



Universidade do Estado de Mato Grosso  
*Campus* de Nova Xavantina  
Programa de Pós-Graduação em Ecologia e  
Conservação



**INDICANDO CAMINHOS: PROPOSTAS CONCEITUAIS E  
METODOLÓGICAS NA MODELAGEM DE ADEQUABILIDADE  
AMBIENTAL**

**João Carlos Pires de Oliveira**

Nova Xavantina-MT  
2019



Universidade do Estado de Mato Grosso  
*Campus de Nova Xavantina*  
Programa de Pós-Graduação em Ecologia e Conservação



**INDICANDO CAMINHOS: PROPOSTAS CONCEITUAIS E  
METODOLÓGICAS NA MODELAGEM DE ADEQUABILIDADE  
AMBIENTAL**

**João Carlos Pires de Oliveira**

Dissertação apresentada à Universidade do Estado de Mato Grosso, como parte dos requisitos do Programa de Pós-graduação em Ecologia e Conservação para obtenção do título de mestre.

Orientador: Prof. Dr. Pedro Vasconcellos Eisenlohr  
Coorientador: Dr. Diogo Souza Bezerra Rocha

Nova Xavantina-MT  
2019

O48i OLIVEIRA, João Carlos Pires de.  
Indicando Caminhos: Propostas Conceituais e Metodológicas na Modelagem de Adequabilidade Ambiental / João Carlos Pires de Oliveira – Nova Xavantina, 2019.  
53 f.; 30 cm.(ilustrações) Il. color. (sim)

Trabalho de Conclusão de Curso  
(Dissertação/Mestrado) – Curso de Pós-graduação Stricto Sensu (Mestrado Acadêmico) Ecologia e Conservação, Faculdade de Ciências Agrárias Biológicas e Sociais Aplicadas, Câmpus de Nova Xavantina, Universidade do Estado de Mato Grosso, 2019.  
Orientador: Pedro Vasconcellos Eisenlohr Coorientador: Diogo Souza Bezerra Rocha

1. Modelos Ecológicos. 2. Cuidados em Modelagem Ecológica. 3. Rotinas de Modelagem. 4. Seleção de Gcms. 5. Análise Multivariada. I. João Carlos Pires de Oliveira.  
II. Indicando Caminhos: Propostas Conceituais e Metodológicas na Modelagem de Adequabilidade Ambiental: .  
CDU 574.4

**INDICANDO CAMINHOS: PROPOSTAS CONCEITUAIS E  
METODOLÓGICAS NA MODELAGEM DE ADEQUABILIDADE  
AMBIENTAL**

**JOÃO CARLOS PIRES DE OLIVEIRA**

DISSERTAÇÃO APRESENTADA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E  
CONSERVAÇÃO DA UNIVERSIDADE DO ESTADO DE MATO GROSSO COMO REQUISITO  
PARCIAL À OBTENÇÃO DO TÍTULO DE “MESTRE”.

APROVADA EM 07 DE FEVEREIRO DE 2019, PELA BANCA EXAMINADORA:

---

Dr. Pedro Vasconcellos Eisenlohr  
Universidade do Estado de Mato Grosso  
Orientador

---

Dr. Fabricius Maia Chaves Bicalho Domingos  
Universidade Federal de Mato Grosso  
Membro Titular

---

Dr<sup>a</sup>. Natália Mundim Tôrres  
Universidade Federal de Uberlândia  
Membro Titular

---

Dr<sup>a</sup>. Ivone Vieira da Silva  
Universidade do Estado de Mato Grosso  
Membro Suplente

---

Dr. Thadeu Sobral de Souza  
Universidade Federal de Lavras  
Membro Suplente

*“Ausência de evidência não é  
evidência de ausência.” - Carl Sagan*

## AGRADECIMENTOS

Ao finalizar este trabalho agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de estudos e pelo suporte.

Agradeço também à Universidade do Estado de Mato Grosso (UNEMAT) pela oportunidade de cursar um curso de Pós-graduação de alto nível e pelo suporte fornecido para meu aprimoramento acadêmico.

Todo o meu respeito e gratidão a todos os integrantes do Laboratório de Ecologia da Universidade do Estado de Mato Grosso do *Campus* de Alta Floresta, com destaque para o Prof. Dr. Pedro Vasconcellos Eisenlohr, meu orientador, por todo o cuidado com minha formação e pela “orientação democrática” que proporciona a liberdade para pensar e desenvolver o meu senso crítico. Da mesma forma agradeço ao meu coorientador Dr. Diogo Souza Bezerra Rocha por todo o suporte técnico e pelas ótimas contribuições para a realização deste trabalho.

Um agradecimento especial vai para meus pais, que sempre me apoiaram e me incentivaram a seguir na carreira acadêmica.

Por fim, agradeço a todos aqueles que sempre me estenderam a mão e que infelizmente por serem muitos não poderei individualizá-los.

## SUMÁRIO

RESUMO.....	VIII
ABSTRACT.....	IX
<b>CAPÍTULO 1 - CONCEITOS E MÉTODOS EM MODELAGEM DE ADEQUABILIDADE AMBIENTAL: UMA IMPORTANTE FERRAMENTA PARA CONSERVAÇÃO FLORESTAL.....</b>	<b>10</b>
1. INTRODUÇÃO.....	11
2. CUIDADOS CONCEITUAIS.....	13
2.1 DIAGRAMA BAM.....	16
3. CUIDADOS METODOLÓGICOS.....	18
3.1 Registros de Ocorrência.....	18
3.2 Varáveis Exploratórias.....	20
4. Uma Proposta de Rotina para ESM.....	21
4.1 ESM Sob Clima Presente.....	22
4.2 ESMs Sob Cenários de Clima Futuro.....	23
5. CONSIDERAÇÕES FINAIS.....	25
6. REFERÊNCIAS.....	26
<b>CAPÍTULO 2 - MENOS PODE SER MAIS: UMA ROTINA OBJETIVA PARA SELECIONAR MODELOS DE CIRCULAÇÃO GERAL E SUAS APLICAÇÕES NA MODELAGEM ECOLÓGICA.....</b>	<b>31</b>
RESUMO.....	32
1. INTRODUÇÃO.....	33
2. MATERIAL E MÉTODOS.....	35
2.1 Banco de Dados.....	35
2.2 Análise dos Dados.....	36
3. RESULTADOS.....	40
4. DISCUSSÃO.....	46

5. CONCLUSÕES.....	48
6. REFERÊNCIAS.....	49

## RESUMO

O conjunto de técnicas conhecidas como Modelagem de Distribuição de Espécies ou Modelagem de Nicho Ecológico possui ampla utilização na Conservação da Biodiversidade. Estas técnicas possuem aplicações que vão desde guias para realização de excursões de campo até detecção de possíveis impactos causados por mudanças no clima futuro. Entretanto, alguns cuidados devem ser tomados antes de iniciarmos a modelagem, e no primeiro capítulo desta dissertação abordamos algumas dessas questões que devem ser consideradas antes e durante o procedimento de modelagem. Esses cuidados são relacionados à nomenclatura designada a essas técnicas e também cuidados metodológicos que devemos ter durante o procedimento de modelagem. Cuidados metodológicos devem ser tomados na coleta dos pontos de ocorrência do grupo de interesse, na seleção das variáveis que serão utilizadas como preditoras dos modelos e também com a interpretação dos resultados obtidos. Neste mesmo capítulo sugerimos que Modelagem de Adequabilidade Ambiental (ESM) é a forma mais adequada para nomear esse conjunto de técnicas. Além disso, sugerimos uma rotina para construir ESM sobre clima atual e passado/futuro. Para a modelagem sob clima futuro é comum utilizarmos projeções de variáveis climáticas para outro intervalo temporal. Essas projeções das variáveis climáticas são produzidas com base em Modelos de Circulação Geral (GCMs) e a escolha desses GCMs para a realização de ESM é geralmente feita de forma arbitrária e pouco transparente. Para tornar mais clara a escolha de subgrupos de GCMs sem que haja perda de desempenho, elaboramos o segundo capítulo desta dissertação e apresentamos uma rotina para selecionar subgrupos de GCMs com maior independência e, assim, evitar dados redundantes dentro dos ESMs. Para isso utilizamos uma abordagem em 4 etapas: 1) análise de ordenação (*PCoA*), para identificar como os objetos (GCMs) seriam ordenados no espaço bidimensional e quais descritores mais influenciariam nessa ordenação; 2) análise de agrupamento, para identificar possíveis grupos por similaridade; 3) análise correlativa, para selecionar um GCM menos correlacionado com os demais integrantes de seus respectivos grupos; e 4) análise de validação, por meio de Espécies Virtuais (VS), da capacidade preditiva da ESM em prever a adequabilidade das VSs. As principais conclusões alcançadas no capítulo 1 são: ESMs são realmente poderosas e podem atuar como apoio na Conservação da Biodiversidade, inclusive das Florestas Tropicais; Modelagem de Adequabilidade Ambiental seja uma melhor nomenclatura para esse conjunto de técnicas. Apesar de serem poderosas ferramentas, alguns cuidados devem ser tomados para evitar equívocos antes, durante e depois do processo de modelagem. As principais conclusões alcançadas no capítulo 2 são: os GCMs são altamente correlacionados e por isso não recomendamos que sejam utilizados grandes grupos na construção de ESMs, pois haveria maior representatividade dos GCMs mais correlacionados; mostramos que, se selecionarmos GCMs que sejam mais independentes, os desempenhos dos ESMs são tão representativos quanto ESMs produzidos sobre grandes grupos de GCMs, e com grupos mais independentes evitamos o viés de os GCMs mais correlacionados terem maior peso dentro da análise.

**Palavras Chave:** Modelos Ecológicos, Cuidados em Modelagem Ecológica, Rotinas de Modelagem, Seleção de GCMs, Análise Multivariada.

## ABSTRACT

The set of techniques known as Species Distribution Modeling or Ecological Niche Modeling has wide use in Biodiversity Conservation. These techniques have applications ranging from guides to field trips to detecting possible impacts caused by changes in the future climate. However, some care must be taken before we begin modeling, and in the first chapter of this master's dissertation we address some of these issues that must be considered before and during the modeling procedure. Such care is related to nomenclature assigning to these techniques and also methodological care that we should have during the modeling procedure. Methodological care must be taken in the collection of the points of occurrence of the interest group, in the selection of variables that will be used as predictors of the models and also with the interpretation of the results obtained. In this same chapter we suggest that Environmental Suitability Modeling (ESM) is the most appropriate way to name this set of techniques. Moreover, we suggest a routine to build ESM on current and past/future climate. In modeling over future climate it is common to use projections of climatic variables for another time interval. These projections of climate variables are produced based on General Circulation Models (GCMs) and the choice of these GCMs to build ESM is usually done in an arbitrary and non-transparent way. To make it easier to select subgroups of GCMs without loss of performance, we elaborate the second chapter of this master's dissertation and present a routine to select subgroups of GCMs with a greater independence to avoid redundant data within the ESMs. For this we used a 4-step approach: 1) ordination analysis (PCoA), to identify how the objects (GCMs) would be ordered at a two-dimensional space and which descriptors would most influence this ordering; 2) cluster analysis, to identify possible groups by similarity; 3) correlative analysis, to select the GCM less correlated with the other members of their respective groups; and 4) validation analysis, through Virtual Species (VS), of the predictive capacity of ESM to predict the adequacy of VS. The main conclusions reached in chapter 1 are: ESMs are really powerful and can act as a support in the Conservation of Biodiversity, including Tropical Forests; Environmental Suitability Modelling is a better nomenclature for this set of techniques; although being powerful tools, some care must be taken to avoid misunderstanding before, during and after the modeling process. The main conclusions reached in chapter 2 are: the GCMs are highly correlated and therefore we do not recommend that large groups be used in the construction of ESMs, since there would be greater representativeness of the most correlated GCMs; the results show that if we select GCMs that are more independent, the performances of the ESMs are as representative as ESMs produced over large groups of GCMs, and with more independent groups we avoid the bias of the more correlated GCMs having greater weight in the analysis.

**Keywords:** Ecological Models, Ecological Modeling Care, Modeling Routines, GCMs Selection, Multivariate Analysis

# **CAPÍTULO 1 - CONCEITOS E MÉTODOS EM MODELAGEM DE ADEQUABILIDADE AMBIENTAL: UMA IMPORTANTE FERRAMENTA PARA CONSERVAÇÃO FLORESTAL**

*João Carlos Pires-Oliveira<sup>1</sup>, Leandro José-Silva<sup>1</sup>,  
Diogo Souza Bezerra da Rocha<sup>2</sup>  
e Pedro V. Eisenlohr<sup>1</sup>*

<sup>1</sup>Laboratório de Ecologia, Universidade do Estado de Mato Grosso,  
Alta Floresta, MT, Brasil

<sup>2</sup>Escola Nacional de Botânica Tropical, Jardim Botânico do Rio de Janeiro, Rio de Janeiro, RJ,  
Brasil

## 1. INTRODUÇÃO

A diversidade biológica em nosso planeta é tão alta que não temos ainda um número preciso para representá-la (Wilson, 2000; Wilson, 2003), mesmo com os avanços em tecnologias para fazer estimativas neste campo (Guralnick et al., 2007; Mora et al. 2011; Larsen et al. 2017). Existem organismos que vivem nos lugares mais improváveis, como as calotas polares, os desertos extremamente áridos (ou seja, o deserto do Atacama) e os oceanos mais profundos (Wynn-Wilhams, 1996; Willerslev et al., 1999; Warren-Rhodes et al., 2006), mas é na região tropical que a maior parte dessa biodiversidade está concentrada (Jenkins, 2003). De fato, a biodiversidade é apresentada em toda a sua grandeza em florestas tropicais (por exemplo, Slik et al., 2015), como a floresta amazônica no continente americano, a floresta tropical do Congo e a floresta tropical do Sudeste Asiático.

