais, se o ponto encontra-se dentro de uma região permitida, ou seja, se o ponto não tenha sido criado numa região onde não há dados disponíveis. Caso esta primeira condição seja atendida uma nova comparação é feita: calcula-se a distância do ponto gerado com todos os pontos de presença fornecidos pelo usuário; se a distância, para todos os pontos de presença, for maior ou igual à distância fornecida pelo usuário, o ponto de pseudo-ausência é um ponto válido. A figura 3.5 ilustra esta idéia.

Deve-se ressaltar que a distância, equivalente ao raio do *buffer*, é fornecida pelo usuário, tendo por base seu conhecimento e experiência sobre a mobilidade e/ou capacidade de disseminação da espécie modelada.

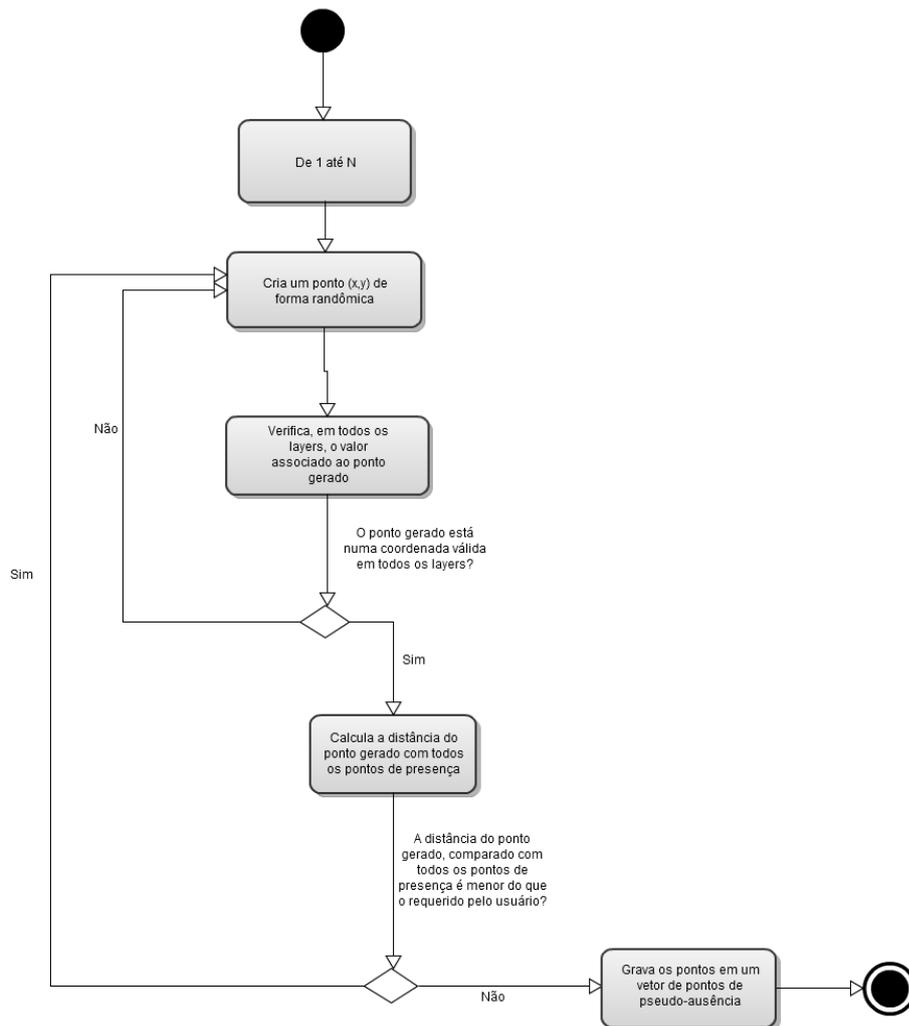


Figura 3.5: Geração de Pontos de Pseudo-Ausência Pelo Método da Distância

3.1.3. Pontos de Pseudo-Ausência Gerados pelo Método BIOCLIM

Assim como nos métodos aleatório e distância, o método apresentado neste tópico, em um primeiro momento, gera pontos de forma aleatória. Entretanto, ele difere dos outros métodos supracitados por utilizar o algoritmo BIOCLIM, que implementa o conceito de envelopes bioclimáticos (Nix, 1986).

Para o algoritmo BIOCLIM, um ponto em um determinado local no espaço pode ser classificado como: inadequado, marginal ou adequado. Para a construção deste método, tomou-se o cuidado, portanto, de obedecer estas três condições propostas pelo BIOCLIM.

Após gerar o ponto de forma aleatória, é verificado se o ponto gerado encontra-se dentro de uma região permitida, ou seja, se não foi criado numa região onde não há dados disponíveis, assim como nos métodos anteriores; caso este ponto esteja em uma região válida, verifica-se se está contido no envelope bioclimático estipulado pelo BIOCLIM; se o ponto estiver dentro dos envelopes, para todos os *layers*, então este é descartado pois trata-se de uma região correspondente à presença.

Caso esteja fora dos envelopes bioclimáticos, mas ainda dentro dos valores máximos e mínimos para todos os *layers*, então este ponto também é descartado, por haver uma probabilidade marginal de presença.

Em último caso, se o ponto estiver fora do intervalo de mínimo e máximo para qualquer uma das variáveis ambientais, o ponto é considerado como inadequado e, sendo assim, válido como um ponto de pseudo-ausência.

A figura 3.6 representa esta idéia.

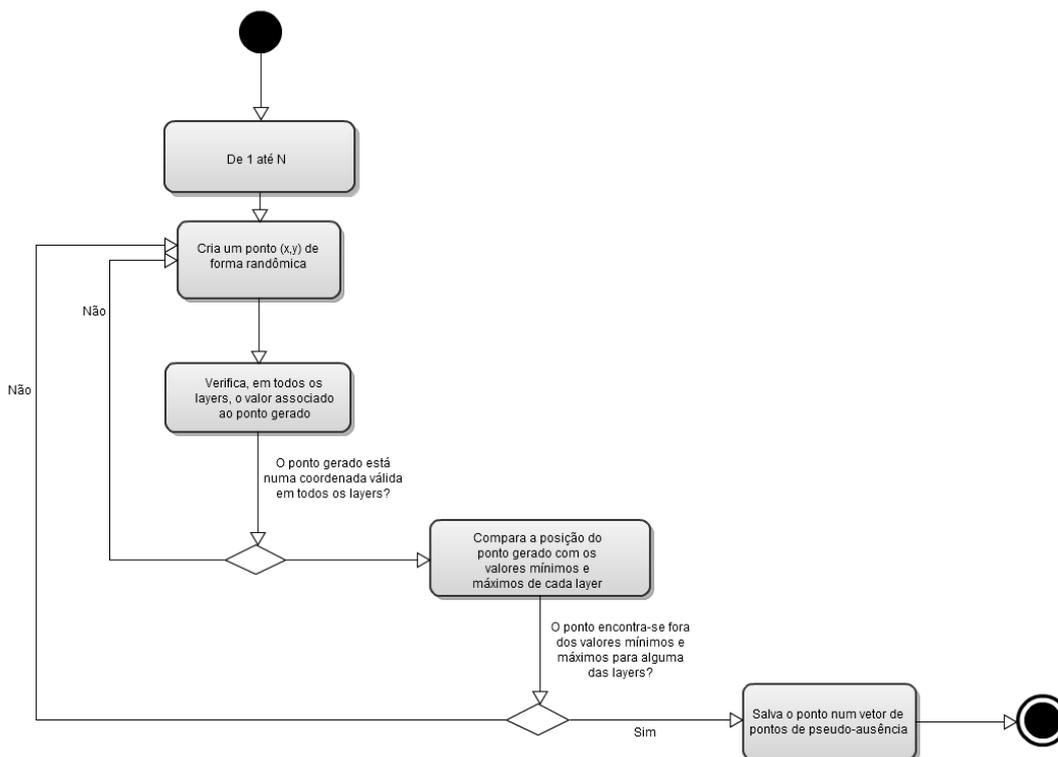


Figura 3.6: Geração de Pontos de Pseudo-Ausência Pelo Método de envelopes bioclimáticos

3.1.4. Delimitação de Área para a Geração dos Pontos de Pseudo-Ausência

Nos SDMs que utilizam dados de presença e ausência, alguns autores afirmam ser necessário que os pontos de ausência (quando não conhecidos) sejam escolhidos de forma satisfatória para um melhor resultado de modelo (VanDerWal, *et. al.*, 2009; Chefaoui & Lobo, 2008).

