
ECOLOGIA

ABNER CASTRO DE CARVALHO

**INFLUÊNCIA DA SELEÇÃO DE VARIÁVEIS
AMBIENTAIS SOBRE A MODELAGEM DE
DISTRIBUIÇÃO DE ESPÉCIES:
UM ESTUDO DE CASO PARA
ESCHWEILERA CORIACEA (DC.) S.A MORI
NA REGIÃO AMAZÔNICA**

ABNER CASTRO DE CARVALHO

INFLUÊNCIA DA SELEÇÃO DE VARIÁVEIS AMBIENTAIS SOBRE A
MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIES: UM ESTUDO DE
CASO PARA *ESCHWEILERA CORIACEA* (DC.) S.A MORI NA REGIÃO
AMAZÔNICA

Orientador: Thiago Sanna Freire Silva

Co-orientador: Bruno Garcia Luize

Trabalho de Conclusão de Curso apresentado
ao Instituto de Biociências da Universidade
Estadual Paulista “Júlio de Mesquita Filho” -
Câmpus de Rio Claro, para obtenção do grau
Ecólogo.

Rio Claro
2017

574.9 Carvalho, Abner Castro de
C331i Influência da seleção de variáveis ambientais sobre a
 modelagem de distribuição de espécies: um estudo de caso
 para *Eschweilera coriacea* (DC.) S.A Mori na região
 amazônica / Abner Castro de Carvalho. - Rio Claro, 2017
 34 f. : il., figs., tabs.

Trabalho de conclusão de curso (Ecologia) - Universidade
Estadual Paulista, Instituto de Biociências de Rio Claro
Orientador: Thiago Sanna Freire Silva
Coorientador: Bruno Garcia Luize

1. Biogeografia. 2. Modelagem de distribuição de espécies
(MDE). 3. MaxEnt. 4. Geotecnologias. 5. Amazônia. I. Título.

Dedico este trabalho a todos que acreditaram em meu potencial e que colaboraram
para que eu o realizasse.

AGRADECIMENTOS

Agradeço a todos os que contribuíram direta ou indiretamente para minha formação. Agradeço à minha família que sempre esteve disposta a ajudar-me e que me manteve em condições de estudar em outra cidade. Agradeço aos amigos que me motivaram e sempre renovaram minhas energias para continuar focado nos estudos. Agradeço também a todos os membros do *Ecosystem Dynamics Observatory* (EcoDyn), laboratório no qual desenvolvi meu Trabalho de Conclusão de Curso, que contribuíram para ampliar meu conhecimento. Em especial, agradeço ao meu orientador Thiago Sanna Freire Silva por ajudar-me a trilhar no caminho acadêmico de maneira eficiente. Também gostaria de lembrar o auxílio do meu co-orientador Bruno Garcia Luize no delineamento e execução deste projeto. Agradeço ao Carlos Leandro Cordeiro que sempre esteve disposto a ajudar e contribuir com seus conhecimentos para meu avanço técnico. Agradeço à bolsa PIBITI – 116121/2016-5 pelo financiamento do projeto.

“Dou importância às coisas desimportantes
E aos seres desimportantes
Prezo insetos mais que aviões.
Prezo a velocidade
das tartarugas mais do que as dos mísseis.
Tenho em mim esse atraso de nascença.”
(BARROS, 2010, p. 47).