Grandes esforços têm sido investidos para preencher lacunas de conhecimento sobre a distribuição geográfica de espécies em florestas tropicais (Kerr, 1997; Myers et al., 2000; Siqueira et al., 2009; Oliveira-filho, 2017). No contexto da conservação da biodiversidade tropical, o conhecimento sobre os padrões ecológicos e de distribuição das espécies é de particular importância, uma vez que o risco de negligenciar uma espécie-chave devido à falta de dados pode comprometer significativamente a tomada de decisão. Neste contexto, um grande número de ferramentas foram desenvolvidas, como as técnicas de modelagem de adequabilidade ambiental (ESM), para prever ou projetar a distribuição potencial de diferentes espécies (Elith et al., 2006; Liu et al., 2011; Zhang et al., 2015). Essas ferramentas utilizam variáveis preditivas (geralmente dados abióticos) que deveriam ser relevantes para determinar a ocorrência de espécies, que são combinadas com dados georreferenciados para gerar um modelo de adequabilidade ambiental. É possível projetar esse modelo de volta ao espaço geográfico, indicando áreas de adequabilidade para a ocorrência da espécie com base no leque de condições ambientais em que ocorre (Elith et al., 2006; Elith & Leathwick, 2009; Peterson et al., 2011). Em poucas palavras, o ESM "transforma" a distribuição geográfica de pontos de ocorrência em uma superfície de distribuição ou área de adequabilidade, com base em variáveis ambientais. Mesmo que alguns modelos particulares apresentem um grau considerável de incerteza, eles ainda representam uma alternativa ou um complemento para os estudos *in situ*, a fim de descrever as faixas

geográficas das espécies (Smith et al., 2006; Liu et al., 2011; Gastón et al., 2014; Zhang et al., 2015).

A distribuição geográfica de uma determinada espécie está relacionada aos requisitos necessários para estabelecer populações persistentes, mas também a fatores relacionados à capacidade e possibilidade de dispersão, bem como relações interespecíficas (isto é, competição, predação, parasitismo) e fatores históricos (Brown et al., 1996; Chave et al., 2002), ou seja, está relacionado ao nicho dos grupos de organismos. Os ESMs procuram representar esse universo de condições e recursos por meio de um subconjunto de camadas ambientais, mas não revelam nada sobre o movimento dos indivíduos, uma vez que os dados de dispersão são raros ou inexistentes para a maioria das espécies. Além disso, fatores históricos também não são considerados na ESM. A ESM encontra apoio na teoria do nicho de Hutchinson (1957), que é o seu pilar central. Essa teoria é baseada nos fatores ecológicos atuais em que as espécies sobrevivem, crescem e se reproduzem, ou em outras palavras, nas condições que a espécie tolera e nos recursos que suprem suas necessidades. Entretanto, nem todos os requisitos do nicho de Hutchinson (1957) são atendidos. Com base nesse conhecimento prévio, são geradas projeções espaciais para identificar áreas adequadas para a ocorrência das espécies (Soberon, 2007; Pearman et al., 2008). Assim, dois pressupostos fundamentais devem ser considerados quando se utilizam ESMs: 1) a distribuição do grupo ecológico considerado no estudo está em equilíbrio com o meio ambiente (De Marco-Junior et al., 2008), e 2) o nicho desses grupos é conservado no espaço e no tempo, uma vez que as condições ambientais são consideradas o principal fator limitante das distribuições das espécies em geral (Pearson & Dawson, 2003, Soberón, 2007, Soberón & Nakamura, 2009).

O ESM tem sido extensivamente utilizado com múltiplas aplicações, incluindo o campo da ciência da conservação, em estudos que visam compreender melhor os padrões de distribuição de espécies na superfície terrestre, considerando cenários do presente, do passado e do futuro (Heikkinen et al., 2006; Franklin et al., 2013; Sillero et al., 2013). Essa é uma ferramenta poderosa, por exemplo, como ponto de partida para a seleção de áreas para excursões exploratórias e inventários de biodiversidade (Jimenez-Valverde et al., 2008). Além disso, pode-se fazer previsões de mudanças nos padrões de distribuição de espécies em resposta aos efeitos da mudança climática global (Pacifi et al., 2015). No entanto, todas as vantagens oferecidas pelos ESMs perdem o

seu valor se algum cuidado não for tomado. Esses cuidados devem ser tomados dos pontos de vista conceitual e metodológico. Apesar da extensa literatura que discute a aplicação ou mesmo a designação desses modelos (Franklin, 2009; Peterson & Sobrerón, 2012; Guisan et al., 2017), pesquisadores que desejam entrar no universo de modelagem de adequabilidade ambiental podem sentir falta de um texto que aborde as questões conceituais e metodológicas centrais ou até mesmo um texto mais conciso, que deve não ser exaustivamente dedicado a fórmulas.

Limitações em questões conceituais e metodológicas que apoiam o uso de ESMs ainda são frequentes. Uma vez identificada essa lacuna de conhecimento, discutiremos questões centrais sobre o uso dessa ferramenta, destacando o que chamamos de “armadilhas da modelagem de adequação ambiental”. Nosso foco é apresentar um quadro teórico para pesquisadores que desejam ter lições preliminares sobre modelagem ecológica. Considerando isso, primeiro separaremos as questões conceituais das metodológicas para facilitar nossa abordagem. Também serão apresentados os erros mais frequentes encontrados em artigos publicados em periódicos científicos e algumas formas de evitar tais erros ou minimizar seus efeitos.

## **2. CUIDADOS CONCEITUAIS**

Conceitos são partes fundamentais da ciência, que devem ser tão claras e universais quanto possível, para que possam ser usadas em todos os lugares e campos da ciência sem perda ou distorção em seu significado. Portanto, antes de começarmos a discutir os erros metodológicos, é necessário esclarecer alguns problemas conceituais que frequentemente aparecem em artigos que usam ESMs. Esses erros podem levar a equívocos e interpretações errôneas dos resultados, tornando o produto final não confiável.

O primeiro erro conceitual está relacionado ao nome da ferramenta de modelagem. Antes de escolher um título ou até mesmo escrever um texto sobre ESMs, é necessário pensar no que está sendo modelado. Sem uma definição clara, os leitores serão confundidos sobre o que os resultados representam e qual é a sua aplicabilidade. Essa dúvida pode levar ao uso indevido dos resultados apresentados pelos modelos, simplesmente porque não há consenso sobre como eles devem ser chamados (Guisan et al., 2017) e isto pode gerar variadas formas de interpretar o processo de modelagem de

adequabilidade (Austin, 2002; Soberón & Peterson, 2005). A fim de padronizar os termos em modelos de adequabilidade, alguns cientistas tentaram propor um termo abrangendo todos os tipos de trabalho que usam os ESMs (Peterson & Soberón, 2012; Guisan et al., 2017). Em uma busca rápida no indexador de trabalho científico popular 'Google Scholar' (<https://scholar.google.com>) é comum encontrar obras que usam os termos “Modelagem de Distribuição de Espécies” (40,81%), “Modelagem de Nicho Ecológico” (40,43%)", e mais recentemente, “Modelagem de Adequabilidade de Habitat” (17, 2%) e “Modelagem de Nicho Ambiental” (1,74%).

Modelagem de Distribuição de Espécies - obtivemos 211 resultados em uma busca rápida usando as palavras-chave "modelagem de distribuição de espécies" como palavras-título no Google Scholar, onde filtramos os artigos dentro do intervalo entre 2010 e 2017. Escolhemos este período porque a grande ascensão das ESMs iniciou-se em 2010, o que pode ser confirmado pelo número crescente de trabalhos publicados no campo, e 2017 foi escolhido como ano limite superior para evitar interferências de artigos publicados no ano em curso (2018). Aqui definimos a modelagem de distribuição de espécies como a real probabilidade de encontrar um indivíduo do grupo modelado, considerando as condições e os recursos necessários para sua persistência na área e também sua capacidade de movimento. Isso inclui a definição utilizada por Franklin et al. (2009) e Guisan et al. (2017).

Modelagem de Nicho Ecológico - Obtivemos 209 registros usando as palavras-chave acima durante nossa busca. De fato, o termo “modelagem de nicho ecológico” foi popularizado após o estudo de Peterson & Soberón (2012), que apresentou excelente discussão sobre qual terminologia é a mais adequada para se referir a modelos de adequabilidade; no entanto, mesmo antes de 2012, essa denominação já estava sendo utilizada (Peterson, 2003; Barve et al., 2011). Quando usamos a palavra "nicho" no texto atual, seguimos Hutchinson (1957), ou seja, consideramos o nicho como um hipervolume composto de  $n$  dimensões onde o conjunto de condições e recursos necessários para sobrevivência do grupo de interesse é satisfeito. Assim, quando nos referimos a ESMs como modelos de nicho ecológico, estaríamos intrinsecamente assumindo que os resultados desses modelos são representativos das condições e recursos que as espécies precisam, e isso não é necessariamente verdade, pois o resultado da modelagem dependerá das variáveis ambientais (preditores) que serão utilizadas, e essas variáveis apenas não conseguem representar a totalidade do

nicho ecológico do grupo de interesse, pois este conta com  $n$  dimensões também. Portanto, se utilizarmos este conceito estaríamos superestimando os resultados da modelagem.

Modelagem de Nicho Ambiental - Este foi o termo menos utilizado para descrever os ESMs nos títulos do artigo, aparecendo apenas nove vezes em nossa pesquisa de um total de 517. Esta baixa representatividade é um pouco incompreensível, uma vez que entre todos os termos mencionados, este é o mais próximo de uma denominação coerente para os ESMs. Essa definição contém um termo-chave para descrever os ESMs: o ambiente. No entanto, enfatizamos que o uso da palavra “nicho” neste contexto pode levar a imprecisões de interpretação, uma vez que o nicho é algo muito complexo para ser expresso apenas em termos ambientais, e as advertências feitas no item anterior sobre o emprego da palavra “nicho” também se aplicam aqui.

Modelagem de Adequabilidade de Habitat - as pesquisas de termos referentes à adequabilidade do habitat retornaram 88 resultados. Este é o último termo proposto para descrever os ESMs e tornou-se mais popular após a publicação de Guisan et al. (2017). Estes autores desenvolveram uma discussão excelente sobre a designação correta para modelos da adequabilidade, mas poderiam ser questionados em definir os modelos como projeções do habitat da espécie. De acordo com o dicionário de Oxford (OD), a definição de habitat é a representação do "ambiente natural", território natural, habitação ou habitação de um organismo (ou seja, animal ou planta). Assim, se considerarmos a definição de OD, o uso de "habitat" não seria uma escolha adequada para descrever o que é de fato produzido pelos ESMs. Os modelos não inferem nada sobre o ambiente natural ou habitação de organismos. O que eles realmente produzem são estimativas das condições ambientais apropriadas com base nos registros de ocorrência.

Entre os 517 artigos, 40,81% utilizaram o termo "modelagem de distribuição de espécies" em seus títulos; 40,43% trabalhavam com "modelagem ecológica de nicho"; 17,2% optaram por "modelagem de adequação de habitat", e apenas 1,74% usaram terminologia de "modelagem de nicho". Consideramos este último uma abordagem mais adequada do que as anteriores, mas a escolha da palavra "nicho" deve ser usada com extrema cautela e com a explicação do que está sendo considerado "nicho ambiental". O conceito de nicho (Hutchinson, 1957; Chase e Leibold, 2003) é um fator central para a

modelagem (Franklin et al., 2009), mas o que os modelos apresentam são, de fato, áreas com adequabilidade ambiental, ou seja, áreas que apresentam condições ambientais (incluindo condições climáticas) para o estabelecimento/desenvolvimento de uma determinada espécie, similar àquelas onde os registros conhecidos da ocorrência estão atuais.

## 2.1 DIAGRAMA BAM

Os valores descritos acima não descrevem os usos reais de cada um dos termos, mas podemos ver quais são os mais utilizados. Cada um destes termos deve ser usado sob uma gama muito estreita de condições (Peterson & Soberón, 2012; Guisan et al., 2017). Além disso, a maioria desses estudos utiliza principalmente dados de precipitação e temperatura como preditores dos modelos (Araújo & Luoto, 2007; Vogler et al., 2013; Váz & Nabout, 2016), e esses descritores não representam a complexidade do nicho ambiental, o nicho ecológico ou a distribuição de espécies (Jiménes-Valverde et al., 2008). Assim, os modelos apenas indicam áreas com condições ambientais semelhantes àquelas em que a espécie foi localizada, considerando apenas as variáveis inseridas no modelo para indicar as áreas onde a espécie poderia ser estabelecida. Qualquer outro tipo de interpretação sobre o que os modelos representam poderá estar errado. O diagrama BAM, inicialmente proposto por Soberón e Peterson (2005), é uma forma muito simples e adequada de representar os principais fatores presentes no trabalho com a distribuição de espécies (Figura 1). Faremos uso desse diagrama aqui, pois temos uma ferramenta muito útil para representar fatores fundamentais para o conhecimento ecológico sobre as espécies e, conseqüentemente, para a produção de ESMs.

Entre os fatores apresentados no diagrama da Figura 1, apenas a intersecção entre dois deles (AM) é utilizada na construção dos ESMs. Como mencionado anteriormente, os modelos contêm apenas variáveis ambientais (abiótico - A) e, em alguns casos raros, o uso de dados de dispersão (movimento M) também é possível, e desta forma teríamos uma melhor representação do G, que corresponde aos locais onde possivelmente encontraríamos indivíduos do grupo de interesse (Soberón & Peterson 2005).

Outro erro conceitual muito frequente aparece quando os autores tratam os valores de adequação gerados pelos modelos como probabilidade de presença da espécie, ou seja, como um valor de probabilidade de encontrar um indivíduo do grupo modelado.

Essa interpretação está errada. Os modelos apresentam projeções derivadas de um espaço ambiental, ou seja, não representam todas as condições no espaço geográfico e também não representam a probabilidade de encontrarmos um indivíduo do grupo de interesse. Na próxima etapa, os dados associados a esses pontos são extraídos das variáveis fornecidas (camadas ambientais) e as projeções são feitas com base nos dados extraídos (Figura 2).