Para contribuir com esta condição, de que os pontos de pseudo-ausência sejam escolhidos de forma satisfatória, criou-se um delimitador de área para a geração destes pontos. Esta área é definida pelo usuário e pode ser pré-visualizada, possibilitando-lhe verificar se a área definida é a desejada. Isto pode ser observado na figura 3.7.



Figura 3.7: Exemplo de delimitação de área para a geração de pontos de pseudo-ausência

Capítulo 4

Estudo de Caso

Aplicou-se a geração automática de pontos de pseudo-ausência no sistema SAHGA SDM na modelagem da distribuição potencial da espécie *Thalurania furcata boliviana*². Ajustaram-se dois modelos para cada método de geração automática (aleatório, distância e envelope bioclimático). O primeiro utiliza 50 pontos de pseudo-ausência enquanto o segundo utiliza 65 pontos de pseudo-ausência. O objetivo do estudo de caso é verificar a influência dos métodos de geração automática de pontos de pseudo-ausência na geração do modelo de distribuição potencial de espécies. O método base para a comparação foi o método aleatório sem área delimitada para a geração dos pontos

Alguns conjuntos com parâmetros genéticos pré-definidos estão disponíveis no SAHGA SDM (Santa Catarina, 2009), a tabela 4.1 apresenta estes conjuntos, assim como os parâmetros genéticos pré-definidos dos mesmos.

Tabela 4.1: Conjunto de Parâmetros pré-definidos do algoritmo SAHGA

Parâmetros	Conjunto de Parâmetros				
	<i>Default</i>	<i>Fast</i>	<i>Hard</i>	<i>Ultra</i>	<i>High Pop</i>
Tamanho da População	50	20	100	200	500
Número de ciclos	10	5	20	50	20
Temp. mínima	0,001	0,001	0,001	0,001	0,001
Temp. máxima	3	3	3	3	3
Constante de resfriamento	0,9	0,9	0,9	0,9	0,75
Número de repetições	5	3	5	10	5
Tamanho Elite	1	1	1	1	1
Taxa cruzamento	80%	80%	80%	80%	80%
Taxa mutação	1%	1%	1%	1%	1%

Fonte: Santa Catarina, 2009

O conjunto de parâmetros pré-definidos utilizados no ajuste de modelos foi o *Hard*, por ser um conjunto de parâmetros genéticos que visa assegurar ao algoritmo qualidade da resposta

² Mais conhecida como beija-flor-tesoura-verde, catalogada por Adophe Boucard em 1894, esta espécie habita a América do Sul (mais especificamente os Andes peruanos).

com tempo de convergência aceitável (Santa Catarina, 2009).

No estudo de caso aplicou-se a construção da GPM que, como dito anteriormente, é responsável por representar os relacionamentos espaciais (Santa Catarina, 2009). Para a criação da GPM utilizou-se a seguinte regra: se dois pontos, de presença ou ausência, estiverem distantes entre si em até 50 km, então estes estão espacialmente relacionados com peso $W_{ij} = 1$; caso os pontos estejam distantes em um intervalo de 50 a 100 km, então é atribuído um peso $W_{ij} = 0,5$; se os pontos estiverem distantes em um intervalo maior que 100 km, não há relacionamento espacial.

Para cada experimento, dividiu-se aleatoriamente o conjunto com os pontos de presença e pseudo-ausência em dois subconjuntos: treino e teste. Cada um destes subconjuntos restou com 50% do total de pontos, sendo 50% de presença e 50% de ausência, do total de pontos.

Optou-se por utilizar o tipo linear de modelagem, por ter se mostrado mais suficiente para relacionar a ocorrência da espécie com as variáveis ambientais, para a espécie escolhida. O método quadrático não foi escolhido por trazer poucos ganhos, comparado ao modelo linear, e inserir complexidade ao processo.

Como função objetivo, o tipo escolhido foi aquele que maximiza o acerto global do modelo, diminuindo os erros de omissão e comissão, e ao mesmo tempo maximiza a qualidade do ajuste do modelo, minimizando a soma dos desvios quadrados. O sistema SAHGA SDM implementa o tipo Min-Both, que implementa a maximização do ajuste do modelo e também o acerto global do modelo sendo, portanto, o escolhido para a realização dos experimentos.

No sistema SAHGA SDM não há sentido em normalizar-se a variável dependente, pois ela só faz sentido quando o algoritmo SAHGA está processando outros tipos de variáveis dependentes, ou seja, aquelas cuja faixa de variação não está no intervalo entre 0 e 1.

4.1. Espécie *Thalurania furcata boliviana* Boucard, 1984

A base de dados *Thalurania furcata boliviana* é disponibilizada com o instalador do *software* openModeller Desktop. A base contém 65 pontos de presença da referida espécie. Também são disponibilizados 08 *layers* geográficos correspondentes às variáveis ambientais da espécie, a saber: precipitação acumulada no trimestre mais úmido, precipitação acumulada no trimestre mais quente, precipitação anual, temperatura média anual, temperatura média no trimestre mais frio, temperatura média no trimestre mais seco, temperatura média no trimestre

mais quente e temperatura média no trimestre mais úmido.

4.2. Resultados

Como explicado anteriormente, para cada método de geração automática de pontos de pseudo-ausência, foram criados dois conjuntos de dados: um com 50 pontos e outro com 65 pontos de pseudo-ausência. Esta estratégia visa verificar se os modelos ajustados pelo sistema SAHGA SDM apresentam qualidades diferentes quando se utilizam porções equitativas e porções não equitativas de pontos de presença/ausência.

Após a escolha dos pontos, ajustaram-se e compararam-se os modelos, visando verificar quais apresentam os melhores resultados.

Para cada método utilizado foi atribuída uma sigla; para o método aleatório que não utilizou área de delimitação para geração de pontos, atribuíram-se as siglas TR50 (50 pontos de pseudo-ausência) e TR65 (65 pontos de pseudo-ausência); ao método aleatório que utiliza área delimitadora para a geração dos pontos atribuíram-se as siglas RD50 e RD65; ao método da distância atribuíram-se as siglas DS50 e DS65; e ao método dos envelopes bioclimáticos, atribuíram-se as siglas BC50 e BC65.

As métricas para a avaliação dos modelos selecionados para o experimento encontram-se na tabela 4.2. As curvas ROC são apresentadas na figura 4.1

Tabela 4.2: Métricas para avaliação dos modelos selecionados para o experimento.

Métrica	TR50	TR65	RD50	RD65	DS50	DS65	BC50	BC65
Acurácia	91,23%	81,25%	87,72%	85,94%	82,46%	93,75%	92,98%	84,37%
Erro de Omissão	0,00%	0,00%	6,06%	9,09%	0,00%	0,00%	0,00%	0,00%
Erro de Comissão	20,83%	38,71%	20,83%	19,35%	41,67%	12,90%	16,67%	32,26%
CCM	0,829	0,670	0,748	0,721	0,669	0,881	0,862	0,721
AUC	0,851	0,877	0,871	0,864	0,891	0,990	0,968	0,926