RESUMO

O avanço da tecnologia tem colaborado com os estudos ecológicos. Em especial, vemos o surgimento de geotecnologias que contribuem com a geração de dados ambientais e auxiliam estudos desta grande área. Ao mesmo tempo, cada vez mais surgem registros de ocorrência de espécies que ajudam a modelar sua área e limites de distribuição. Neste trabalho, desenvolvemos seis modelos de distribuição potencial da espécie *Eschweilera coriacea* para a região amazônica, utilizando dados de temperatura e precipitação de fontes distintas, uma vez que estas variáveis são reportadas como as mais comumente usadas na literatura. Um dos modelos se baseou também nas variáveis ambientais de elevação, evapotranspiração e biomassa, além da temperatura e precipitação, com o objetivo de analisar se a inclusão de outras variáveis importantes para as plantas mudam a predição da distribuição da espécie. Os modelos gerados mostraram que as mesmas variáveis, quando obtidas através de fontes de dados geoespaciais distintas, podem prever a distribuição da espécie de formas diferentes. A área não predita para a distribuição da espécie em nenhum dos modelos corresponde a 29,19% do total. A área predita por apenas um modelo corresponde a 20,59%. A área predita por dois modelos foi igual a 8,31%. A área predita por três modelos foi igual a 6%. A área predita por quatro modelos foi igual a 6,60%. A área predita por cinco modelos correspondeu a 13,17% e a área predita por todos os modelos foi de apenas 9,84%. Sendo assim, menos de 10% do total de áreas preditas foram comuns em todas as modelagens. Todos os modelos tiveram $AUC > 0.5$, mas só o modelo que utilizou diferentes variáveis ambientais além de precipitação e temperatura teve um $AUC > 0.7$. Assim, demonstramos que a seleção das variáveis ambientais a serem utilizadas na elaboração de modelos de distribuição potencial de espécies é um aspecto importante e que pode influenciar significativamente o resultado da predição da distribuição das espécies estudadas, além da confiabilidade estatística dos modelos.

Palavras-chave: Modelagem de distribuição de espécies (MDE). MaxEnt. Geotecnologias. Amazônia

ABSTRACT

The advancement of technology has collaborated with ecological studies. In particular, we can see the emergence of geotechnologies that present environmental data and benefits studies of this broad area. At the same time, there are more and more records of occurrences of species that help to shape the distribution of species. In this work, we conducted six distribution modeling of the *Eschweilera coriacea* species for the Amazon region, using only temperature and precipitation, that are reported as the most common in literature, but coming from five distinct sources of information. One of the models was also based on the environmental variables elevation, evapotranspiration and biomass, in addition to temperature and precipitation, in order to analyze if the inclusion of other important variables for the plants changes the prediction of the species distribution. The generated models showed that temperature and precipitation data coming from different sources may lead to distinct predictions of the species distribution. The models with coarse resolution variables were the ones that most varied in the prediction of the species distribution. The area not predicted for the distribution of the species in any of the models corresponds to 29.19% of the total. The area predicted by only one model corresponds to 20.59%. The area predicted by two models was equal to 8.31%. The area predicted by three models was equal to 6%. The area predicted by four models was equal to 6.60%. The area predicted by five models corresponded to 13.17% and the area predicted by all models was only 9.84%. Thus, less than 10% of the total areas predicted were common in all models. All models had AUC > 0.5, but only one model using different environmental variables besides precipitation and temperature had an AUC > 0.7. Thus, we have shown that the selection of environmental variables to be used in the elaboration of potential species distribution models is an important aspect that can significantly influence the prediction of the distribution of the studied species, as well as the statistical reliability of the models.

Keywords: Species Distribution Model (SDM). MaxEnt. Geotechnology. Amazon

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVOS	12
3	MATERIAIS E MÉTODOS	13
3.1	Área de estudo	13
3.2	Escolha da espécie	14
3.3	Dados de ocorrência	14
3.4	Dados ambientais	15
3.5	Modelagem de distribuição da espécie	18
4	RESULTADOS E DISCUSSÃO	21
5	CONCLUSÃO	30
	REFERÊNCIAS	31

1 INTRODUÇÃO

A ecologia pode ser definida como o estudo da distribuição e abundância das espécies (BEGON et al., 2009). O mesmo objetivo é compartilhado por outras disciplinas como, por exemplo, a macroecologia, a filogeografia e a biogeografia. A biogeografia pode ser definida como uma ciência que se preocupa em documentar e compreender os padrões e processos que determinam a distribuição presente e pretérita dos seres vivos (BROWN; LOMOLINO, 2006). A compreensão da distribuição e abundância das espécies é cada vez mais relevante, visto nossa urgência em prever os efeitos que mudanças ambientais e pressões antrópicas poderão ter sobre a biodiversidade (HORTAL et al., 2015). Neste sentido, temos obtido grandes avanços tecnológicos, os quais vem auxiliando nossa compreensão sobre a distribuição da vida na Terra.