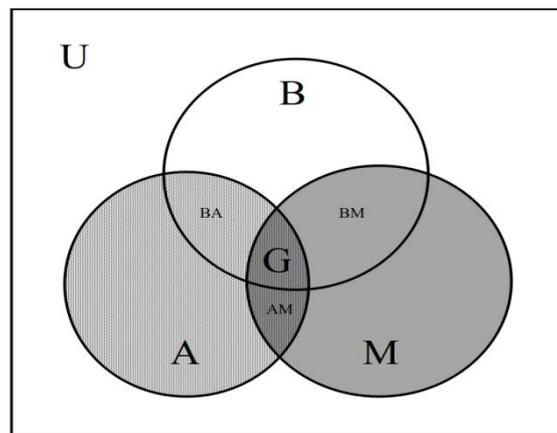


Figura 1. O diagrama BAM é uma forma de representar os principais grupos de fatores que influenciam os padrões de distribuição das espécies, onde B representa aqueles fornecidos por interações bióticas, A representa as condições abióticas exercidas pelo ambiente e M consiste na influência exercida pelo deslocamento, capacidade da espécie, ou movimento. Todos esses componentes estão dentro de um espaço chamado “universo”, representado aqui pela letra U. O intercepto BA representa condições bióticas e abióticas favoráveis, mas as áreas são inacessíveis (-M). BM consiste em condições bióticas favoráveis e acessíveis, mas condições abióticas favoráveis não estão presentes; AM descreve locais com condições abióticas favoráveis e acessíveis, mas interações favoráveis com outras espécies não estão presentes. Finalmente, a área representada por G consiste no nicho realizado, isto é, a área que é realmente ocupada pelo grupo de interesse (Soberón & Peterson 2005).

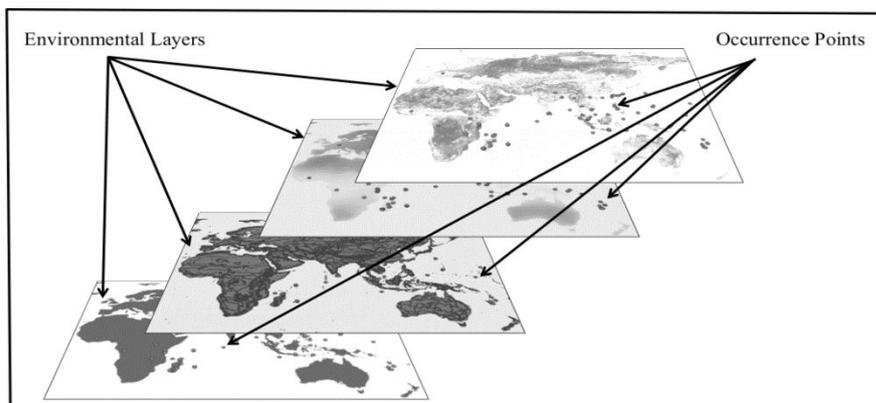


Figura 2. Camadas ambientais com pontos de ocorrência. Cada ponto traz os dados das camadas ambientais correspondentes; esses dados serão usados pelos algoritmos para criar ESMs.

Em seguida, esses dados são projetados no espaço ambiental, onde os dados correspondentes a cada uma das variáveis serão tratados em um plano bidimensional. Em seguida, os algoritmos correlacionam os pontos com os dados ambientais associados a eles para encontrar condições ambientais similares (Figura 3).

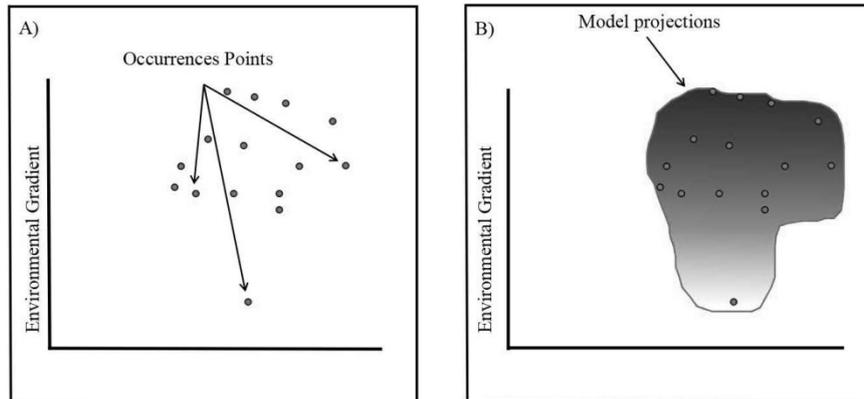


Figura 3. Diagrama representando o espaço ambiental ao longo de um gradiente ambiental (Temperatura no eixo X e Precipitação no eixo Y) e as projeções do modelo referente a esses pontos. A) Pontos de ocorrência dispostos em um plano bidimensional; B) Pontos de ocorrência dispostos em (e influenciados por) um gradiente ambiental.

### 3. CUIDADOS METODOLÓGICOS

Como mencionado anteriormente, o uso de ESMs aumentou, mas mesmo com esses avanços, não se deve negligenciar aspectos centrais da metodologia. O primeiro aspecto que abordaremos aqui está relacionado à compilação de dados de ocorrência de espécies. O simples ato de se pesquisar dados de ocorrências na Web ou mesmo em museus de história natural e outras coleções individuais não garante um conjunto de dados robusto. Eisenlohr e Oliveira-Filho (2015) apresentam algumas técnicas para aumentar a confiabilidade dos metadados utilizados na construção dos modelos. Vamos mencionar aqui algumas dessas técnicas, propor outras, e fazer algumas sugestões para evitar problemas relacionados aos dados de ocorrência e às variáveis preditoras dos modelos.

#### 3.1 Registros de Ocorrência

Um dos primeiros passos para aqueles que querem usar ESMs para responder suas perguntas é encontrar dados de ocorrência de espécies ou mesmo grupos de interesse. No entanto, isso pode se tornar uma tarefa mais árdua do que o planejado originalmente, uma vez que em trabalhos que visam investigar espécies raras nem sempre há registros

de ocorrência dessas espécies devido à dificuldade de encontrar espécimes em bancos de dados.

Os dados sobre a ocorrência de espécies são frequentemente pesquisados em grandes bancos de dados de espécies como o GBIF (<https://www.gbif.org/>). Há também bancos de dados nacionais relevantes que devem ser consultados, como, por exemplo, o SpeciesLink (<http://www.splink.org.br>) para a biota brasileira. Usar todos os registros de ocorrência obtidos de tais bancos de dados para construir ESMs é um problema sério (Beck et al., 2014). Outro problema é que nem sempre podemos ter certeza de que os dados registrados no banco de dados correspondem às mesmas espécies que estamos estudando, ou seja, podemos ter um viés taxonômico (Troudet et al., 2017). Uma alternativa para contornar esse problema é procurar dados de ocorrência em bancos de dados que forneçam imagens e/ou determinadores. Se esses dados forem insuficientes ou se for necessária uma amostragem mais completa, as buscas poderão ser realizadas em periódicos especializados abrigando o grupo que será modelado (espécie, gênero ou até mesmo um grupo ecológico), e esta abordagem é recomendada para todos os grupos. Embora o banco de dados ganhe confiabilidade, não há garantia de que o pesquisador tenha dados taxonômicos limpos e totalmente confiáveis, mas esse procedimento reduz a probabilidade de se trabalhar com dados altamente tendenciosos.

Mesmo quando temos confiança na identificação das espécies e/ou podemos verificar se essa identificação está correta, ainda temos a possibilidade de que o georreferenciamento dos pontos de ocorrência, quando presentes, não seja muito preciso (Beck et al., 2014). Nessa situação podemos adotar alguns procedimentos para ganhar confiabilidade que os dados georreferenciados correspondam à localidade informada. Um desses procedimentos é verificar se os coletores fizeram uma descrição do local onde a amostragem foi obtida on-line, pois, mesmo com georreferenciamento de baixa qualidade, ainda podemos recuperar as coordenadas da localidade através de softwares ou plataformas como o Google Earth (Guralnick et al., 2007). Essa não é a melhor maneira de realizar uma coleta de dados ou limpeza, mas no caso de alguns pontos, pode ser a única maneira de tornar o trabalho viável. Outra maneira de ter mais confiança em nossos dados é seguir a recomendação de Eisenlohr e Oliveira-Filho (2015), que consiste em traçar pontos de ocorrência em um *shapefile* usando um ambiente GIS (ou seja, QGIS, DIVA-GIS ou até mesmo funções em R meio Ambiente). Após plotar os pontos é recomendado selecionar apenas pontos que caiam em zonas que

contenham condições ambientais mínimas para a ocorrência do grupo de interesse e desconsiderar aqueles pontos que aparecem em zonas que são muito distintas ambientalmente daquelas zonas onde os demais pontos estão localizados. Também indicamos o uso de filtros ambientais, que podem reduzir o viés de amostragem. Para mais detalhes sobre o uso de filtros ambientais, veja Varela et al. (2014).

### **3.2 Variáveis Exploratórias**

Depois de concluir o processo de limpeza dos dados de ocorrência, é hora de escolher quais variáveis serão usadas como preditoras dos ESMs. Esse passo pode parecer trivial comparado aos descritos acima, pois teoricamente trata-se de uma fase intuitiva e que não necessitaria de muitos cuidados, mas isso não é verdade. Esses modelos devem apresentar informações relevantes sobre as curvas de resposta das espécies, as quais são apresentadas aqui sob a forma de adequação ambiental da espécie (Austin, 2002; Chunco et al., 2013). Por esse motivo, o uso de variáveis que têm uma influência relevante nos processos que condicionam a distribuição é substancial. Nesta seção, apresentaremos alguns cuidados que devemos tomar ao selecionar variáveis para a construção do modelo e também apresentaremos métodos para lidar com cada um dos possíveis problemas.

A escolha de quantas variáveis devem ser usadas em nossos modelos desempenha um papel central na construção dos ESMs, e não existe uma “fórmula mágica” para executar tal tarefa. Os pesquisadores devem usar, sempre que possível, um conjunto de variáveis com o objetivo de representar as condições ambientais mais importantes para o estabelecimento da espécie da forma mais completa possível, desde que haja informações suficientes sobre a biologia e ecologia do grupo de interesse (Peterson et al., 2011). Além disso, usar um conjunto muito grande de variáveis ambientais pode resultar em modelos com alta complexidade e interpretação difícil (Guisan & Zimmermann, 2000).

Outra preocupação comum está relacionada à colinearidade ou multicolinearidade, ou seja, um alto grau de redundância entre as variáveis preditoras (Dormann et al., 2013). A modelagem gerada com variáveis muito colineares mostra a inflação de erros de comissão, ou seja, o modelo prevê uma área como adequada para a espécie quando esta não é (Pearson, 2007). Por esses motivos, recomendamos nunca usar variáveis sem verificar a colinearidade entre as variáveis. Essa verificação pode ser realizada através

do cálculo do VIF (fator de inflação de variância) (Booth et al. 1994). VIFs acima de 20 indicam forte colinearidade, mas VIFs acima de 10 devem ser examinados e evitados (Borcard et al., 2011). Entretanto, se for necessário ou importante inserir muitas variáveis, sugerimos utilizar os eixos de uma PCA (Análise de Componentes Principais) como preditores, uma vez que eles são ortogonais e não apresentam colinearidade (Dormann et al., 2013).

Também é preferível que as variáveis preditoras estejam na mesma escala, garantindo que os resultados realmente representem a unidade que desejamos apresentar (Guisan et al., 2017). Variáveis em diferentes escalas podem ter seus valores interpretados de maneira diferente, pois variáveis com valores numéricos mais elevados serão tratadas como variáveis que exercem maior influência nas variáveis de resposta (Quinn & Keough, 2002).

#### **4. UMA PROPOSTA DE ROTINA PARA ESM**

Para obter um modelo de aplicabilidade robusto, confiável e prático, sugerimos duas rotinas disponíveis em <<https://github.com/pedroeisenlohr>>, que apresentam uma forma de desenvolvimento sequencial e clara de aplicar ESMs. Apresentaremos cada uma das rotinas de modelagem separadamente para melhor compreensão. Todas as rotinas foram totalmente desenvolvidas no ambiente estatístico R (R Development Core Team, 2018). As rotinas possuem duas versões: uma para o desenvolvimento de ESMs para o tempo atual, que possui um conjunto de variáveis compiladas de diversas fontes, e a outra desenvolvida para criar projeções para o futuro (2050 e 2070) com base nas projeções do quinto relatório do Painel Intergovernamental sobre Mudanças do Clima - IPCC (IPCC, 2007).

Para construir este framework foi utilizado o pacote “biomod2” (Thuiller et al., 2013), que atualmente possui 11 algoritmos (Modelo Linear Geral - GLM; Modelo Aditivo Generalizado - GAM; Modelo Generalizado de Aprimoramento - GBM; Análise de Árvore de Classificação - CTA; Redes Neurais Artificiais - ANN; Envelope de Superfície - SER (também conhecido como BIOCLIM); Análise Discriminante Flexível - FDA; Múltiplas *Splines* Adaptativas de Regressão - MARS; *Random Forest* - RF; e Máxima Entropia - MAXENT-Phillips e MAXENT-MAXENT.Tsuruoka (Phillips et al.,

2006; Thuiller et al., 2009) O “biomod2” também fornece sete métricas para avaliar a capacidade preditiva dos modelos: Área Sob a Curva de Características Operacionais Relativas - AUC, True Skill Statistic - TSS, Taxa de Sucesso - SR, Taxa de Falso Alarme - FAR, Probabilidade de Detecção - POD, PRECISÃO e Escore de Viés - BIAS (Thuiller et al. 2016). Somente os dois primeiros serão utilizados nas rotinas propostas, uma vez que são os mais adequados para o tipo de modelagem aqui abordado (Allouche et al. 2006; Peterson et al. 2008; van Zonneveld et al., 2014). No entanto, essas métricas avaliativas devem ser restritas aos casos em que muitos pontos de ocorrência ( $> 25$ ) estão disponíveis, como sugerido por Pearson et al. (2007). A capacidade preditiva de modelos produzidos a partir de pequenos conjuntos ( $<25$  pontos de ocorrência) deve ser avaliada usando uma estrutura Jackknife (leave-one-out) (Pearson et al. 2007).