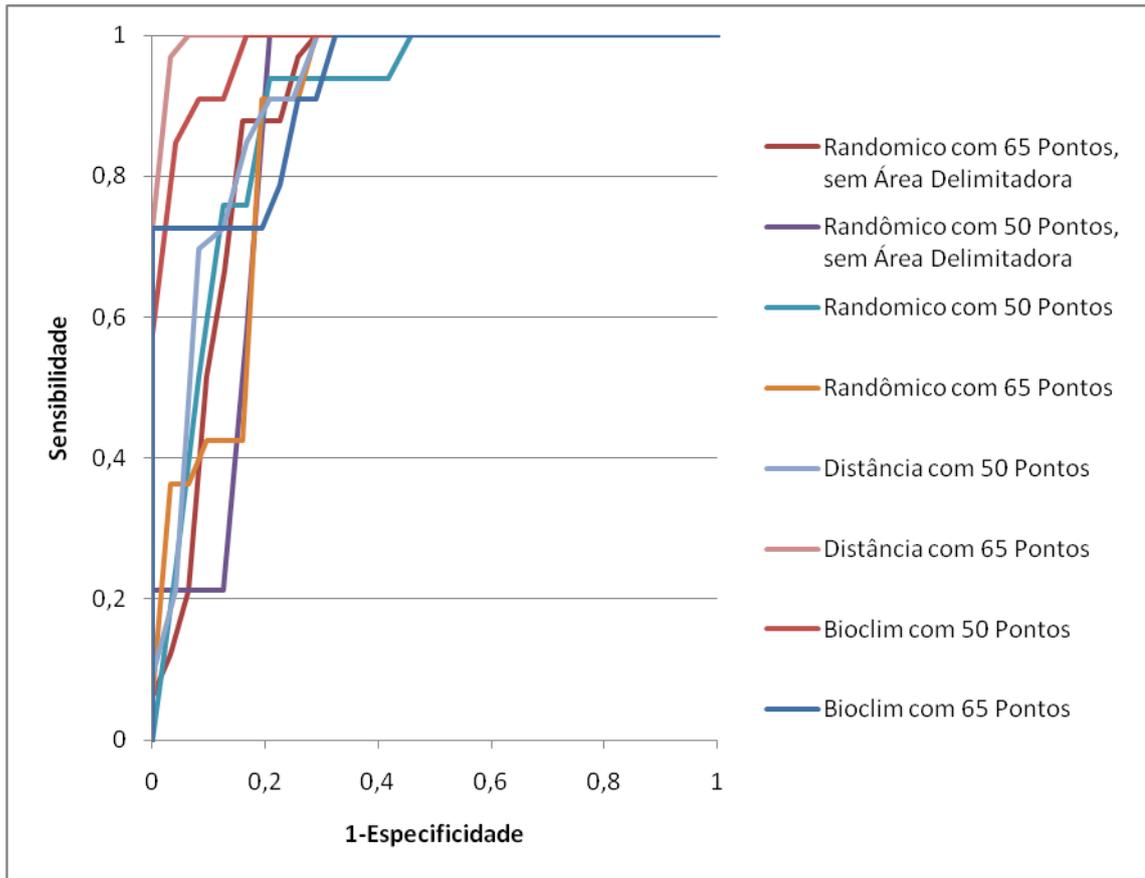


Figura 4.1: Curvas ROC para os modelos selecionados para o experimento

Ao observar as métricas de avaliação para todos os modelos (tabela 4.2), percebe-se que os modelos têm qualidades distintas, entretanto podem ser descritos, de modo geral como bons modelos, por apresentarem AUC superiores a 0,85. As curvas ROC para cada um dos experimentos encontram-se no anexo A deste trabalho.

Pode-se perceber através dos resultados mostrados na tabela 4.2, que modelos que utilizaram um número desigual de pontos (mais presenças do que ausências) tendem a apresentar melhor desempenho quando comparados aos métodos que utilizaram porções iguais de pontos de presença e pseudo-ausência (TR50, RD50 e BC50). A exceção foi o método da distância que utiliza 65 pontos de pseudo-ausência (DS65) que, comparado ao método que utiliza 50 pontos (DS50), apresentou melhores resultados.

O método que apresentou melhor resultado dentre todos os avaliados, foi o método da distância (DS65) com $AUC = 0,990$, indicando elevada capacidade discriminatória. Entretanto, ao analisar a tabela 4.3, pode-se perceber que o método que possui melhor média para o AUC, utilizando as previsões de 50 e 65 pontos, é o método de geração de pontos

utilizando envelopes bioclimáticos. Os modelos ajustados com pontos de pseudo-ausência gerados por este método apresentam menor índice de erro de omissão e comissão e, conseqüentemente, maior coeficiente de correlação de Matthews. Além disso, apresentaram a maior média para a área sob a curva ROC (AUC), indicando que os modelos ajustados possuem elevada capacidade discriminatória.

Tabela 4.3: Métricas para avaliação das médias dos modelos selecionados para o experimento.

Métrica	TR	RD	DS	BC
Acurácia	86,24%	86,83%	88,10%	88,68%
Erro de Omissão	0,00%	7,58%	0,00%	0,00%
Erro de Comissão	29,77%	20,09%	27,29%	19,97%
CCM	0,750	0,735	0,775	0,792
AUC	0,864	0,868	0,940	0,947

Comparado ao método de geração aleatória de pontos de pseudo-ausência, todos os outros métodos apresentaram melhores resultados sendo, portanto, possível afirmar que, para o conjunto de dados *Thalurania furcata boliviana*, a geração de pontos de pseudo-ausência pelo método aleatório, da distância e envelopes bioclimáticos, quando há área delimitadora para a geração dos pontos, apresentam resultados satisfatórios.

Os mapas de distribuição potencial para a espécie *Thalurania furcata boliviana*, previsto por todos os modelos gerados, encontram-se no Anexo B deste trabalho. As taxas de células preditas como presença ou ausência nos mapas que constam no Anexo B são apresentadas na tabela 4.4 para o método aleatório sem área delimitadora (TR50 e TR65), 4.5 para o método aleatório com área delimitadora (RD50 e RD65), 4.6 para o método da distância com área delimitadora (DS50 e DS65) e 4.7 para o método dos envelopes bioclimáticos com área delimitadora (BC50 e BC65).

4.4: Taxas de presença e ausência segundo os modelos TR50 e TR65

Classes	Modelo TR50		Modelo TR65	
	Número Células	%	Número Células	%
Presença	21153	32,7%	14472	22,37%
Ausência	43528	67,3%	50209	77,63%
Total	64681	100%	64681	100%

4.5: Taxas de presença e ausência segundo os modelos RD50 e RD65

Classes	Modelo RD50		Modelo RD65	
	Número Células	%	Número Células	%
Presença	15767	24,38%	16416	25,38%
Ausência	48914	75,62%	48265	74,62%
Total	64681	100%	64681	100%

Tabela 4.6: Taxas de presença e ausência segundo os modelos DS50 e DS65

Classes	Modelo DS50		Modelo DS65	
	Número Células	%	Número Células	%
Presença	16198	25,04%	16848	26,05%
Ausência	48483	74,96%	47833	73,95%
Total	64681	100%	64681	100%

Tabela 4.7: Taxas de presença e ausência segundo os modelos BC50 e BC65

Classes	Modelo BC50		Modelo BC65	
	Número Células	%	Número Células	%
Presença	24846	38,41%	24234	37,47%
Ausência	38835	61,59%	40447	62,53%
Total	64681	100%	64681	100%

Pode-se perceber que todos os métodos, exceto os métodos BC50 e BC65, geraram uma média percentual para pontos preditos de presença e ausência parecida, isto é, predisseram aproximadamente 25% para pontos de ausência e 75% para pontos de presença. Este resultado deve-se à natureza do algoritmo BIOCLIM, que naturalmente isola a área predita como presença da área predita como ausência.

Mais uma vez é possível ratificar a eficiência do método baseado em envelopes bioclimáticos porque, mesmo gerando uma área de presença maior e, por consequência aumentando a possibilidade de erros de comissão, foi o método que menos errou neste quesito.

Capítulo 5

Conclusões

O algoritmo SAHGA – Spatially Aware Hybrid Genetic Algorithm – algoritmo heurístico híbrido com representação explícita de relacionamentos espaciais foi desenvolvido para acomodar a representação de associações e relacionamentos espaciais. Como estratégia, o algoritmo SAHGA utiliza o conceito de Matriz de Proximidade Generalizada.

O sistema SAHGA SDM, como modelador de distribuição potencial de espécies, necessita de pontos de ausência para a execução do modelo, entretanto, este não implementava qualquer estratégia para a geração destes pontos. Deste modo, desenvolveram-se três estratégias para a geração de pontos de pseudo-ausência, a saber: aleatório, distância e envelopes bioclimáticos. Além disso, implementou-se, também, a estratégia de delimitar a área onde seriam gerados os pontos de pseudo-ausência.

Os modelos de distribuição da espécie *Thalurania furcata boliviana*, ajustados pelo sistema SAHGA SDM e utilizando os critérios de geração automática de pontos de pseudo-ausência apresentaram boas métricas de avaliação, sendo que, comparadas ao método de geração aleatória sem área delimitadora para geração dos pontos de pseudo-ausência, apresentaram melhores resultados. Em alguns casos, por exemplo o experimento em que foram gerados 65 pontos de pseudo-ausência para o método da distância, as métricas de avaliação apresentaram resultados excelentes, com resultados de áreas sob a curva ROC (AUC) de 0,990, o que indica uma exímia capacidade discriminatória. Contudo, com uma média de AUC de 0,947, o método baseado em envelopes bioclimáticos se mostrou o método mais adequado para a geração dos pontos.