O sensoriamento remoto é um destes avanços, e pode ser compreendido como a ciência da obtenção de informação sobre um objeto sem contato físico direto, ou como uma tecnologia que pode ser usada para medir e monitorar aspectos biofísicos e de atividades humanas (JENSEN, 2009). Os dados de sensoriamento remoto se mostram importantes na classificação dos ecossistemas e nas descrições das condições ambientais, e são essenciais para estudar a complexidade ecológica nos processos do Sistema Terrestre, contribuindo também para o entendimento da diversidade biológica (MATSON, 1991).

Nas últimas três décadas, as tecnologias de sensoriamento remoto evoluíram, operando e abrangendo variadas escalas espaciais, espectrais e temporais. Somado a isso, o fácil acesso a dados históricos e a redução de custo dos dados de satélites colaboram para a inclusão destas informações nos estudos ambientais (ROGAN, 2004). Atualmente, os sensores remotos coletam e analisam dados a partir de plataformas terrestres, atmosféricas e em órbita terrestre, conectadas a sistemas de navegação por satélites (*Global Navigation Satellite Systems*, GNSS), oferecendo suporte para o desenvolvimento de modelos espaciais quantitativos e Sistemas de Informação Geográfica (SIGs) (FRANKLIN, 2001). Isso faz do sensoriamento remoto uma ferramenta muito valiosa para captar informações sobre a superfície da Terra (ROGAN, 2004).

Assim, temos cada vez mais disponibilidade de informações ambientais sobre a superfície do planeta. Estas informações são monitoradas e compartilhadas em rede e atingem ordem de Petabytes (i.e. Milhões de Gigabytes), estando disponíveis em resoluções espaciais cada vez mais refinadas e abrangendo uma ampla gama de aspectos ambientais (PETERSON et al., 2011). Concomitantemente, presenciamos um crescente compartilhamento de dados sobre a ocorrência e as características das muitas espécies que habitam nosso planeta (e.g. SpeciesLink <www.splink.cria.org.br/>; *Global Biodiversity Information Facility* (GBIF) <www.gbif.org/>; TRY <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627314/>>), chegando à ordem de bilhões de registros, os quais estão sendo digitalizados e compartilhados por museus de todo o mundo. A ecologia entra assim na era da ciência de grandes dados, e se fixa como uma ciência multidisciplinar e colaborativa.

O ferramental disponível para compreender a distribuição das espécies e variações em suas abundâncias vem acompanhando os avanços na quantidade de informações de qualidade e na sua acessibilidade. Desde o final do século passado, a capacidade e acessibilidade computacional tem crescido exponencialmente, e o desenvolvimento de softwares específicos propicia a realização de estimativas mais robustas sobre a distribuição de espécies (PETERSON et al., 2011). Vemos, então, o surgimento de métodos para o desenvolvimento de modelos estatísticos e computacionais para predizer a distribuição de espécies, capazes de relacionar informações georreferenciadas sobre as condições ambientais e sobre a ocorrência das espécies, por meio de algoritmos baseados no referencial teórico construído por ecólogos ao longo do século XX. Estes métodos são conhecidos como Modelos de Nicho Ecológico (*Ecological Niche Models* – ENMs) ou Modelos de Distribuição de Espécies (*Species Distribution Models* – SDMs) (PETERSON et al., 2011; FRANKLIN, 2009). Neste âmbito, as geotecnologias, em especial o uso do sensoriamento remoto e dos sistemas de informações geográficas (SIGs), têm contribuído imensamente para a aquisição, armazenamento, análise e visualização destas informações espaciais e ambientais.