#### **4.1 ESM Sob Clima Presente**

Depois de ter verificado e corrigido possíveis problemas com a matriz de ocorrências na rotina de modelagem atual e também selecionar apenas registros de ocorrência espacialmente exclusivos, para que possamos ter o número de graus de liberdade verdadeiro e também para calibrarmos os dados de teste e treino corretamente, uma Análise de Componentes Principais (PCA) pode ser aplicada para evitar possíveis problemas de colinearidade. Aqui, um número de eixos representando uma fração muito expressiva (por exemplo, 90%) da variação contida nesses dados é retido; esses eixos serão usados como variáveis preditoras dos modelos. Assim, como mencionado anteriormente, eliminamos qualquer colinearidade entre as variáveis, uma vez que os eixos da PCA são matematicamente independentes um do outro (Dormann et al., 2013). Nesta rotina, um conjunto de aproximadamente 70 variáveis é usado. Essas variáveis incluem descritores climáticos, de solo, de cobertura vegetal e de relevo e topografia, mas o pesquisador que pretende utilizar essa rotina pode utilizar o conjunto de variáveis que melhor atenda às suas questões de trabalho.

Após todas as verificações nos dados de ocorrência e nas variáveis preditoras, é hora de formatar os dados para iniciar a modelagem. Nesta etapa, prepararemos os dados a serem usados nos modelos, ou seja, estabeleceremos os parâmetros necessários para realizar a modelagem. Primeiramente, selecionamos parte dos dados de ocorrência para a construção dos modelos e parte para fins de avaliação dos modelos (aqui usamos 70%

e 30%, respectivamente; Fielding et al., 1997). Em seguida, separamos o conjunto de 10 algoritmos em dois grupos, pois três algoritmos (GBM, CTA e RF) exigem configurações diferentes (Barbet-Massin et al., 2012). Esses autores recomendam uma calibração para esses três algoritmos de modo que o número de pseudoausências (PAs) seja igual ao número de registros de ocorrência da espécie. No entanto, outros algoritmos (GLM, GAM, ANN, SRE, FDA, MARS e MAXENT) podem usar uma calibração com 10.000 PAs (Barbet-Massin et al., 2012).

Nesta etapa da rotina, os modelos serão construídos com base nos pontos de ocorrência e variáveis ambientais fornecidos. Após essa etapa, os modelos são avaliados por dois critérios: Área Sob a Curva ROC (Característica Operacional Relativa) - AUC e True Skill Statistic - TSS. Após a avaliação descartamos os modelos que não apresentam um bom desempenho de acordo com as métricas de avaliação, os quais definimos aqui como apresentando valores de TSS < 0,4, porque apresentam baixa acurácia (Allouche et al., 2006; Zhang et al., 2015). O próximo passo é criar as projeções, ou seja, os mapas de adequabilidade ambiental. Essas projeções são inicialmente separadas por algoritmos e “empilhadas” para criar um modelo de consenso contínuo com as projeções dos modelos selecionado.

Além de um modelo contínuo, essa rotina também oferece a possibilidade de criação de modelos binários (presença-ausência) para aqueles que precisam tomar uma decisão de “ocorre ou não ocorre”. Estes modelos são produzidos com a aplicação de um limiar que determina que áreas serão consideradas adequadas (isto é, valores acima deste limiar) e áreas que serão consideradas inadequadas (isto é, valores abaixo do limiar). Optamos por utilizar o limiar que maximiza a sensibilidade e a especificidade (limiar ROC), o que implica em menor superpredição (Liu et al., 2005).

## **4.2 ESMs Sob Cenários de Clima Futuro**

A segunda rotina de construção de ESMs que apresentamos é capaz de gerar projeções para os anos 2050 (2041-2060) e 2070 (2061-2080) usando as variáveis disponíveis no banco de dados bioclimático CHELSA 1.2 (ver <http://chelsa-climate.org/>). Estes dados são constituídos de um conjunto de 19 variáveis, das quais 11 são variáveis de temperatura e 8 são variáveis de precipitação. Nesta base de dados existem conjuntos de variáveis que representam as condições climáticas esperadas para o

futuro da Terra sob vários cenários das *Representative Concentration Pathways* (RCP), conforme apresentado pelo quinto relatório do IPCC. Aqui usamos dados de sete modelos de circulação geral (GCM). Os GCMs simulam o comportamento da circulação atmosférica planetária e, portanto, podem ser usados para inferir respostas climáticas em larga escala (Mechoso e Arakawa, 2015). Atualmente, o banco de dados CHELSA contém 30 GCMs, mas nesta rotina usamos apenas 7 (CCMSM4, CMCC-CM, CSIRO-Mk3-5-0, GFDL-CM3, HadGE M2-CC, MIROC5 e MIROC-ESM), pois eles apresentam a maior variação nos dados de acordo com a recomendação disponível na página web do CHELSA.

No caso dessa rotina, trabalhamos apenas com o cenário de emissão de carbono considerado pessimista (rcp 8.5), que se aproxima das observações reais. Essa rotina segue todas as etapas da verificação dos dados de ocorrência descritos anteriormente. O método de verificação de colinearidade das variáveis ambientais usadas aqui é o Variance Inflation Factor - VIF, mas o uso de PCA também é possível (por exemplo, Zwiener et al., 2017). A análise tem como objetivo detectar a colinearidade entre os 19 preditores e selecionar um subconjunto cujas variáveis apresentam um VIF inferior a 10 (Quinn e Keough, 2002; Borcard et al., 2011). Verificação dupla deve ser feita no conjunto de variáveis para detectar variáveis com informações redundantes. A primeira verificação de colinearidade é realizada nas variáveis associadas aos pontos, e a segunda, nas variáveis extraídas de 10.000 pontos distribuídos aleatoriamente ao longo do background. Adotando este *framework*, nós controlamos o problema da colinearidade entre as variáveis preditoras. Para o desenvolvimento dessas rotinas, foram utilizados os dados disponíveis no banco de dados CHELSA (Karger et al., 2017), mas com poucas mudanças pode-se usar qualquer conjunto de variáveis que forneçam projeções para o intervalo de tempo requerido. O procedimento para formatar os dados e criar os modelos é descrito na seção 4.1, e eles não diferem da calibração usada aqui. O próximo passo é avaliar esses modelos; a partir de agora, apenas modelos com TSS > 0,4 serão utilizados (Zhang et al., 2015).

Em seguida, os modelos consensuais são gerados para cada algoritmo individualmente. Então, as projeções são geradas para um determinado período (por exemplo, anos 2050 e 2070). Os procedimentos para a geração de projeções para 2050 e 2070 anos não diferem. Por fim, essas projeções futuras são agrupadas por algoritmos

ou por GCM, para gerar o consenso médio somado de cada algoritmo ou de cada GCM, binário ou contínuo.

## **5. CONSIDERAÇÕES FINAIS**

ESMs são ferramentas muito poderosas e possuem flexibilidade considerável. No entanto, devemos estar atentos aos conceitos empregados durante a fase inicial da produção do ESM. Se ESMs não forem construídos considerando uma base teórica sólida e com métodos adequados, seus resultados serão nada mais que um conjunto de figuras “bonitas” com pouca relevância para o mundo científico. Acreditamos também que a Modelagem de Adequabilidade Ambiental é a melhor nomenclatura para descrever o que tais modelos são de fato.

Destacamos também que ESMs são desenvolvidos usando subconjuntos de variáveis preditoras e correlacionam dados de ocorrência de espécies com dados ambientais para prever áreas com adequabilidade ambiental para um grupo de interesse. Dessa forma, os mapas de saída estão diretamente relacionados à seleção dessas variáveis. No entanto, mesmo com dados de presença confiáveis, a seleção de variáveis não é uma tarefa simples, pois cada espécie precisa de uma infinidade de condições e recursos ecológicos, e essa informação nem sempre está disponível. Assim, deve-se ter cautela para minimizar os efeitos desses fatores.

Aqui listamos erros que, em nossa opinião, são os mais comuns encontrados na produção de ESMs. No entanto, muitos outros podem ocorrer. Assim, é de extrema importância que todos aqueles que trabalham com conservação biológica ou qualquer outra área relacionada que desejem se aventurar no campo dos ESMs tenham suas perguntas-guia bem estabelecidas. Sem isso, a chance de extrair soluções inúteis dessa imensidão de formas e fórmulas é muito grande, tornando o pesquisador mais propenso a cair em armadilhas como as mencionadas aqui.

Para finalizar, apresentamos uma rotina robusta que atua contra os possíveis erros metodológicos aqui abordados. Ressaltamos que os ESMs são ferramentas muito bem-vindas para ajudar no contexto da conservação da biodiversidade, não apenas para as florestas tropicais, mas também para qualquer outro grupo ou ecossistema.

## 6. REFERÊNCIAS

- Allouche, O., Tsoar A., Kadmon R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232. doi: 10.1111/j.1365-2664.2006.01214.x.
- Araújo, M. B., Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Glob Ecol Biogeogr* 16:743–753. doi: 10.1111/j.1466-8238.2007.00359.x.
- Austin, M. P. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecol Modell* 157:101–118. doi: 10.1016/S0304-3800(02)00205-3.
- Barbet-Massin, M., Jiguet, F., Albert, C. H., Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods Ecol Evol* 3:327–338. doi: 10.1111/j.2041-210X.2011.00172.x.
- Barve, N., Barve, V., Jiménez-Valverde, A., et al. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol Modell* 222:1810–1819. doi: 10.1016/j.ecolmodel.2011.02.011.
- Beck, J., Böller, M., Erhardt, A., Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol Inform* 19:10–15. doi: 10.1016/j.ecoinf.2013.11.002.
- Booth, G. D., Niccolucci, M. J., Schuster, E. G. (1994). Identifying proxy sets in multiple linear-regression - an aid to better coefficient interpretation. *USDA For Serv Intermt Res Stn Res Pap* 1–13.
- Borcard, D., Gillet, F., Legendre, P. (2011). *Numerical Ecology with R, 1<sup>a</sup>*. Springer US, New York.
- Brown, J. H., Stevens, G. C., Kaufman, D. M. (1996). The geographic range: size, shape, and internal structure. *Annu Rev Ecol Syst* 27:597–623. doi: 10.1146/annurev.ecolsys.27.1.597.
- Chase, J. M., Leibold, M. A. (2008). Ecological niches. *Linking classical and contemporary approaches*.
- Chave, J., Leigh, E. G. (2002). A Spatially Explicit Neutral Model of  $\beta$ -Diversity in Tropical Forests. *Theor Popul Biol* 62:153–168. doi: 10.1006/tpbi.2002.1597.
- Chunco, A. J., Phimmachak, S., Sivongxay, N., Stuart, B. L. (2013). Predicting environmental suitability for a rare and threatened species (Lao Newt, *Laotriton laoensi*) using validated species distribution models. *PLoS One* 8:1–13. doi: 10.1371/journal.pone.0059853.

- De Marco-Júnior, P., Diniz-Filho, J. A. F., Bini, L. M. (2008). Spatial analysis improves species distribution modelling during range expansion. *Biol Lett* 4:577–580. doi: 10.1098/rsbl.2008.0210.
- Dormann, C. F., Elith, J., Bacher, S., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop)* 36:027–046. doi: 10.1111/j.1600-0587.2012.07348.x.
- Eisenlohr, P. V., Oliveira-Filho, A. T. (2015). Obtaining metadata for phytogeographic of synthesis works and the NeotropTree database as a case study. In: Eisenlohr PV, Felfili JM, Melo MM de RF de, et al. (eds) *Phytosociology in Brazil: Methods and Case Studies*, Vol. 2. Ed. UFV, Viçosa, pp 387–411.
- Elith, J., Graham, C., Anderson, R., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography (Cop)* 29:129–151. doi: 10.1111/j.2006.0906-7590.04596.x.
- Elith, J., Leathwick, J. R. (2009). Species Distribution Models: Ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst* 40:677–697. doi: 10.1146/annurev.ecolsys.110308.120159.
- Fielding, A. H., Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49. doi: 10.1017/S0376892997000088.
- Franklin, J. (2009). *Mapping species distributions*, 1<sup>a</sup>. Cambridge University Press, Cambridge.
- Franklin, J., Davis, F. W., Ikegami, M., et al. (2013). Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Glob Chang Biol* 19:473–483. doi: 10.1111/gcb.12051.
- Gastón, A., Garcia-Vinas, J. I., Bravo-Fernandez, A. J., et al. (2014). Species distribution models applied to plant species selection in forest restoration: are model predictions comparable to expert opinion? *New For* 45:641–653. doi: DOI 10.1007/s11056-014-9427-7.
- Guisan, A., Thuiller, W., Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models*. Cambridge University Press, Cambridge.
- Guisan, A., Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecol Modell* 135:147–186. doi: 10.1016/S0304-3800(00)00354-9.
- Guralnick, R. P., Hill, A. W., Lane, M. (2007). Towards a collaborative, global infrastructure for biodiversity assessment. *Ecol Lett* 10:663–672. doi: 10.1111/j.1461-0248.2007.01063.x.
- Heikkinen, R. K., Luoto, M., Araújo, M. B., et al. (2006). Methods and uncertainties in bioclimatic envelope modelling under climate change. *Prog Phys Geogr* 30 (6):751–777.
- Hutchinson, G. E. (1957). Population studies: animal ecology and demography. *Bull Math Biol* 53:415–247. doi: 10.1101/SQB.1957.022.01.039.

IPCC (2007) Climate change: Synthesis report. IPCC, Geneva

Jenkins, M. (2003). Prospects for Biodiversity. *Science* (80-) 302:1175–1177. doi: 10.1126/science.1088666.

Jiménez-Valverde, A., Lobo, J. M., Hortal, J. (2008). Not as good as they seem: The importance of concepts in species distribution modelling. *Divers Distrib* 14:885–890. doi: 10.1111/j.1472-4642.2008.00496.x.

Karger, D. N., Conrad, O., Böhner, J., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Sci Data* 4:1–20. doi: 10.1038/sdata.2017.122.

Kerr, J. T. (1997). Species richness, endemism, and the choice of areas for conservation. *Conserv Biol* 11:1094–1100. doi: 10.1046/j.1523-1739.1997.96089.x.

Larsen BB, Miller EC, Rhodes MK, Wiens JJ (2017) Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life. *Q Rev Biol* 92:229–265. doi: 10.1086/693564

Liu, C., Berry, P. M., Dawson, T. P., Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography (Cop)* 28:385–393. doi: 10.1111/j.0906-7590.2005.03957.x.