Entretanto, apesar dos resultados promissores, não se pode estabelecer uma conclusão generalizada. Seriam necessários mais testes com outros conjuntos de dados, visando atestar a eficiência dos métodos de geração de pontos de pseudo-ausência.

Como resultado secundário é pertinente dizer que a interface gráfica e multiplataforma do sistema SAHGA SDM foi concluída, facilitando a utilização do mesmo pelo usuário.

5.1. Trabalhos Futuros

Após a conclusão desta pesquisa, verificou-se que o sistema SAHGA SDM pode ter algumas funcionalidades adicionadas, caracterizando novos desafios científicos:

1. A implementação do método *Best Subset* para a obtenção de um modelo com melhores métricas de qualidade;
2. O desenvolvimento de novas estratégias para a geração automática de pontos de pseudo-ausência, confrontando os resultados obtidos por estas com os resultados apresentados neste trabalho.

Anexo A

Curvas ROC

Nas figuras A.1, A.2, A.3, A.4, A.5, A.6, A.7 e A.8, encontram-se as curvas ROC de cada experimento realizado.

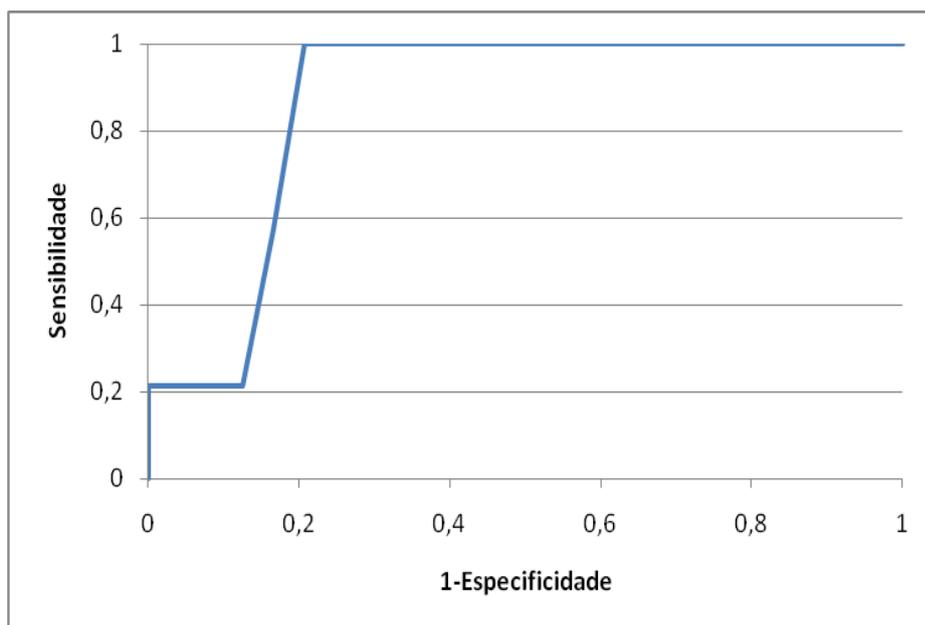