Os SDMs são um dos métodos mais populares atualmente usados para estudar impactos sobre a biodiversidade (GUISAN et al., 2013). Para alcançar o melhor resultado possível, os SDMs geralmente são calibrados, parcial ou totalmente, a partir de variáveis climáticas, dada a influência direta do clima sobre a distribuição de espécies (DEBLAUWE, 2016; GUISAN; ZIMMERMANN, 2000).

Uma das maiores limitações de todos os SDMs é a impossibilidade de se obterem dados de ausência. Sendo assim, é preciso trabalhar só com dados de presença, que requer métodos específicos. Um destes métodos é o modelo de máxima entropia (MaxEnt). O MaxEnt é um algoritmo que modela a distribuição de espécies com base em registros de ocorrências, considerando somente dados confirmados da presença da espécie (ELITH et al., 2010). O MaxEnt integra a densidade de pontos de ocorrência e localização geográfica de espécies para modelar as distribuições, fornecendo uma probabilidade para avaliar a contribuição das camadas de dados geotecnológicos (PHILLIPS et al., 2005).

O modelo gera previsões das áreas adequadas para distribuição geográfica potencial da espécie (PHILLIPS et al., 2015). Desta forma, prevê-se que uma espécie com ocorrência observada junto a um conjunto de variáveis ambientais venha ocorrer em áreas semelhantes, mesmo a espécie não sendo registrada nestes locais (STEEGE et al., 2013). Disponível a partir de 2004, o algoritmo MaxEnt tem sido muito utilizado nos estudos de SDM, com o objetivo de mapear e prever a distribuição potencial de espécies, com aplicações ecológicas, evolutivas e de conservação (ELITH et al., 2010). Comparado com outros modelos existentes, o MaxEnt possui uma série de características que o tornam muito útil para trabalhos de modelagem de distribuição de espécies (ELITH et al., 2006; PHILLIPS et al., 2005).

Pouca atenção tem sido dada à origem e qualidade dos dados climáticos utilizados (DEBLAUWE, 2016; FERNANDEZ et al., 2013). A seleção correta das variáveis ambientais a serem utilizadas e o conjunto mais adequado é essencial (DORMANN, 2007). Modelos de distribuição são ferramentas muito poderosas para estudar e prever a distribuição das espécies, mas a qualidade e precisão da modelagem dependem fortemente dos dados utilizados como entrada, assim como para a maioria dos modelos estatísticos e computacionais (MOD et al., 2016). Atualmente, a grande maioria dos estudos em modelos de distribuição de plantas incluem variáveis de temperatura e variáveis relacionadas à precipitação (cerca de 88,5% dos estudos), mas mais da metade destes se baseia unicamente em descritores de variações de temperatura e precipitação (ou a estas duas categorias e mais uma variável adicional), potencialmente negligenciando outros aspectos ecofisiológicos importantes (MOD et al., 2016). Adicionalmente, alguns estudos têm considerado o balanço hídrico (precipitação menos evapotranspiração), uma medida

mais precisa de água disponível no caso de organismos vegetais do que simplesmente o dado de precipitação (MOD et al., 2016).

A fonte de dados climáticos mais comumente utilizada é o conjunto de dados globais do *dataset* WorldClim (HIJMANS et al., 2005), que possui acesso livre (DEBLAUWE, 2016). Os dados WorldClim são calculados com base na interpolação de observações de estações climáticas próximas, ponderado pelo relevo. Em áreas tropicais, a interpolação recobre grandes áreas, visto que a densidade de estações meteorológicas nesta região é baixa, podendo levar a previsões inconsistentes de padrões climáticos. (DEBLAUWE, 2016; WASHINGTON et al., 2013; NEW et al., 2002; FERNANDEZ et al., 2013.)