Liu, C., White, M., Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography (Cop)* 34:232–243. doi: 10.1111/j.1600-0587.2010.06354.x.

Mora C, Tittensor DP, Adl S, et al (2011) How Many Species Are There on Earth and in the Ocean? *PLoS Biol* 9:e1001127. doi: 10.1371/journal.pbio.1001127

Mechoso, C. R., Arakawa, A. (2015). *Numerical Models: General Circulation Models*, Second Edi. Elsevier.

Oliveira-Filho, A. T. (2017). NeoTropTree, Flora arbórea da Região Neotropical: Um banco de dados envolvendo biogeografia, diversidade e conservação. *Universidade Federal de Minas Gerais*. (<http://www.neotroptree.info>).

Pacifici, M., Foden, W. B., Visconti, P., et al. (2015). Assessing species vulnerability to climate change. *Nat Clim Chang* 5:215–225. doi: 10.1038/nclimate2448.

Pearman, P. B., Guisan, A., Broennimann, O., Randin, C. F. (2008). Niche dynamics in space and time. *Trends Ecol Evol* 23:149–158. doi: 10.1016/j.tree.2007.11.005.

Pearson, R. G. (2007). Species's distribution modelling for conservation educators and practitioners. *Lesson Conserv* 54–89.

Pearson, R. G., Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Glob Ecol Biogeogr* 12:361–371. doi: 10.1046/j.1466-822X.2003.00042.x.

- Peterson, A. T., Nakazawa, Y. (2008). Environmental data sets matter in ecological niche modelling: An example with *Solenopsis invicta* and *Solenopsis richteri*. *Glob Ecol Biogeogr* 17:135–144. doi: 10.1111/j.1466-8238.2007.00347.x.
- Peterson, A. T. (2003). Predicting the geography of species' invasions via ecological niche modeling. *Q Rev Biol* 78:419–33. doi: 10.1086/ 378926.
- Peterson A. T., Soberon, J. (2012). Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *Nat Conserv* 10:102–107. doi: 10.4322/natcon.2012.019.
- Peterson, A. T., Soberón, J., Pearson, R. G., et al. (2011). *Ecological niches and geographic distributions*.
- Phillips, S. J., Anderson, R. P., Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190:231–259. doi: 10.1016/j.ecolmodel.2005.03.026.
- Quinn, G., Keough, M. (2002). *Experimental Design and Data Analysis for Biologists*, 2<sup>a</sup>. Cambridge University Press, Cambridge.
- R Core Team (2018). *R: A language and environment for statistical computing*.
- Sillero, N., Carretero, M. A. (2013). Modelling the past and future distribution of contracting species. The Iberian lizard *Podarcis carbonelli* (Squamata: Lacertidae) as a case study. *Zool Anz* 252:289–298. doi: 10.1016/j.jcz.2012.08.004.
- Siqueira, M. F. de, Durigan, G., de Marco-Júnior, P., Peterson, A. T. (2009). Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *J Nat Conserv* 17:25–32. doi: 10.1016/j.jnc.2008. 11.001.
- Slik, J. W. F., Arroyo-Rodríguez, V., Aiba, S-I., et al. (2015). An estimate of the number of tropical tree species. *Proc Natl Acad Sci* 112:7472–7477. doi: 10.1073/pnas.1423147112.
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecol Lett* 10:1115–1123. doi: 10.1111/j.1461-0248.2007.01107.x.
- Soberon, J., Nakamura, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proc Natl Acad Sci* 106:19644–19650. doi: 10.1073/pnas.0901637106.
- Soberón, J., Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Informatics* 2:1–10.
- Thuiller, W., Georges, D., Engler, R., Breiner, F. (2016). *Ensemble Platform for Species Distribution Modeling*.
- Thuiller, W., Lafourcade, B., Engler, R., Araújo, M. B. (2009). BIOMOD - A platform for ensemble forecasting of species distributions. *Ecography (Cop)* 32:369–373. doi: 10.1111/j.1600-0587.2008.05742.x.
- TrouDET, J., Grandcolas, P., Blin, A., et al. (2017). Taxonomic bias in biodiversity data and societal preferences. *Sci Rep* 7:1–14. doi: 10.1038/s41598-017-09084-6.

- van Zonneveld, M., Castañeda, N., Scheldeman, X., et al. (2014). Application of consensus theory to formalize expert evaluations of plant species distribution models. *Appl Veg Sci* 17:528–542. doi: 10.1111/avsc.12081.
- Varela, S., Anderson, R. P., García-Valdés, R., Fernández-González, F. (2014). Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography (Cop)* 37:1084–1091. doi: 10.1111/j.1600-0587.2013.00441.x.
- Vaz, Ú. L., Nabout, J. C. (2016). Using ecological niche models to predict the impact of global climate change on the geographical distribution and productivity of *Euterpe oleracea* Mart. (Arecaceae) in the Amazon. *Acta Bot Brasilica* 30:1–6. doi: 10.1590/0102-33062016abb0036.
- Vogler, R. E., Beltramino, A. A., Sede, M. M., et al. (2013). The giant *African snail*, *Achatina fulica* (Gastropoda: Achatinidae): Using bioclimatic models to identify South American areas susceptible to invasion. *Am Malacol Bull* 31:39–50. doi: 10.4003/006.031.0115.
- Warren-Rhodes, K. A., Rhodes, K. L., Pointing, S. B., et al. (2006). Hypolithic cyanobacteria, dry limit of photosynthesis, and microbial ecology in the hyperarid Atacama Desert. *Microb Ecol* 52:389–398. doi: 10.1007/s00248-006-9055-7.
- Willerslev, E., Hansen, A. J., Christensen, B., et al. (1999). Diversity of Holocene Life Forms in Fossil Glacier Ice. *Source Proc Natl Acad Sci United States Am* 96:8017–8021. doi: 10.1073/pnas.96.14.8017.
- Wilson, E. (2000). *Sociobiology : the new synthesis*. Belknap Press of Harvard University Press.
- Wilson, E. O. (2003). The encyclopedia of life. *Trends Ecol Evol* 18:77–80. doi: 10.1016/s0169-5347(02)00040-x.
- Wynn-Williams, D. D. (1996). Antarctic microbial diversity: The basis of polar ecosystem processes. *Biodivers Conserv* 5:1271–1293. doi: 10.1007/BF00051979.
- Zhang, L., Liu, S., Sun, P., et al. (2015). Consensus Forecasting of Species Distributions: The Effects of Niche Model Performance and Niche Properties. *PLoS One* 10:1–18. doi: 10.1371/journal.pone.0120056.
- Zwiener, V. P., Lira-Noriega, A., Grady, C. J., et al. (2017). Climate change as a driver of biotic homogenization of woody plants in the Atlantic Forest. *Glob Ecol Biogeogr* 27:298-309. doi: 10.1111/geb.12695.

**CAPÍTULO 2 - MENOS PODE SER MAIS: UMA ROTINA  
OBJETIVA PARA SELECIONAR MODELOS DE CIRCULAÇÃO  
GERAL E SUAS APLICAÇÕES NA MODELAGEM ECOLÓGICA**

*João Carlos Pires-Oliveira, Pedro V. Eisenlohr & Diogo S. B. Rocha*

## RESUMO

A modelagem ecológica (ME) apresenta a possibilidade de se criar projeções espaciais e/ou temporais de superfícies de adequabilidade climática/ambiental de organismos ou grupos de interesse. Os pilares para a aplicação desse tipo de ferramenta são constituídos por inferências matemáticas sobre o clima na forma de modelos de circulação geral (GCMs). Porém, devido ao grande número de GCMs disponíveis, a escolha de um conjunto que não contenha redundância, ou que contenha a menor redundância possível, não é trivial. Neste trabalho, desafiamos a necessidade de se utilizar muitos GCMs em ME e propomos uma rotina para realizar a seleção de GCMs utilizando três técnicas complementares (ordenação, agrupamento, correlação linear e espécies virtuais). Aplicamos essa rotina sobre os GCMs disponíveis na base de dados CHELSA para identificar quantos e quais GCMs podemos utilizar sem perder informação relevante. Com o nosso método, selecionamos três GCMs dos 33 disponíveis na CHELSA e avaliamos desempenho deles em prever a adequabilidade de espécies virtuais e os indicamos para serem utilizados em ME, pois eles são menos correlacionados com os demais e, dessa forma, contribuem com maior variação nos dados. Nosso trabalho aponta direções importantes para a ME em diferentes escalas temporais face ao obstáculo da subjetividade na escolha de modelos climáticos de projeções pretéritas e/ou futuras.

**Palavras-chave:** Análise Multivariada, GCM, Modelagem de Adequabilidade Ambiental, Validação de Modelos, Espécies Virtuais.

## 1. INTRODUÇÃO

A Modelagem Ecológica (ME) apresenta grande diversidade de técnicas, alta flexibilidade em suas aplicações (Siqueira et al., 2009, Vogler et al., 2013, Gastón et al., 2014, Souza & De Marco Jr., 2014) e retorna resultados significativamente robustos com aplicações potenciais em diversas áreas da biologia e ciências afins (Chunco et al., 2013, Gastón et al., 2014). A utilização de ferramentas de ME para responder a questões sobre distribuição geográfica de organismos, dentre as quais se destaca a Modelagem de Adequabilidade Ambiental (do inglês *Environmental Suitability Modeling* - ESM, por vezes referidos como Modelagem de Nicho Ecológico ou Modelagem de Distribuição Potencial de Espécies; Peterson e Soberón, 2012; Guisan et al., 2017), tem particularmente crescido nas últimas décadas, devido à sua capacidade de detectar regiões adequadas ao estabelecimento dos grupos de interesse (Diniz-Filho et al., 2009, Guisan et al., 2017). A ESM dispõe de algoritmos das mais diversas complexidades, variando desde aqueles simples que analisam a distância ambiental entre áreas de ocorrência do grupo de interesse, como, por exemplo, o Bioclim (Rusby, 1991), a algoritmos complexos, como, por exemplo, as Redes Neurais Artificiais (Zhang e Li, 2017). Além de um arcabouço de técnicas, a ESM possui muitas aplicações, como a possibilidade de construir projeções sobre áreas no tempo presente e também para outros intervalos temporais (passado e futuro) (Pearson et al., 2003, Waltari et al., 2007, Diniz-Filho et al., 2009).

Um procedimento comum em ESM é utilizar variáveis climáticas, como temperatura e precipitação para identificar e compreender os padrões de distribuição dos organismos na superfície do planeta, visto que está bem estabelecida a importância do clima nos padrões de distribuição de espécies (Meadows, 1999, Faleiro et al., 2013, Rezende et al., 2014). Esse procedimento ganhou relevância com os alertas emitidos pelo Painel Intergovernamental sobre Mudanças Climáticas (em inglês, *Intergovernmental Panel on Climate Change* – IPCC) de iminentes alterações nos padrões climáticos em um futuro próximo (até 2100) devido à emissão descontrolada de gases de efeito estufa (IPCC, 2013). Embasados nos alertas do IPCC, os organizadores do Projeto de Intercomparação de Modelos Acoplados (em inglês, *Coupled Model Intercomparison Project* – CMIP) projetaram, sobre um conjunto de variáveis bioclimáticas do presente, simulações para tentar prever como serão as condições climáticas no futuro próximo. Dessa forma, os Modelos de Circulação Geral (em inglês,

*General Circulation Models* – GCMs) (IPCC, 2013, Mechoso et al., 2015) passaram a ser utilizados com a finalidade de prever alterações no clima futuro.

Devido à sua proposta de prever mudanças nos padrões climáticos futuros, os GCMs são frequentemente utilizados para identificar alterações espaciais e temporais em ESM (Diniz-Filho et al., 2009, Guisan et al., 2017, Zwiener et al., 2018). Com efeito, uma das etapas cruciais em ESM é selecionar os GCMs para as projeções futuras ou passadas. Sabendo que eles apresentam resultados distintos de predição, muitos autores não utilizam critérios, ou não os informam, para a escolha dos modelos utilizados em seus trabalhos (Waltari et al., 2007, Franklin et al., 2013, Guisan et al., 2013, Guisan et al., 2017), o que pode tornar a escolha questionável devido à proporção de variação contida no grupo selecionado. Uma seleção aleatória de GCMs aumentaria o volume de dados a serem processados, sem acrescentar informação útil à ESM ou, até mesmo, gerando modelos enviesados.

Neste contexto, propusemos a seguinte pergunta: i) Como podemos otimizar o número de GCMs a serem utilizados e reduzir o volume de dados a serem processados sem haver perda de informações relevantes para os modelos? Acreditamos que não seja necessário utilizar todos os GCMs na ESM para gerar bons modelos, pois conforme demonstrado por Knutti et al. (2013), Masson & Knutti (2011) e Sanderson et al. (2015), os GCMs são semelhantes. Esses trabalhos indicam que os GCMs derivados de um mesmo relatório do IPCC também possuem elevada similaridade, especialmente aqueles produzidos pelas mesmas instituições ou instituições próximas (Basson & Knutti, 2011, Knutti et al., 2013). Esse conjunto inflado de dados apenas resultaria em maior esforço na ESM sem que houvesse ganho nas predições. Além disso, utilizar conjunto de GCMs correlacionados resultaria em projeções enviesadas, pois os modelos tenderiam a dar maior importância aos GCMs mais semelhantes e acabariam por desconsiderar os menos correlacionados e isto não é desejado. Porém, os trabalhos que indicaram direções para a seleção de GCMs não consideraram suas implicações nas projeções realizadas a partir de ESM (Overland et al., 2011, Knutti et al. 2013, Evans et al., 2013, Sanderson et al., 2015). Guiados pela questão e lacunas acima, buscamos quantificar a correlação entre os GCMs e apresentar uma rotina alternativa para obter um balanço favorável entre utilização de baixo número de GCMs selecionados e retenção de elevada proporção de variação nos dados climáticos. Avaliamos também o desempenho de grupos constituídos por diferentes números de GCMs quando utilizados na ESM.