Figura A.1: Curva ROC correspondente ao experimento TR50

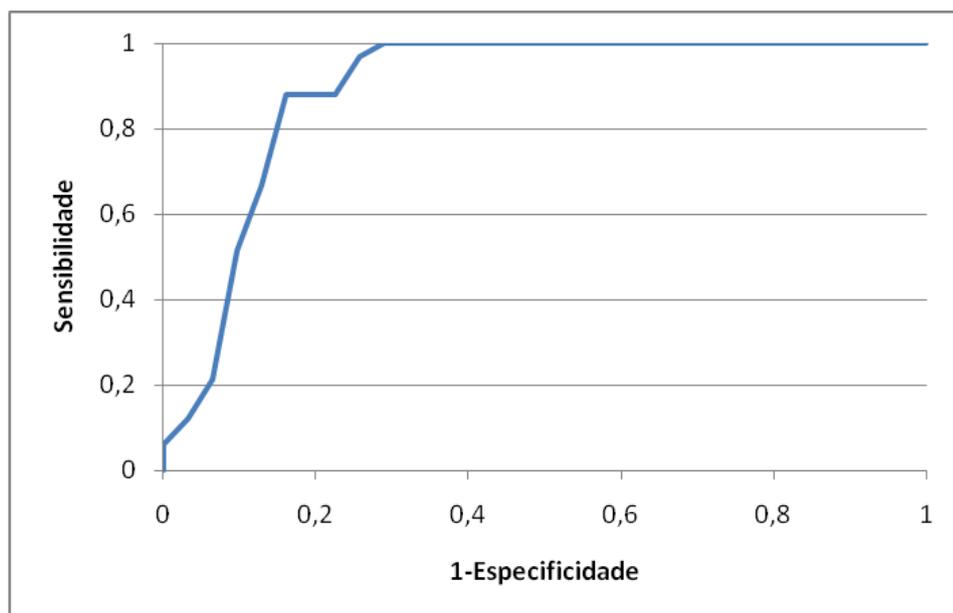


Figura A.2: Curva ROC correspondente ao experimento TR65

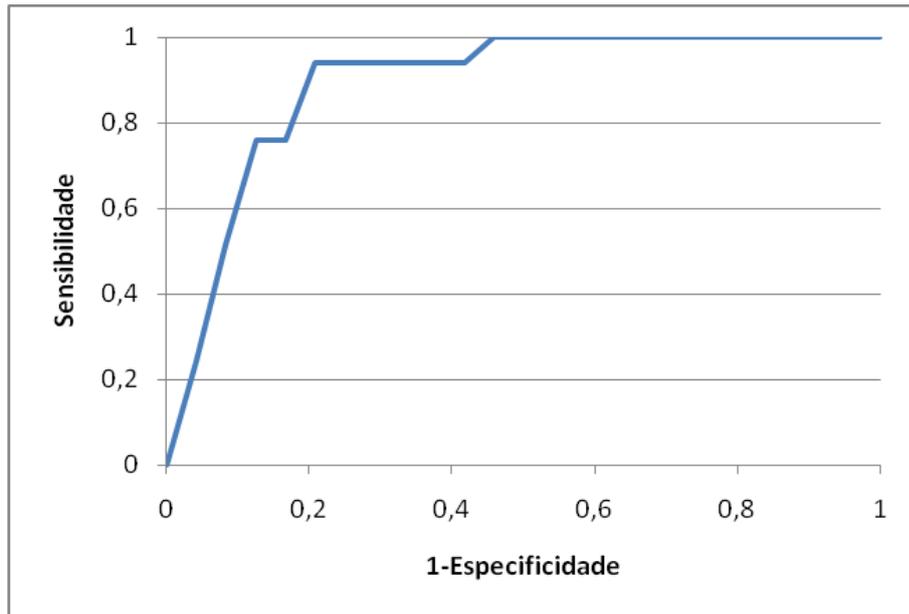


Figura A.3: Curva ROC correspondente ao experimento RD50

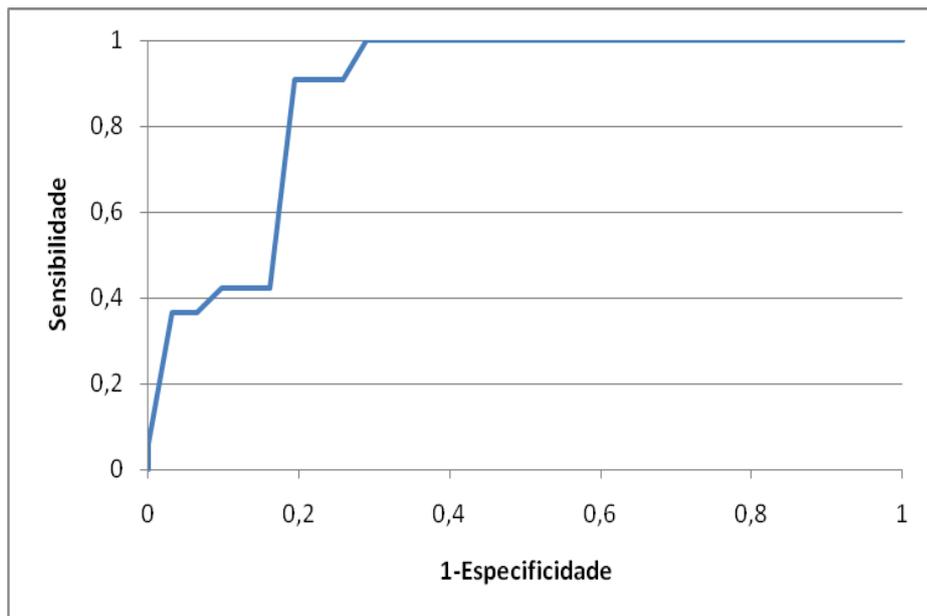


Figura A.4: Curva ROC correspondente ao experimento RD65

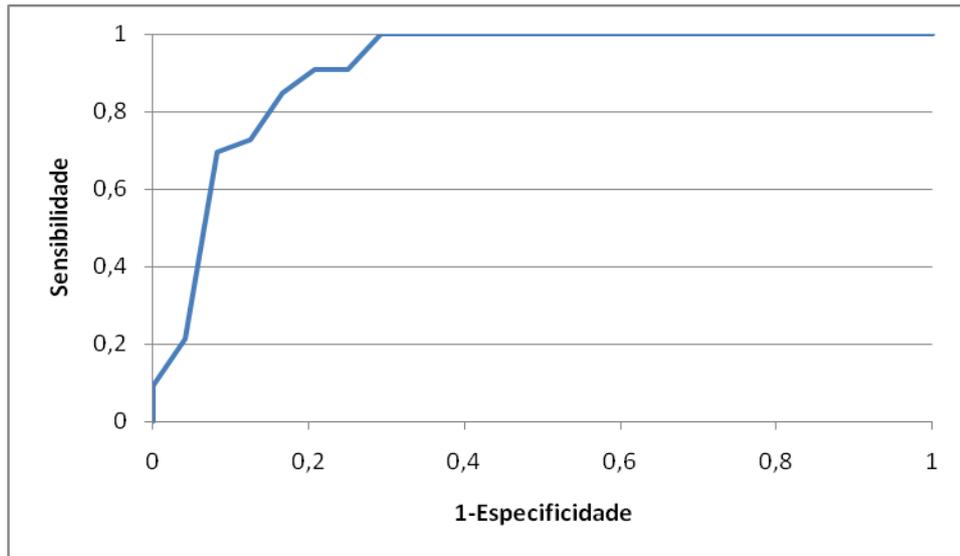


Figura A.5: Curva ROC correspondente ao experimento DS50

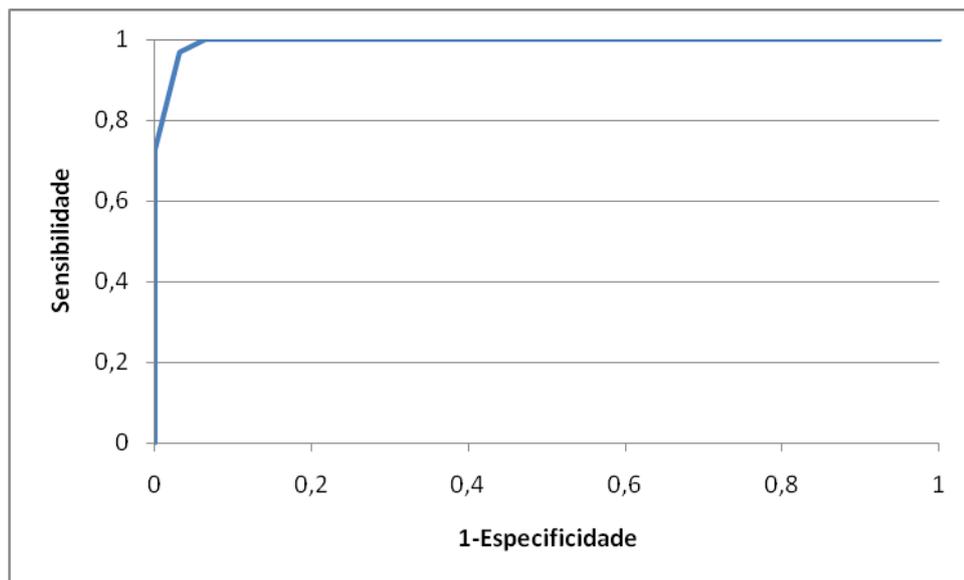


Figura A.6: Curva ROC correspondente ao experimento DS65

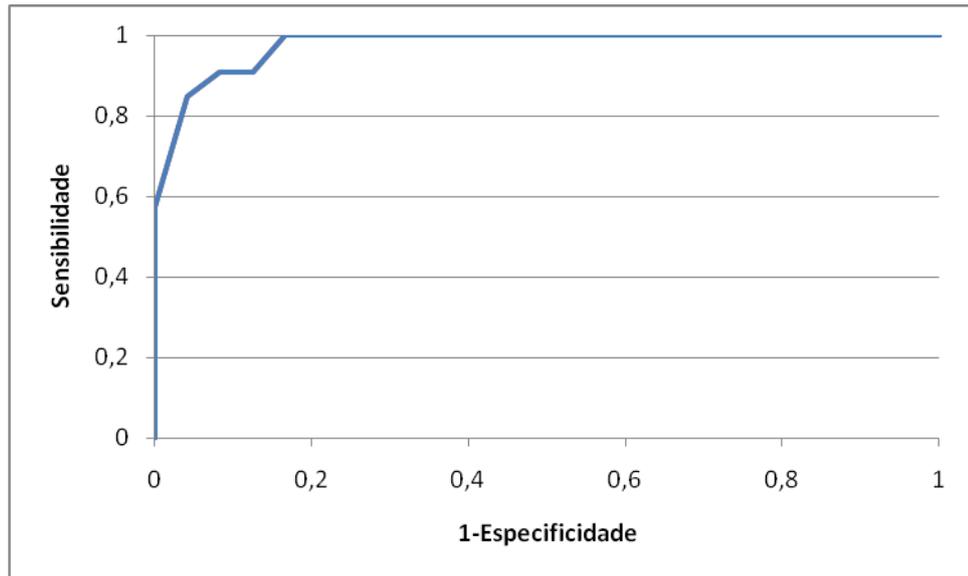


Figura A.7: Curva ROC correspondente ao experimento BC50

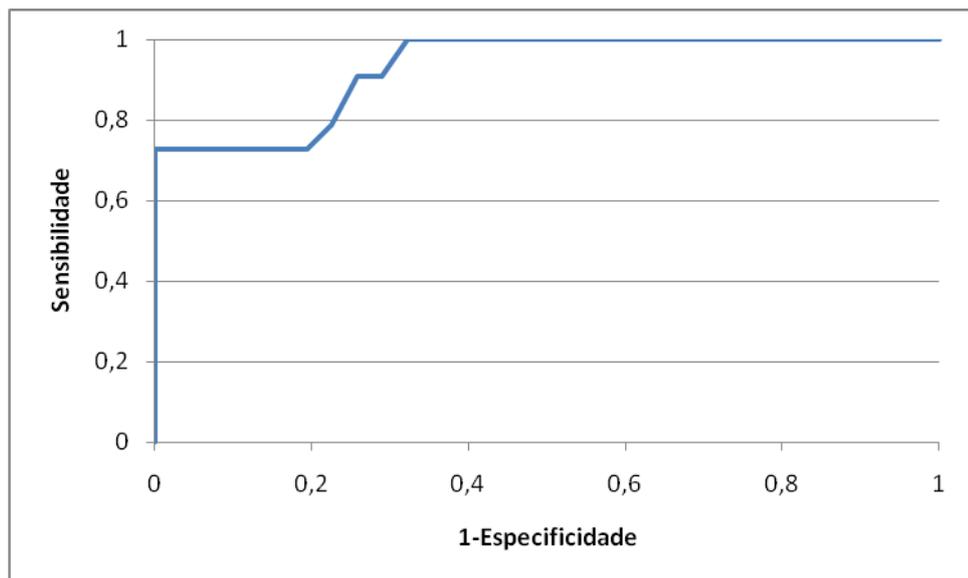


Figura A.8: Curva ROC correspondente ao experimento BC65

Anexo B

Mapas de Distribuição Potencial da Espécie *Thalurania furcata boliviana*

Neste anexo encontram-se os mapas de distribuição potencial da espécie *Thalurania furcata boliviana*, preditos pelo sistema SAHGA SDM. Os círculos verdes são indicados como presença, os círculos amarelos, como pseudo-ausência.