Para as plantas, sete fatores ambientais são tidos como essenciais: temperatura, água, nutrientes, luz, interações bióticas e concentração atmosférica de CO₂ (KÖRNER, 2014; GUISAN; ZIMMERMANN, 2000; AUSTIN; NIEL, 2011). Por isso, mais esforço deve ser feito para a seleção e avaliação das variáveis preditoras utilizadas para construir os SDMs. Isto requer um estudo mais profundo para escolher os melhores preditores ambientais necessários para atingir o objetivo do estudo, assim como desenvolver novos preditores que se façam necessários (MOD et al., 2016). Assim, uma boa distribuição, realista e precisa, só pode ser alcançada quando todos os fatores importantes que definem o nicho da espécie são contabilizados e incluídos no estudo (PEARSON; DAWSON, 2003; MCGILL, 2010).

2 OBJETIVOS

Os objetivos deste trabalho foram:

- a) Modelar a distribuição da espécie de árvore amazônica *Eschweilera coriacea* (Matamatá, Matamatá Branco), utilizando dados de temperatura e precipitação obtidos de fontes diferentes, e gerados através de algoritmos distintos, e analisar o efeito da seleção destas variáveis sobre a predição da distribuição desta espécie;
- b) Modelar a distribuição de *E. coriacea* utilizando variáveis ambientais adicionais e comparar a qualidade dos modelos com aqueles baseados apenas em temperatura e precipitação.

O interesse foi responder às seguintes questões:

- a) A utilização de variáveis similares de temperatura e precipitação, obtidas através de fontes e algoritmos diferentes afetam o resultado da predição da distribuição de *E. coriacea*?
- b) A inclusão de outras variáveis menos utilizadas na literatura, e muitas vezes omitidas, muda e/ou melhora os resultados a modelagem de distribuição da espécie selecionada?

3 MATERIAS E MÉTODOS

3.1 Área de estudo

Este estudo abrangeu as áreas úmidas florestais da Amazônia sensu-latíssimo, que compreende a bacia hidrográfica dos rios Amazonas e Tocantins (Amazônia sensu-stricto) e quatro sub-regiões periféricas: Andes, Guianas, Gurupi e Planalto. Essa nova delimitação considera critérios hidrográficos da região amazônica, critérios ecológicos que exercem influência direta ou indireta na região das terras baixas e planícies amazônicas, critérios biogeográficos e históricos da floresta amazônica (Figura 1) (EVA et al., 2005).

A maior parte da região é coberta por florestas tropicais, mas também inclui vegetação não florestal, como as savanas amazônicas, campinas, campinaranas e florestas inundáveis (e.g várzeas e igapós) (EVA et al., 2005). Em termos de cobertura, a região possui aproximadamente 80% de florestas e inclui 97% das florestas tropicais do continente e 83% das florestas alagadas (EVA et al., 2005).