## 2. MATERIAL E MÉTODOS

### 2.1 Banco de Dados

Há diversas opções de plataformas que disponibilizam dados para modelagem para o presente e também GCMs com projeções dessas variáveis para o futuro, das quais as mais relevantes são *WorldClim* (Hijmans et al., 2005; Fick e Hijmans, 2017) e, recentemente, *CHELSA* (Karger et al., 2017). Ambas oferecem informações climáticas considerando quatro cenários diferentes: “mais otimista” – rcp2.6; “menos otimista” – rcp4.5; “intermediário” – rcp6.0 e “pessimista” – rcp8.5. Consideramos neste estudo a plataforma *CHELSA* por apresentar uma série temporal das projeções mais atual (1979-2013). A *CHELSA* disponibiliza 33 GCMs para o intervalo temporal de 2061-2080 no cenário rcp8.5, que consiste nas previsões mais alarmantes e, provavelmente, mais realistas (Riahi et al., 2011).

Nosso banco de dados consistiu de 19 variáveis bioclimáticas derivadas de temperatura e precipitação (*BIOS*), distribuídas nos 33 GCMs com projeções futuras do clima disponíveis na *CHELSA* (Tabela 1). Utilizamos apenas os GCMs que disponibilizavam variáveis para o intervalo de 2061-2080 no cenário rcp8.5. Os GCMs que não contemplaram esses critérios foram desconsiderados em nossas análises. Apesar da escolha do intervalo temporal e do cenário ser um critério arbitrário, a mesma não enviesava nossas análises, pois o objetivo deste trabalho está direcionado à seleção de GCMs e não às variações temporais e/ou entre cenários de emissões de gases.

Circunscrevemos as variáveis de todos os GCMs à região neotropical e as reescalamos para a resolução de ~10km. Essa resolução, apesar de perdermos em precisão por utilizar tamanhos de células maiores, possui algumas vantagens: i) otimiza a eficiência no esforço computacional quando comparada a resoluções mais finas (ex: ~1 km); ii) é compatível com a realidade geral da ESM, uma vez que os dados de registros de ocorrência de organismos, particularmente aqueles obtidos de plataformas *online*, não são habitualmente tão precisos quanto ~1km (Pires-Oliveira et al., 2019); iii) minimiza a autocorrelação espacial em ESM (Pearson et al., 2007).

Tabela 1. Lista com nomes informados na página da plataforma *CHELSEA* (IN) e as respectivas siglas (SG) para identificar os GCMs. Nas figuras dos nossos resultados, apenas essas siglas são apresentadas.

IN	SG	IN	SG	IN	SG
ACCESS1-0	ACC	bcc-csm1-1	BCC	BNU-ESM	BNU
CanESM2	CAN	CCSM4	CCS	CESM1-BGC	CEB
CESM1-CAM5	CEC	CMCC-CM	CMM	CMCC-CMS	CMS
CNRM-CM5	CNM	CSIRO-Mk3-6-0	CSK	FGOALS-g2	FGG
FIO-ESM	FIM	GFDL-ESM2G	GFG	GFDL-CM3	GFM
GFDL-ESM2M	GFS	GISS-E2-R-CC	GIC	GISS-E2-H	GIE
GISS-E2-H-CC	GIH	GISS-E2-R	GIR	HadGEM2-CC	HAC
HadGEM2-AO	HAO	inmcm4	INM	IPSL-CM5A-LR	IPL
IPSL-CM5A-MR	IPM	MIROC5	MIC	MIROC-ESM-CHEM	MIE
MIROC-ESM	MIM	MPI-ESM-LR	MPL	MPI-ESM-MR	MPM
MRI-CGCM3	MRM	MRI-ESM1	MRS	NorESM1-M	NON

## 2.2 Análise dos Dados

Há muitos métodos disponíveis na literatura para selecionar GCMs (Wilby et al., 1999, Wilby et al., 2002, Pierce et al., 2009, Sanderson et al., 2015, Breach et al., 2016), porém com objetivos que não contemplam a sua aplicação em ESM. Aqui, optamos pela seguinte rotina: 1) realizar uma análise de ordenação (PCoA) para identificar quais descritores mais influenciam a dissimilaridade entre GCMs; 2) gerar um dendrograma utilizando os autovetores da PCoA como descritores; e 3) verificar a formação de grupos grandes ( $\geq 3$  GCMs) e, caso isso ocorra, realizar uma análise de correlação múltipla com índice de *Spearman* par a par entre os integrantes desses grupos grandes para extrair a média dos valores de  $\rho$  e, então, selecionar o GCM com o menor valor de  $\rho$  médio ( $\rho_m$ ); 4) após termos selecionado um grupo de GCMs, utilizamos ESM para avaliar: (i) as projeções produzidas a partir do grupo que selecionamos, (ii) as projeções geradas com todos os GCMs da base *CHELSEA* e (iii) as projeções de ESM produzidas sobre o conjunto sugerido por Sanderson et al. (2015) (CESM1-BGC,

CESM1-CAM5, CMCC-CM, FIO-ESM, GISS-E2-H, INMCM4, IPSL-CM5A-MR, MIROC5 e MPI-ESM-LR). Escolhemos o trabalho de Sanderson et al. (2015), pois os autores apresentam uma forma robusta de seleção de GCMs para selecionarmos apenas os GCMs menos interdependentes conforme desejamos fazer neste trabalho. O método sugerido pelos referidos autores consiste em seleção por meio do critério *stepwise*; e 5) avaliar o desempenho dos ESMs comparando com Espécies Virtuais (VS).

Acessamos os valores de todas as 1.114.060 células de todas as 19 camadas climáticas para os 33 GCMs utilizando a função ‘*values*’ contida no pacote *raster* (Hijmans et al., 2018). Então, extraímos os valores mínimos e máximos e também a mediana e a média de cada *BIO* para todos os GCMs utilizando a função ‘*summary*’ do pacote *base* (R Core Team, 2018). Organizamos a matriz com os GCMs nas linhas e os descritores (valores máximos, mínimos, média e mediana das *BIOs*) nas colunas. Optamos por utilizar essas métricas devido ao fato de que valores unicamente de tendência central (*i.e.*, média e mediana) não representam a amplitude da variação dos dados. Normalizamos os dados para média 0 e desvio padrão 1 e submetemos a matriz resultante a uma análise de coordenadas principais (PCoA) aplicando a função ‘*pcoa*’ do pacote *ape* (Paradis & Schliepcom, 2018) com Distância Euclidiana, conforme sugerido por Borcard et al. (2011) para matrizes ambientais, e Distância de Manhattan, conforme efetuado por Knutti et al. (2013). Com a PCoA, observamos a ordenação dos GCMs em função dos descritores (*i.e.* valores mínimos, máximos, médios e a mediana de cada *BIO*), conforme ilustrado na Tabela 2.

Tabela 2. Representação da estrutura da planilha utilizada para as análises de PCoA e de agrupamento, utilizando os GCMs como objetos (primeira coluna). As colunas contêm, respectivamente, os valores mínimos, a mediana, a média e os valores máximos das variáveis de cada GCM. Para as análise de *cluster* as colunas comportam os eixos produzidos pela PCoA.

GCM	bio01_gcm-x1 _min	bio01_gcm-x1 _median	bio01_gcm-x1_ mean	bio01_gcm-x1 _max
gcm-x1	x	xxxxx	xxxxx	x
gcm-x2	x.x	xxxx	xxx	xx
gcm-x3	x.xx	xxx	x	xxx

Em seguida, realizamos uma análise de agrupamento por meio da função ‘*hclust*’ do pacote *stats*. Para realizarmos esta análise, utilizamos como descritores os eixos produzidos pela PCoA. Nessa análise, utilizamos quatro métodos alternativos de agrupamento (Ligação Simples, Ligação Completa, Ward e Médias Não Ponderadas – UPGMA), todos eles processados tanto com Distância Euclidiana quanto com Distância de Manhattan, sendo selecionado aquele que apresentou o maior coeficiente de correlação cofenética (Borcard et al., 2011). Depois de escolhermos o método de agrupamento, selecionamos o número ideal de grupos por meio da comparação da matriz de distância produzida com uma matriz binária calculada a partir do corte do dendrograma em múltiplos níveis (correspondendo aos múltiplos grupos) por meio da função ‘*grpdist*’ (ver Borcard et al., 2011). Depois, avaliamos a correlação entre essas duas matrizes de distância utilizando o índice de correlação de *Spearman* (adaptado de Borcard et al., 2011), sendo o número de grupos escolhido no nível em que a correlação entre as duas matrizes foi mais alta (veja Figura 2).

De posse dos grupos formados, realizamos uma análise de correlação entre os elementos de grupos com três ou mais GCMs, pois com apenas dois GCMs o  $\rho_m$  seria o mesmo. Em grupos com dois GCMs, efetuamos uma análise de correlação múltipla destes dois integrantes com o vizinho mais próximo e selecionamos aquele que mais diferiu deste. Esse procedimento foi útil para identificarmos o modelo com o menor índice de correlação de *Spearman* –  $\rho$  (i.e. o GCM com menor valor de  $\rho_m$ , com  $p\text{-value} \leq 0.05$ ). O GCM com o menor  $\rho_m$  corresponde àquele que mais diferiu dos demais, sendo, portanto, o GCM selecionado. Essa escolha é lógica, pois esse modelo contém a menor porção de informação redundante e, desta forma, acrescenta informação relevante ao grupo de GCMs selecionados ( $g\text{-GCM}_s$ ). Para obter o  $g\text{-GCM}_s$  geramos um valor de  $\rho$  para cada GCM. Então, obtivemos um valor de  $\rho_m$  de cada modelo com os demais. Finalmente, em cada grupo, o GCM que obteve o menor valor de  $\rho$  foi selecionado para compor o  $g\text{-GCM}_s$ . Realizamos uma Análise de Variância Multivariada Não-Paramétrica (NPMANOVA) com 999 permutações, baseada em distância euclidiana, para testar a existência de diferença entre os grupos, ao nível de significância de 5%.

Posteriormente, para investigar a eficácia da nossa estratégia de seleção de GCMs, utilizamos três partições de GCMs compostas por: Partição 1) os modelos selecionados com  $n = 3$ ; Partição 2) todos os 33 GCMs da CHELSA; e Partição 3) os GCMs indicados por Sanderson et al. (2015), com  $n = 9$ . O GCM ACCESS1-3 foi indicado

pelos referidos autores, porém este não consta na base de dados CHELSA e, por isso, não foi utilizado. Obtivemos o consenso médio somado de cada uma das 19 *Bio* dos GCMs que compõem cada partição, de modo a termos 19 variáveis para cada particionamento. Posteriormente, geramos três VS sobre os *rasters* dos cinco primeiros eixos de PCA obtidos das 19 *Bio* (consenso somado) de cada uma das partições. Geramos as VS com três amplitudes de distribuição (VS 1 – Distribuição intermediária, com prevalência igual a 0.2; VS 2 – Distribuição restrita, com prevalência de 0.02; e VS 3 – Distribuição ampla, com prevalência de 0.5). Utilizamos as configurações de prevalências mencionadas anteriormente para gerar as VS em todas as partições. Criamos as VS por meio da função ‘*generateSpFromPCA*’ do pacote *virtualspecies* (Leroy et al., 2015), ajustada para que as condições nas quais as VS ocorrem fossem idênticas às condições para uma espécie verdadeira na área de estudo. O passo seguinte foi converter os mapas contínuos de adequabilidade das VS para mapas de presença-ausência utilizando uma função probabilística logística. Esta função calcula a probabilidade de obter uma possível “presença da espécie” em um determinado pixel. A probabilidade depende da adequabilidade da espécie nesse pixel. Depois, extraímos 500 pontos aleatórios da área de adequabilidade máxima (*i.e.*, presença) para representar as possíveis ocorrências por meio da função “*sampleOccurrences*” (veja Leroy et al., 2015 e <http://borisleroy.com/virtualspecies/> para mais detalhes).

Realizamos ESM utilizando como dados de ocorrência os pontos obtidos de cada uma das três VS, em cada partição, para a região neotropical. Dessa forma, eliminamos os vieses de coleta associados aos pontos de ocorrência de espécies reais. Utilizamos nove dos 11 algoritmos disponíveis no pacote *biomod2* do R, calibrando os modelos com cinco *runs*, cinco conjuntos de pseudoausências e calibrando 70% dos pontos para treino e 30% para teste. Utilizamos o mesmo número de pseudoausências para todos os algoritmos, sendo igual ao número de pontos de ocorrência espacialmente únicos. Para a avaliação dos modelos, utilizamos a métrica TSS - *True Skill Statistics* (Allouche et al., 2006). Os algoritmos com TSS < 0,4 (Zhang et al., 2015) foram descartados do consenso médio somado seguindo a rotina apresentada por Pires-Oliveira et al. (2019). Os modelos foram produzidos usando as três partições de GCMs mencionadas anteriormente. Finalmente, comparamos os resultados das saídas dos ESMs para cada um dos três tratamentos acima avaliando se o consenso médio somado binário gerado na ESM correspondia à área binária indicada pela VS correspondente. Essa comparação foi feita por meio da função ‘*nicheOverlap*’ do pacote *dismo*, utilizando o índice de



Para o método de agrupamento selecionado (UPGMA, tanto para Distância Euclidiana como para Distância de Manhattan; Tabela 3), obtivemos três grupos estatisticamente consistentes (Figura 2), com todos os grupos contendo mais de dois GCMs (Figura 3).

Tabela 3. Coeficiente cofenético para os cinco diferentes métodos de agrupamento utilizados para avaliar o grau de similaridade dos GCMs da base de dados *CHELSEA*, computados com coeficientes de Distância Euclidiana e Distância de Manhattan. Com ambos os coeficientes de distância o método de agrupamento selecionado foi UPGMA, pois este resultou nos maiores coeficientes cofenéticos.