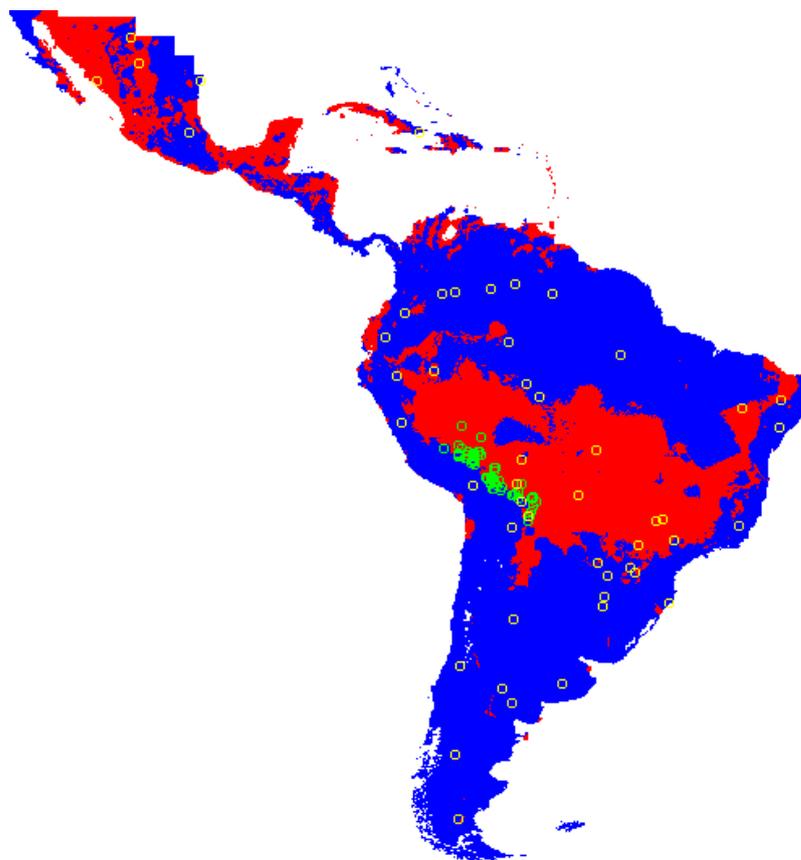


Figura B.1: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo TR50)

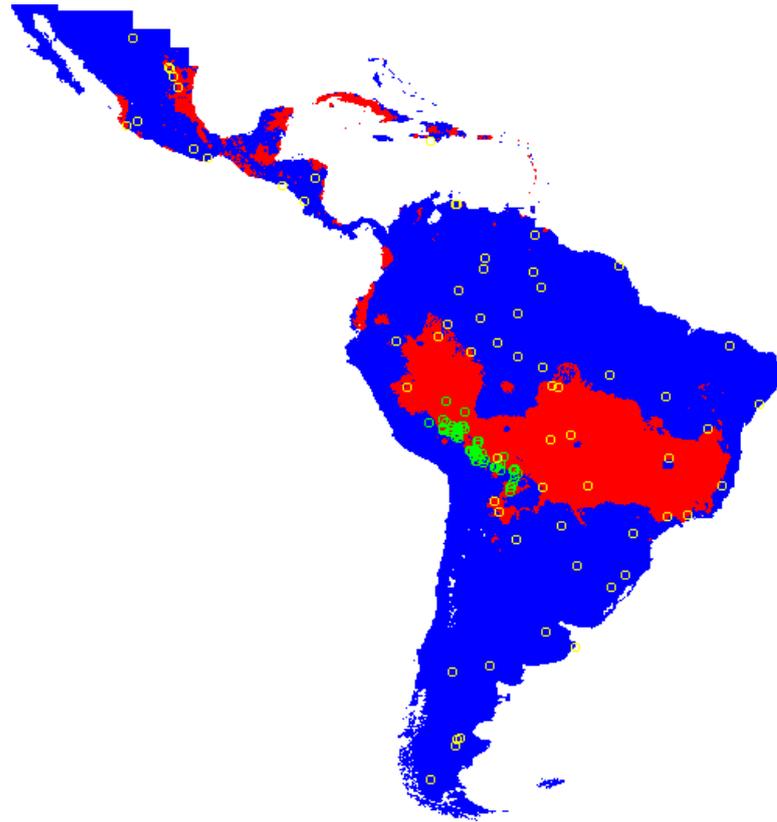


Figura B.2: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo TR65)

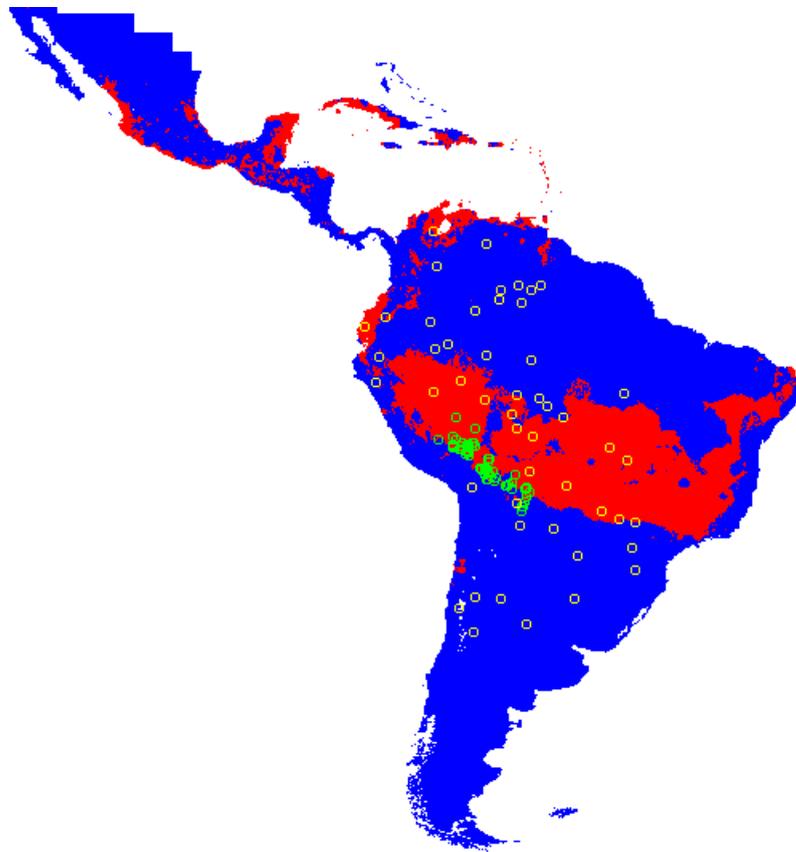


Figura B.3: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo RD50)

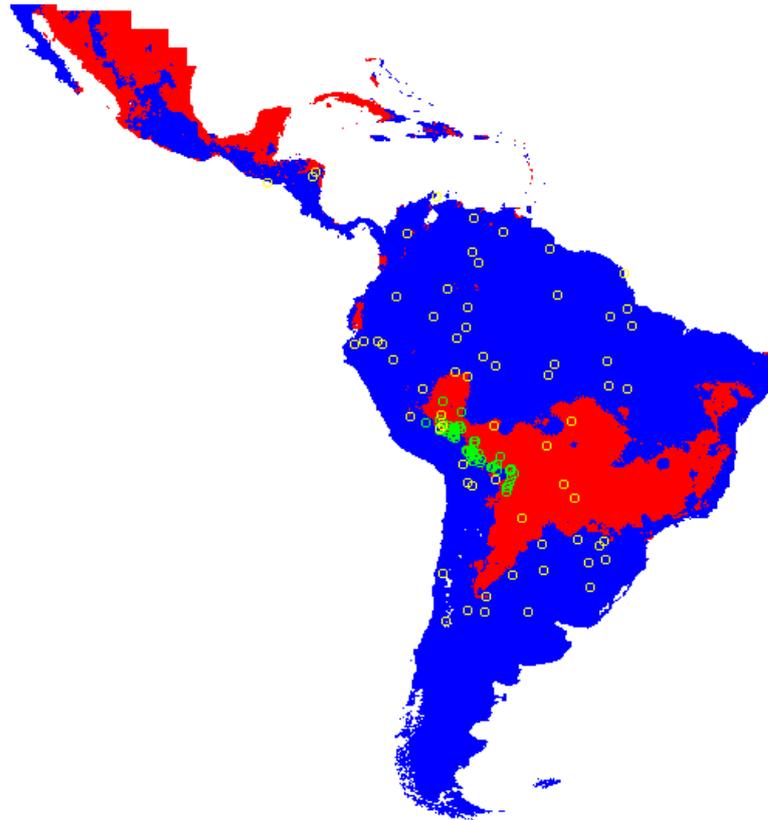


Figura B.4: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo RD65)