Distância Euclidiana		Distância de Manhattan	
Método	Coef. Cofenético	Método	Coef. Cofenético
Ligação Simples	0.7035	Ligação Simples	0.7607
Ligação Completa	0.8256	Ligação Completa	0.7657
Ward	0.7525	Ward	0.6117

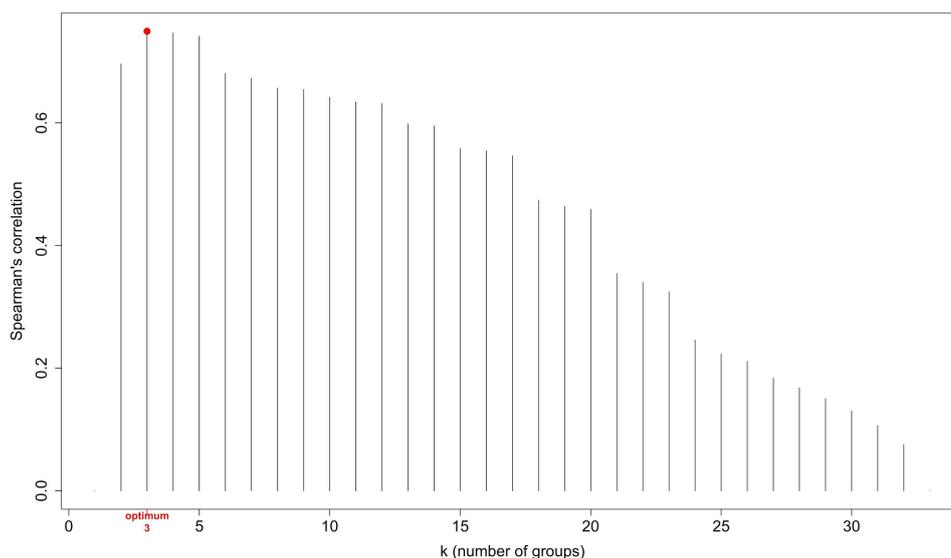


Figura 2. Representação gráfica da correlação entre a matriz de distância original (com distância euclidiana) e a matriz binária gerada a partir do corte do dendrograma em múltiplos níveis. Em nossa análise obtivemos o número ótimo definido como sendo cinco grupos, indicados no gráfico de barras pelo ponto vermelho no topo da barra. Neste ponto obtivemos o maior valor de correlação entre as duas matrizes ( $\rho = 0.73$ ).

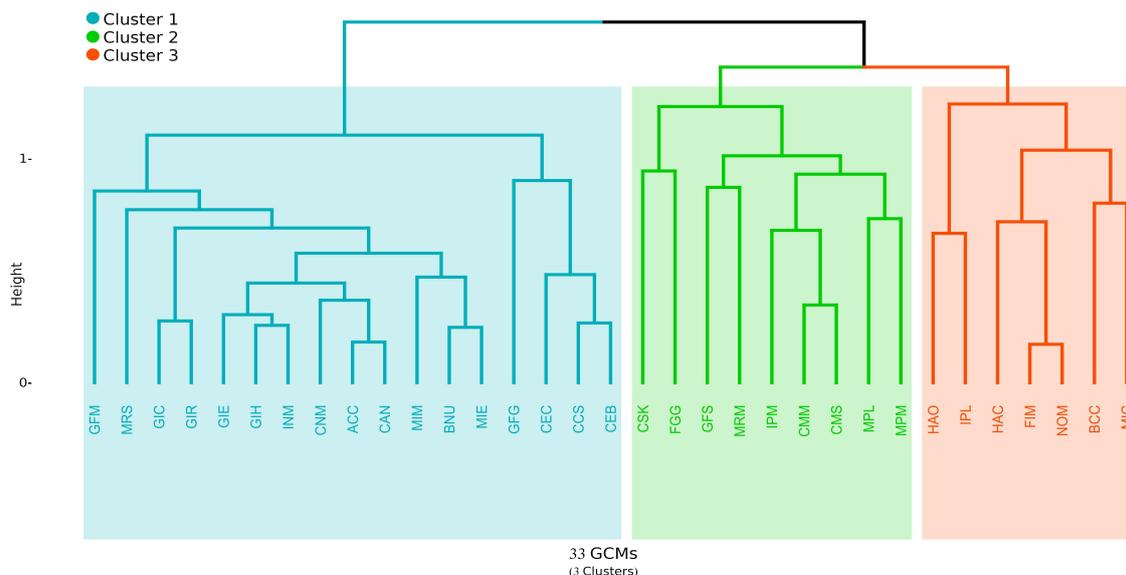


Figura 3. Dendrograma construído com os 33 GCMs da plataforma *CHELSA* com projeções para o intervalo temporal 2061-2080 sob o cenário rcp8.5 de emissão de gases de efeito estufa. Utilizamos os 14 eixos originados da análise *PCoA* como os descritores para produzirmos o agrupamento.

O GCM que conteve a menor correlação com os outros integrantes do grupo, conforme demonstra a Figura 4, foi selecionado em cada grupo. O *g-GCM* totalizou três GCMs (Grupo 1 – CanESM2, Grupo 2 – CSIRO-Mk3-6-0 e Grupo 3 – IPSL-CM5A-LR). Esses foram, segundo as nossas análises, os GCMs mais divergentes entre si dentre os disponíveis na base de dados *CHELSA*. A NPMANOVA revelou significativa diferença entre esses três grupos ( $F = 17.58$ ;  $p < 0.001$ ), com 54% da variância explicada pelo modelo, indicando que os três grupos devem ser representados na ESM.

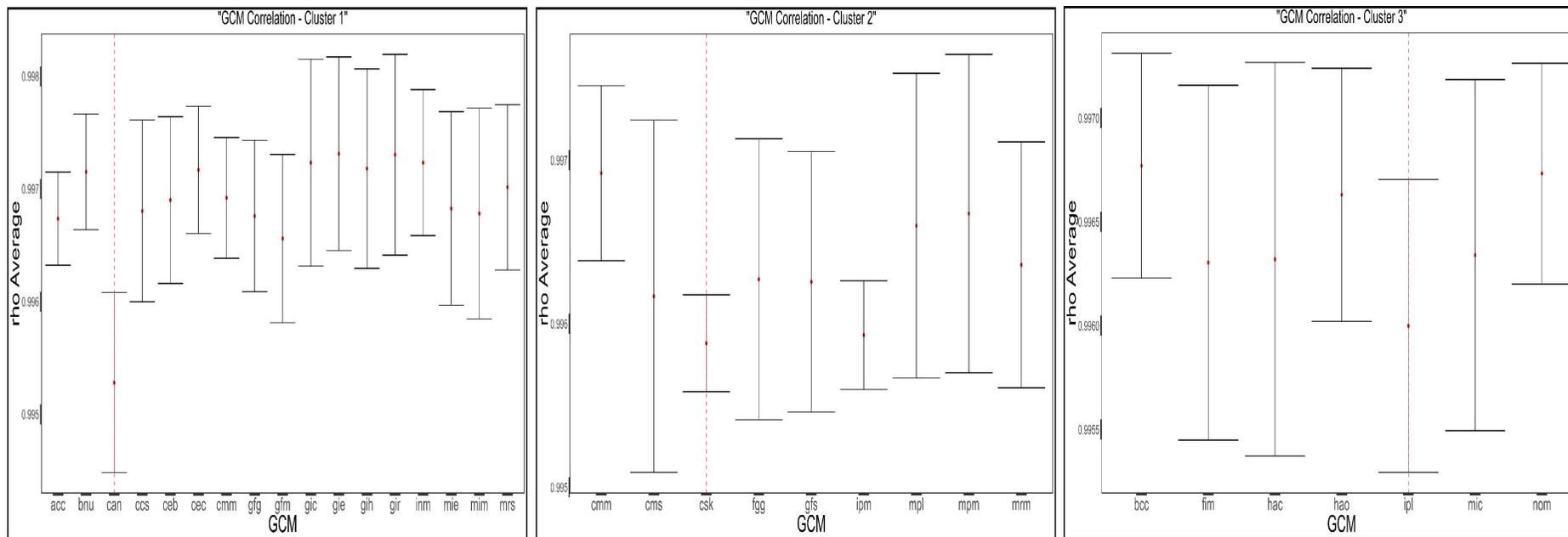


Figura 4. Plot dos valores de  $\rho_m$  (Eixo Y) para cada um dos GCMs do rcp8.5 para o intervalo 2061-2080 (Eixo X). As barras indicam o desvio-padrão da média. A linha pontilhada indica o modelo que obteve o menor  $\rho_m$ . Note a proximidade dos valores de  $\rho_m$  entre todos os GCMs.

A análise de similaridade entre os mapas binários de adequabilidade produzidos pelas VS e por ESM demonstrou diferença entre os desempenhos dos modelos (Figura 5). Apesar da diferença visual (Figura 5 - coluna *Delta*), não houve diferença significativa entre as diferentes partições de GCMs (Kruskal-Wallis;  $p = 0.56$ ). Porém, quando comparamos o desempenho dos ESMs em prever a área de adequabilidade das espécies com diferentes amplitudes de distribuição encontramos maior precisão dos ESMs em prever as áreas de adequabilidade das Espécies 1 e 3, com distribuição intermediária e ampla, respectivamente. A precisão dos ESMs em prever a área de adequabilidade da Espécie 2 (distribuição restrita) foi significativamente inferior em relação ao desempenho dos modelos produzidos para as outras duas espécies (ANOVA One-Way,  $F = 162.6$ ,  $p > 0.001$ ).

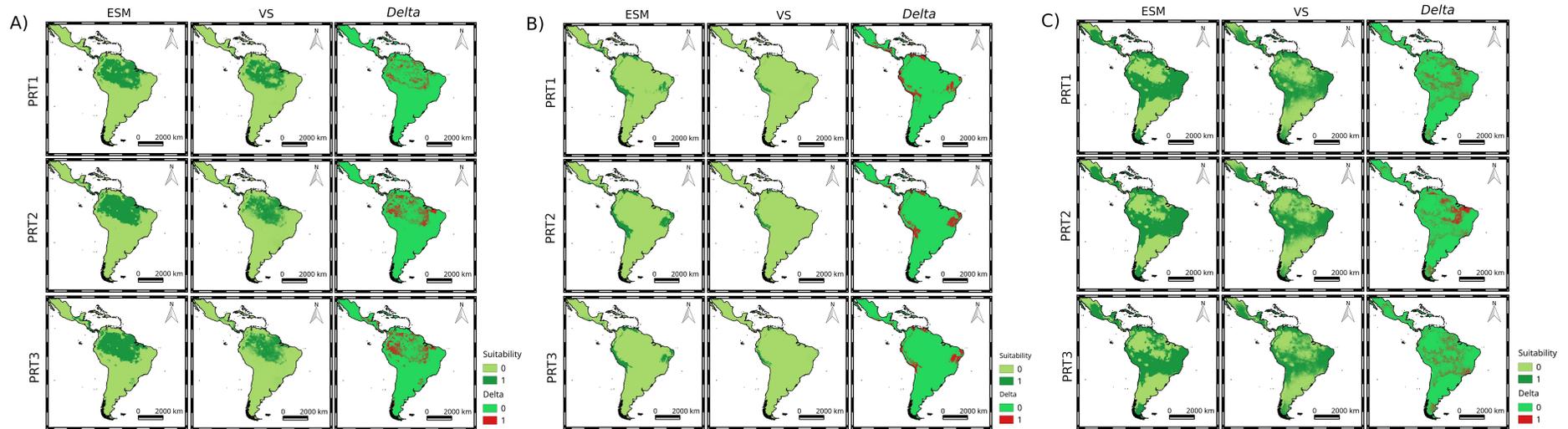


Figura 5. Mapas com as adequabilidades produzidas por ESMs na coluna da esquerda, Espécies Virtuais (VS) na coluna central e a diferença entre os mapas produzidos pelas VS e pelos ESM ( $\Delta = VS - ESM$ ) na coluna da direita. Mapas de adequabilidade gerados com 3 GCMs (PRT1), 33 GCMs da *CHELSA* (PRT2) e com 9 GCMs sugeridos por Sanderson et al. (2015). A) Espécie 1 (Distribuição Intermediária). B) Espécie 2 (Distribuição Restrita). C) Espécie 3 (Distribuição Ampla).

## 4. DISCUSSÃO

Nossa hipótese de que não há necessidade de utilizarmos grande número de GCMs na modelagem de adequabilidade ambiental foi corroborada. Também enfatizamos a importância de se evitar a escolha aleatória de GCMs, pois muitos deles possuem alto grau de correlação, conforme apontado por Knutti et al. (2013). Porém, conforme já mencionamos, os referidos autores utilizaram apenas os parâmetros de médias de precipitação e temperatura anuais. Esses parâmetros são comumente utilizados em estudo para detectar as variações na precisão de previsões de mudanças climáticas realizadas pelos GCMs (Overland et al., 2011, Knutti et al., 2013, Vano et al., 2016), e essas variações são de fundamental importância em modelagens ecológicas e/ou biogeográficas, porém em nossa análise de ordenação, demonstramos que variáveis que indicam a amplitude também são importantes, o que fica evidente pelo comprimento dos autovetores no *plot* da *PCoA*.

Knutti et al. (2013) forneceram importantes sugestões sobre como esses modelos são similares, comparando versões diferentes (CMIP2, CMIP3 e CMIP5) dos GCMs para demonstrar como a versão mais recente é mais eficiente na predição de mudanças climáticas. Entretanto, em nossa análise de agrupamento também encontramos grande similaridade entre os GCMs, com formação de apenas três grupos compostos por 17 GCMs (grupo 1), 9 GCMs (grupo 2) e 7 GCMs (grupo 3), todos com alto grau de correlação entre os constituintes de cada grupo (veja figura 3 e 4), o que demonstra grande similaridade entre os GCMs.

Outros estudos também buscaram otimizar a seleção dos GCMs, mas focaram em avaliar outros processos e aplicações inerentes aos GCMs e não na correlação entre eles (Zubler et al. 2016, Vano et al., 2015), exceção feita a Sanderson et al. (2015). Trabalhos que buscam a seleção de GCMs estão surgindo com diferentes abordagens e metodologias e sempre com enfoque na variação climática binária em suas aplicações (Breach et al. 2016, Zubler et al. 2016). Porém, intrinsecamente todos buscam selecionar subconjuntos de GCMs que sejam representativos do conjunto completo. Destacamos aqui, além de Knutti et al. (2013), os trabalhos realizados por Pierce et al. (2009), Masson e Knutti (2011), Evans et al. (2013), Sanderson et al. (2015) e Vano et al. (2015), que apresentam discussões e metodologias relevantes para a seleção de subconjuntos de GCMs representativos de um conjunto maior de GCMs, considerando a acurácia dos modelos em suas diversificadas abordagens.