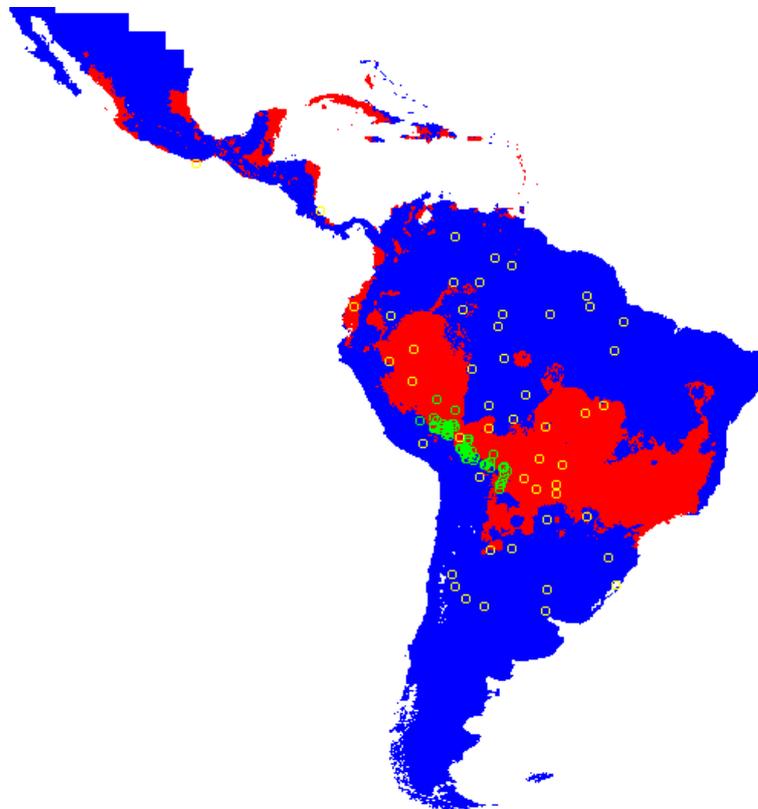


Figura B.5: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo DS50)

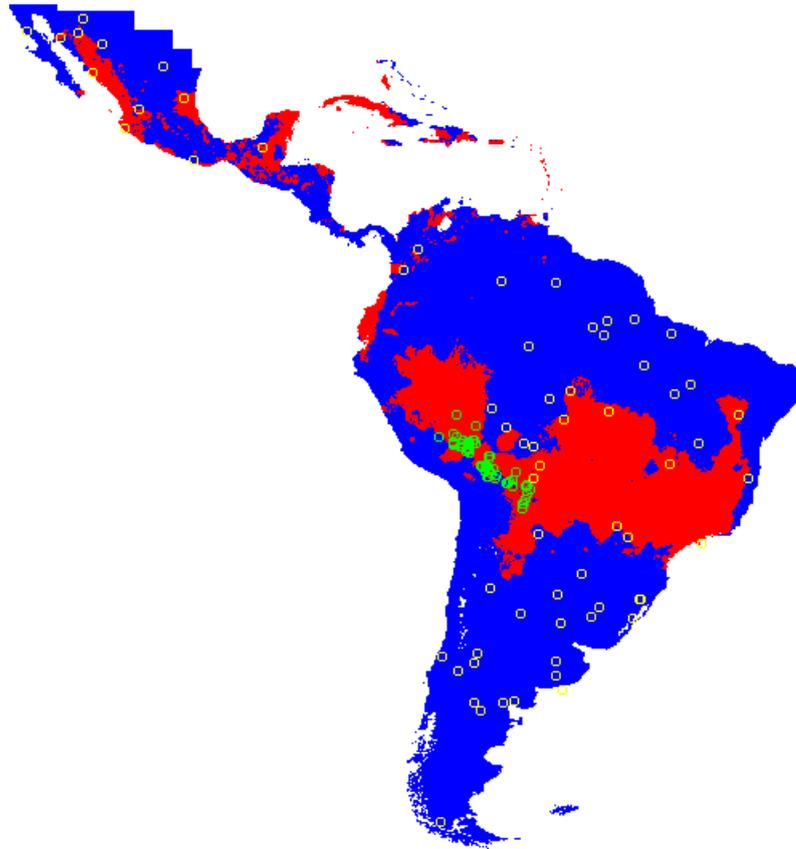


Figura B.6: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo DS65)

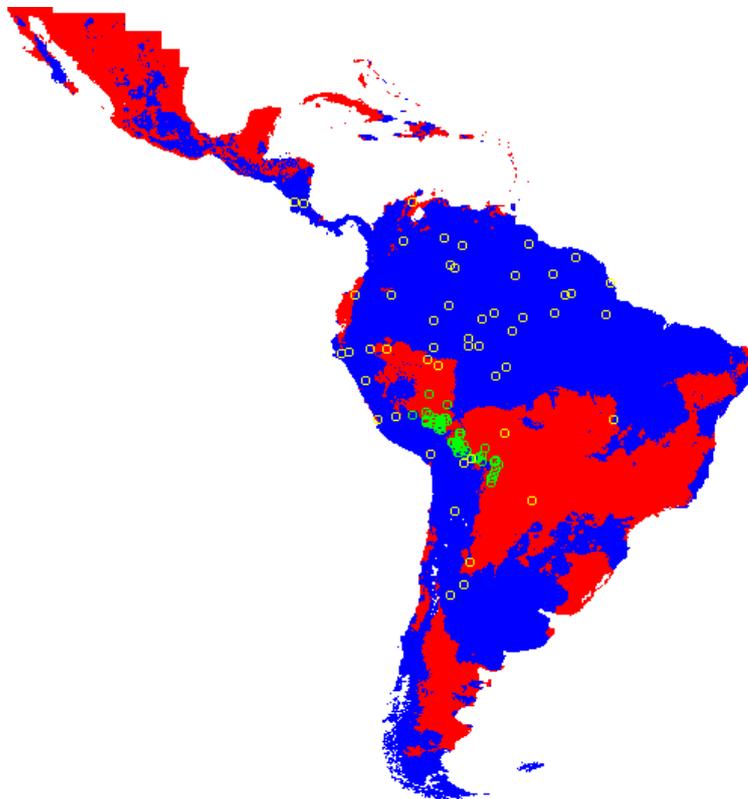


Figura B.7: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo BC50)

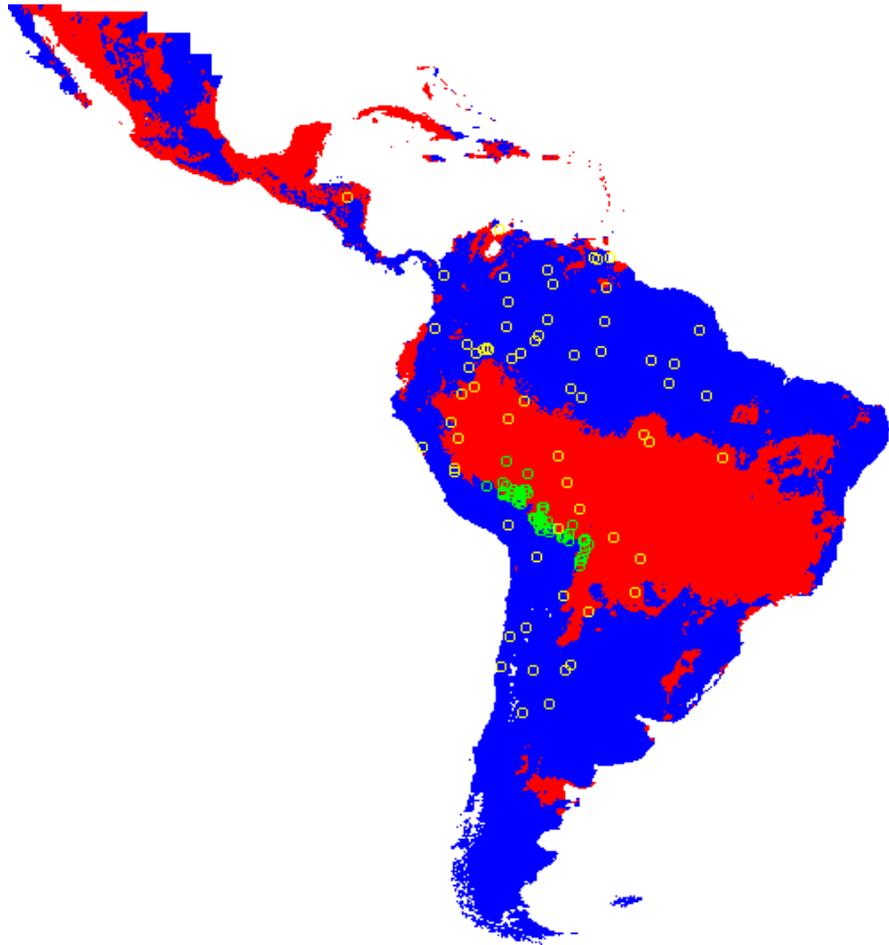


Figura B.8: Distribuição potencial da espécie *Thalurania furcata boliviana* (modelo BC65)

Referências Bibliográficas

AGUIAR, A. P. D. et al. **Modeling spatial relations by generalized proximity matrices.** In: **Brazilian Symposium on Geoinformatics**, 5, 2003. Campos do Jordão - SP. Anais eletrônicos... São José dos Campos: INPE, Nov. 2003. Disponível em: <<http://www.geoinfo.info/geoinfo2003/papers/geoinfo2003-11.pdf>>. Acesso em: 04/07/2006.

ARAÚJO, M. B.; GUISAN, A. **Five (or so) challenges for species distribution modelling.** *Journal of Biogeography*, v. 33, n.10. p. 1677-1688, Oct. 2006.

BALDI, P. et al. **Assessing the accuracy of prediction algorithms for classification: an overview.** *Bioinformatics*, v. 16, n. 5, p. 412-424, Maio 2000.

BRAGA, A. C. S. **Curvas ROC: aspectos funcionais e aplicações.** 2000. 243 p. Tese (Doutorado em Engenharia de Produção e Sistemas) - Universidade do Minho, Braga.

BRENLLA, R. **Historia das Interfaces Gráficas de Usario.** AGNIX_. 2005

BROTONS, L.; THUILLER, W.; ARAUJO, M. B.; HIRZEL, A.H. **Presence-absence versus presence-only modelling methods for predicting bird habitat suitability.** *Ecography*. V. 27, p. 437-448. 2004.

BUSBY, J.R. **BIOCLIM – A bioclimatic analysis and predictive system.** In **Nature Conservation: Cost Effective Biological Surveys and Data Analysis** (Eds. C.R. Margules and M.P. Austin.) pp. 64-68. CSIRO: Canberra. 1991.

CHEFAOUI, R.M.; LOBO, J.M. **Assessing the effects of pseudo-absences on predictive distribution model performance.** *Ecological Modelling*. v. 210. p. 478-486. 2008.

CUNNINGHAM, D. C.; ANDERSON, W.R.; WALSH, K.B. **Ecology and biogeography of *Cassia brewsteri*: assessment of potential sites for cultivation.** *Australian Journal of Agriculture*, Melbourne, v. 42, p. 1071-1080, 2002.

DAUBENMIRRE, R. **Plant Communities: a textbook of plant synecology.** New York. Harper & Row, 1968.

ELITH, J.; GRAHAM, C.H.; NCEAS Modeling Group. **Novel methods improve prediction of species? Distributions from occurrence data.** *Ecography*, v.29, n. 2. p. 129-151, Apr. 2006.

ENGLER, R.; GUIBAN, A.; RECHSTEINER, L. **An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data.** *Journal of Applied Ecology*, v. 41, n. 2, p. 263-274, Abr. 2004.

FIELDING, A. H.; BELL, J. F. **A review of methods for the assessment of prediction errors in conservation presence/absence models.** *Environmental Conservation*, v. 24, n. 01, p. 38-49, Mar. 1997.

GODOWN, M. E.; PETERSON, A. T. **Preliminary distributional analysis of US endangered bird species.** *Biodiversity and Conservation*, London, v. 9, p. 1312-1322, 2000.

GUIBAN, A.; THUILLER, W. **Predicting species distribution: offering more than simple habitat models.** *Ecology Letters*, v. 8, n. 9, p. 993-1009, Jun. 2005.

GUIBAN, A.; ZIMMERMAN, N. E. **Predictive habitat distribution models in ecology.** *Ecological Modelling*, v. 135, n. 2, p. 147-186, Dec. 2000.

HIRZEL, A.H.; HELFER, V.; METRAL, F. **Assessing habitat suitability models with a virtual species.** *Ecological Modelling*, v. 145, p. 111–121. 2001.

HUTCHINSON, G. E. **Concluding Remarks.** *Cold Spring Harbor Symp. Quantitative Biol.*, v. 22, p. 415-427, 1957.

ISO 9241-1, **Ergonomic requirements for office work with visual display terminals (VDTs): General introduction.** 1992.

IWASHITA, F. **Sensibilidade de modelos de distribuição de espécies a erros de posicionamento de dados de coleta.** 2007. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos.

MATEO, R. G.; CROAT, T.B.; FELICÍSIMO, Á.M.; MUÑOZ, J. **Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections.** *Diversity and Distributions*, v.16, p.84-94. 2010.

NIELSEN, J. **Usability Engineering.** Boston: Academic Press Professional, 1993.

NIX, H. A. **A biogeographic analysis of Australian elapid snakes.** In: LONGMORE, R. (Ed.) **Atlas of elapid snakes of Australia.** Canberra: Australian government publishing service, v.7, 1986. p. 4–15. (Australian flora and fauna series).

- OLIVEIRA, A. L. L.; SANTOS, E. T. **Estudo de usabilidade da interface gráfica do programa de geometria dinâmica iGEOM**. 16o Simpósio Nacional de Geometria Descritiva e Desenho Técnico. 2003.
- PEARCE, J.; BOYCE, M. **Modelling distribution and abundance with presence-only data**. Journal of Applied Ecology, v. 43, p. 405–412. 2006.
- PEARCE, J.; FERRIER, S. **An evaluation of alternative algorithms for fitting species distribution models using logistic regression**. Ecological Modelling, 128: 127-147. 2000.
- PEDROSA, B. M. **Ambiente computacional para modelagem dinâmica**. 2003. 71 p. Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos.
- PHILLIPS, S.J.; ANDERSON, R.P.; SCHAPIRE, R.E, **Maximum entropy modeling of species geographic distributions**. Ecological Modelling, Amsterdam, v. 190, p.231-259, 2006.
- PONTIUS Jr., G, **Quantification error versus location error in comparison of categorical maps**. Photogrammetric engineering and remote sensing, v. 66, n. 8 p 1011-1016, Agosto de 2000
- PONTIUS Jr, G; SCHNEIDER, L.C., **Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA**. Agriculture, Ecosystems and Environment. V. 85, n.1, p. 239-248, Junho de 2001.
- RAVEN, P. H; EVERT, R. F; EICHHORN, S. E. **Biologia Vegetal**. 6ª ed. Rio de Janeiro: Guanabara Koogan, 2001.
- SANTA CATARINA, A. **SAHGA – Um algoritmo genético híbrido com representação explícita de relacionamentos espaciais para análise de dados geoespaciais**. São José dos Campos: INPE, 2009.
- SIQUEIRA, M. F. **Uso de modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais**. 2005. 107 p. Tese (Doutorado em Ciências de Engenharia Ambiental) - Universidade de São Paulo, São Carlos.
- SUTHERST R. W.; MAYWALD G. F.; RUSSEL B. L. **Estimating vulnerability under climate change**. Agri Ecosyst Environ. v. 82. p. 303–319. 2000.
- VANDERWAL, J.; SHOO, L. P.; GRAHAM, C.; WILLIAMS, S. E. **Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know?**. Ecological Modelling. v. 220. p. 589-594. 2009.
- VIVO, M. D.; CARMIGNOTTO, A. P. **Holocene vegetation change and the mammal faunas of South America and Africa**. 2004. Journal of Biogeography, v. 31, n. 6 p. 943-957.

WISZ, M. S.; GUIBAN A. **Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data.** BMC Ecology. 2009.

ZANIEWSKI, A.E.; LEHMANN, A.; OVERTON, J. M. **Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns.** Ecological Modelling, v. 157, p. 261–280. 2002.