Nossa abordagem soma-se a todos esses trabalhos como uma forma de avaliar a capacidade de predição dos GCMs, com enfoque na ESM. Apresentamos uma alternativa prática e confiável na utilização de GCMs para produzir predições sobre adequabilidade ambiental de grupos biológicos. Reconhecemos que as projeções espaciais e temporais da modelagem de adequabilidade produzidas com um consenso médio de GCMs podem retratar de forma mais apropriada as condições ambientais para o intervalo temporal de interesse (Pierce et al., 2009). Porém, argumentamos que, para um conjunto de GCMs ser efetivo, ou seja, para que proporcione a construção de modelos ecológicos robustos, esse conjunto deve carregar a maior porção de variação do conjunto total de dados. Assim, GCMs que apresentam redundância nos padrões de variação dos dados devem ser descartados e apenas os menos correlacionados devem ser utilizados. Dessa forma, cada GCM tem sua porção de contribuição efetiva na produção de mapas de adequabilidade, e isto só se torna plausível se considerarmos a correlação entre eles, e não apenas as variações em micro e macro escala. Destacamos aqui que tanto a seleção de GCMs com baixa correlação entre si quanto de GCMs com maior percentagem de acerto nas previsões de mudanças climáticas acrescentarão poder de predição aos modelos ecológicos.

Nossos resultados sobre o desempenho dos ESMs em diferentes partições de GCMs demonstraram que a sutil diferença encontrada no desempenho dos ESMs produzidos com números diferentes de GCMs não encontra suporte estatístico, possivelmente devido à grande semelhança entre grande parte dos GCMs, conforme descrito por Knutti et al. (2013), Sanderson et al. (2015) e também corroborado por nossos resultados. Isto elimina a necessidade de utilização de um conjunto composto por muitos GCMs na Modelagem de Adequabilidade. Quando verificamos a precisão dos ESMs em detectar as distribuições de espécies com diferentes amplitudes de distribuição (i.e Intermediária - Espécie 1, Restrita - Espécie 2 e Ampla - Espécie 3), os ESMs não apresentaram bom desempenho em prever a distribuição de espécies com amplitudes restritas (Espécie 2). Entretanto, esse valor não se mostra influenciado pelo número de GCMs em cada partição, visto que o desempenho para a Espécie 2 se mostrou baixo nas três partições.

Nossa maior contribuição foi demonstrar o quanto os GCMs são semelhantes e também como estão correlacionados. Demonstramos como o desempenho dos ESMs produzidos sobre diferentes partições com diferentes números de GCMs atuam em prever a adequabilidade ambiental de espécies com três amplitudes de distribuição. Constatamos que

não há perda de desempenho dos ESMs influenciado pelo número de GCMs, o que torna injustificável utilizar um grande número de GCMs em ESM. Além disso, se gerarmos mapas de consenso - algo comum e recomendado em ESMs - com projeções futuras/pretéritas produzidos com grandes números de GCMs, o mapa de consenso terá o viés associado à correlação entre GCMs, pois a maior parte dos GCMs são correlacionados, e desta forma haverá uma maior representatividade dos GCMs correlacionados no consenso da ESM. Entretanto, com um grupo de GCMs não correlacionados, como o que selecionamos em nossa rotina, obtivemos ESMs eficazes, com resultados semelhantes aos obtidos quando utilizamos um número maior número de GCMs, porém com ganho significativo de tempo e redução no conjunto de dados a serem processados, e evitamos que modelos mais correlacionados tenham maior influência no consenso.

Apesar de serem amplamente utilizados em ESM, raramente os autores informam o motivo que os levaram a selecionar um grupo de GCMs, deixando arbitrária a escolha de um grupo que colabore para a construção de modelos mais fidedignos. Apresentamos uma alternativa lógica para a seleção de quais GCMs de projeções pretéritas e/ou futuras deve-se utilizar. Com o método que propomos, torna-se possível selecionar GCMs de forma objetiva e com suporte teórico e matemático. Também encorajamos aqui que sejam justificados os procedimentos utilizados para selecionar os GCMs na aplicação dos modelos ecológicos, para que outros pesquisadores possam também usufruir dessas técnicas e, desta forma, contribuir de forma efetiva para o avanço das pesquisas envolvendo modelagem ecológica.

## **5. CONCLUSÕES**

Este trabalho se apresenta como uma alternativa para a utilização dos GCMs para predições sobre adequabilidade ambiental de espécies. Aqui demonstramos como os GCMs são similares e como selecionar os Modelos de Circulação Geral mais independentes para serem utilizados em ESM. Também demonstramos que, se a seleção dos GCMs menos similares for feita da forma correta, o desempenho dos ESMs não difere de ESMs produzidos com grandes grupos de GCMs, e por isso se torna desnecessário utilizar grandes grupos de GCMs na modelagem de adequabilidade ambiental. Além disso, com grupos menores evitamos problemas de sobre-representatividade de modelos redundantes no consenso da ESM.

## 6. REFERÊNCIAS

- Borcard, D., Gillet, F., Legendre, P., Legendre, P., 2011. Numerical Ecology with R, 1<sup>a</sup>. ed, Springer. Springer US, New York. <https://doi.org/10.1017/CBO9781107415324.004>
- Breach, P.A., Simonovic, S.P., Yang, Z., 2016. Global Climate Model Selection for Analysis of Uncertainty in Climate Change Impact Assessments of Hydro-Climatic Extremes. *Am. J. Clim. Chang.* 05, 502–525. <https://doi.org/10.4236/ajcc.2016.54036>
- Busby, J.R., 1991. BIOCLIM: A bioclimate analysis and prediction system. *Nat. Conserv. cost Eff. Biol. Surv. data Anal.* 6.
- Chunco, A.J., Phimmachak, S., Sivongxay, N., Stuart, B.L., 2013. Predicting environmental suitability for a rare and threatened species (Lao Newt, *Laotriton laoensi*) using validated species distribution models. *PLoS One* 8, 1–13. <https://doi.org/10.1371/journal.pone.0059853>
- Diniz-Filho, J.A.F., Bini, L.M., Loyola, R.D., Rangel, T.F.L.V.B., 2009. Conservation biogeography and climate change in the brazilian cerrado. *Nat. Conserv.* 7, 100–112.
- Evans, J.P., Ji, F., Abramowitz, G., Ekström, M., 2013. Optimally choosing small ensemble members to produce robust climate simulations. *Environ. Res. Lett.* 8. <https://doi.org/10.1088/1748-9326/8/4/044050>
- Faleiro, F. V., Machado, R.B., Loyola, R.D., 2013. Defining spatial conservation priorities in the face of land-use and climate change. *Biol. Conserv.* 158, 248–257. <https://doi.org/10.1016/j.biocon.2012.09.020>
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>
- Franklin, J., Davis, F.W., Ikegami, M., Syphard, A.D., Flint, L.E., Flint, A.L., Hannah, L., 2013. Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Glob. Chang. Biol.* 19, 473–483. <https://doi.org/10.1111/gcb.12051>
- Gastón, A., Garcia-Vinas, J.I., Bravo-Fernandez, A.J., Lopez-Leiva, C., Oliet, J. a, Roig, S., Serrada, R., 2014. Species distribution models applied to plant species selection in forest restoration: are model predictions comparable to expert opinion? *New For.* 45, 641–653. <https://doi.org/DOI 10.1007/s11056-014-9427-7>

- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. *Habitat Suitability and Distribution Models*. Cambridge University Press, Cambridge.  
<https://doi.org/10.1017/9781139028271>
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16, 1424–1435. <https://doi.org/10.1111/ele.12189>
- Hijmans, R.J., 2018. raster: Geographic Data Analysis and Modeling.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.  
<https://doi.org/10.1002/joc.1276>
- IPCC, 2013. What is a GCM? [WWW Document]. URL  
[http://www.ipcc-data.org/guidelines/pages/gcm\\_guide.html](http://www.ipcc-data.org/guidelines/pages/gcm_guide.html) (accessed 11.23.18).
- IPCC, 2007. *Climate change: Synthesis report*. IPCC, Geneva.
- Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4, 1–20. <https://doi.org/10.1038/sdata.2017.122>
- Knutti, R., Masson, D., Gettelman, A., 2013. Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.* 40, 1194–1199.  
<https://doi.org/10.1002/grl.50256>
- Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. virtualspecies, an R package to generate virtual species distributions. *Ecography (Cop.)*. 39, 599–607.  
<https://doi.org/10.1111/ecog.01388>
- Masson, D., Knutti, R., 2011. Climate model genealogy. *Geophys. Res. Lett.* 38, n/a-n/a.  
<https://doi.org/10.1029/2011GL046864>
- Meadows, M.E., 1999. Biogeography: changing places, changing times. *Prog. Phys. Geogr.* 23, 257–270. <https://doi.org/10.1177/030913339902300206>
- Overland, J.E., Wang, M., Bond, N.A., Walsh, J.E., Kattsov, V.M., Chapman, W.L., 2011. Considerations in the Selection of Global Climate Models for Regional Climate

- Projections: The Arctic as a Case Study\*. *J. Clim.* 24, 1583–1597.  
<https://doi.org/10.1175/2010JCLI3462.1>
- Paradis, E., Schliep, K., 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. *Bioinformatics* xx, xxx–xxx.
- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* 12, 361–371. <https://doi.org/10.1046/j.1466-822X.2003.00042.x>
- Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *J. Biogeogr.* 34, 102–117.  
<https://doi.org/10.1111/j.1365-2699.2006.01594.x>
- Peterson, A.T., Soberon, J., 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *Nat. Conserv.* 10, 102–107.  
<https://doi.org/10.4322/natcon.2012.019>
- Pierce, D.W., Barnett, T.P., Santer, B.D., Gleckler, P.J., 2009. Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci.* 106, 8441–8446.  
<https://doi.org/10.1073/pnas.0900094106>
- Pires-Oliveira, J.C., José-Silva, L., Rocha, D.S.B., Eisenlohr, P.V., 2019. Concepts and methods in environmental suitability modeling, an important tool for forest conservation, in: Eisenlohr, P. V (Ed.), *Forest Conservation: Methods, Management and Challenges*. Nova Publishers Ltda, New York, pp. 269–290.
- R Core Team, 2018. R: A language and environment for statistical computing.
- Rezende, V.L., Eisenlohr, P. V., Gasper, A.L. de, Vibrans, A.C., Oliveira-Filho, A.T. de, 2014. Toward a better understanding of the subtropical Atlantic Forest in the state of Santa Catarina, Brazil: tree sampling accuracy, species richness and frequency of rare species. *Acta Bot. Brasilica* 28, 382–391. <https://doi.org/10.1590/0102-33062014abb3146>
- Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., Rafaj, P., 2011. RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Clim. Change* 109, 33–57. <https://doi.org/10.1007/s10584-011-0149-y>
- Sanderson, B.M., Knutti, R., Caldwell, P., 2015. A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble. *J. Clim.* 28, 5171–5194.  
<https://doi.org/10.1175/JCLI-D-14-00362.1>

- Siqueira, M.F. de, Durigan, G., de Marco Júnior, P., Peterson, A.T., 2009. Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *J. Nat. Conserv.* 17, 25–32. <https://doi.org/10.1016/j.jnc.2008.11.001>
- Souza, R.A. de, de Marco Júnior, P., 2014. The use of species distribution models to predict the spatial distribution of deforestation in the western Brazilian Amazon. *Ecol. Modell.* 291, 250–259. <https://doi.org/10.1016/j.ecolmodel.2014.07.007>
- Vano, J.A., Kim, J.B., Rupp, D.E., Mote, P.W., 2015. Selecting climate change scenarios using impact-relevant sensitivities. *Geophys. Res. Lett.* 42, 5516–5525. <https://doi.org/10.1002/2015GL063208>
- Vogler, R.E., Beltramino, A.A., Sede, M.M., Gregoric, D.E.G., Núñez, V., Rumi, A., Vogler, R.E., Beltramino, A.A., Sede, M.M., Gregoric, D.E.G., Núñez, V., Rumi, A., 2013. The giant African snail, *Achatina fulica* (Gastropoda: Achatinidae): Using bioclimatic models to identify South American areas susceptible to invasion. *Am. Malacol. Bull.* 31, 39–50. <https://doi.org/10.4003/006.031.0115>
- Waltari, E., Hijmans, R.J., Peterson, A.T., Nyari, A.S., Perkins, S.L., Guralnick, R.P., 2007. Locating pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS One* 2, e563. <https://doi.org/10.1371/journal.pone.0000563>
- Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* (N. Y.). 62, 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>
- Wilby, R., Dawson, C., Barrow, E., 2002. sdsms — a decision support tool for the assessment of regional climate change impacts. *Environ. Model. Softw.* 17, 145–157. [https://doi.org/10.1016/S1364-8152\(01\)00060-3](https://doi.org/10.1016/S1364-8152(01)00060-3)
- Wilby, R.L., Hay, L.E., Leavesley, G.H., 1999. A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. *J. Hydrol.* 225, 67–91. [https://doi.org/10.1016/S0022-1694\(99\)00136-5](https://doi.org/10.1016/S0022-1694(99)00136-5)
- Zhang, J., Li, S., 2017. A Review of Machine Learning Based Species' Distribution Modelling, in: 2017 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII). IEEE, pp. 199–206. <https://doi.org/10.1109/ICIICII.2017.76>
- Zubler, E.M., Fischer, A.M., Fröb, F., Liniger, M.A., 2016. Climate change signals of CMIP5 general circulation models over the Alps - impact of model selection. *Int. J. Climatol.* 36, 3088–3104. <https://doi.org/10.1002/joc.4538>

Zwiener, V.P., Lira-Noriega, A., Grady, C.J., Padial, A.A., Vitule, J.R.S., 2018. Climate change as a driver of biotic homogenization of woody plants in the Atlantic Forest. *Glob. Ecol. Biogeogr.* 27, 298–309. <https://doi.org/10.1111/geb.12695>