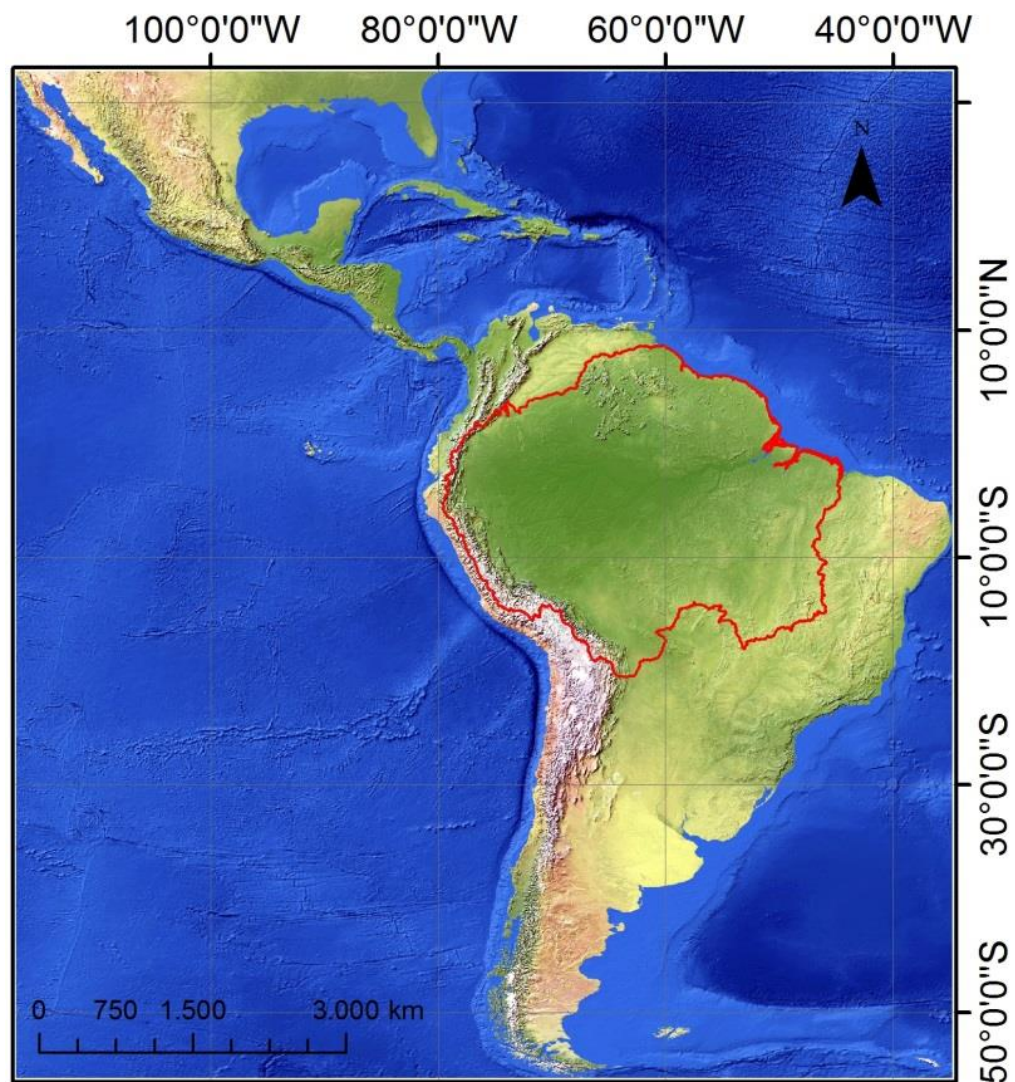


Figura 1. Área de estudo em vermelho: Amazônia sensu-latíssimo, proposto por Eva et al. (2005). (Fonte: Elaborado pelo autor).

3.2 Escolha da espécie

A espécie escolhida, *Eschweilera coriacea*, conhecida popularmente como Matamatá, é uma planta pertencente à família Lecythidaceae, tida como uma das espécies hiperdominantes da Amazônia (STEEGE et al., 2013). O que motivou a escolha de *E. coriacea* foi sua ampla abundância e distribuição na região amazônica, estando distribuída de forma pan-amazônica. A espécie está em terceiro lugar na classificação de árvores de maior abundância na Amazônia (TER STEEGE et al., 2013), e na Amazônia central, a espécie chega a uma densidade de 1.519 árvores de um total de 7.791 amostradas em um terreno de 100 ha (MORI et al., 2001).

3.3 Dados de ocorrência

Os dados de ocorrência de *E. coriacea* foram obtidos a partir da fonte: GBIF (<<http://www.gbif.org/>>). O GBIF é um banco de dados aberto e destinado ao fornecimento de dados sobre a ocorrência de todos os tipos de vida na Terra. Foram encontrados 1.840 registros da espécie na região amazônica, que depois de serem filtrados resultaram em um total de 347 pontos de ocorrências (Figura 2). Os dados de ocorrência da espécie foram filtrados retirando os pontos de mesma localidade, pontos de ocorrência inexistente e/ou fora da área de estudo (e.g pontos que caíam no oceano), pontos de sede de municípios e pontos mais antigos que 1950, pois os dados consideram os anos acima de 1950.

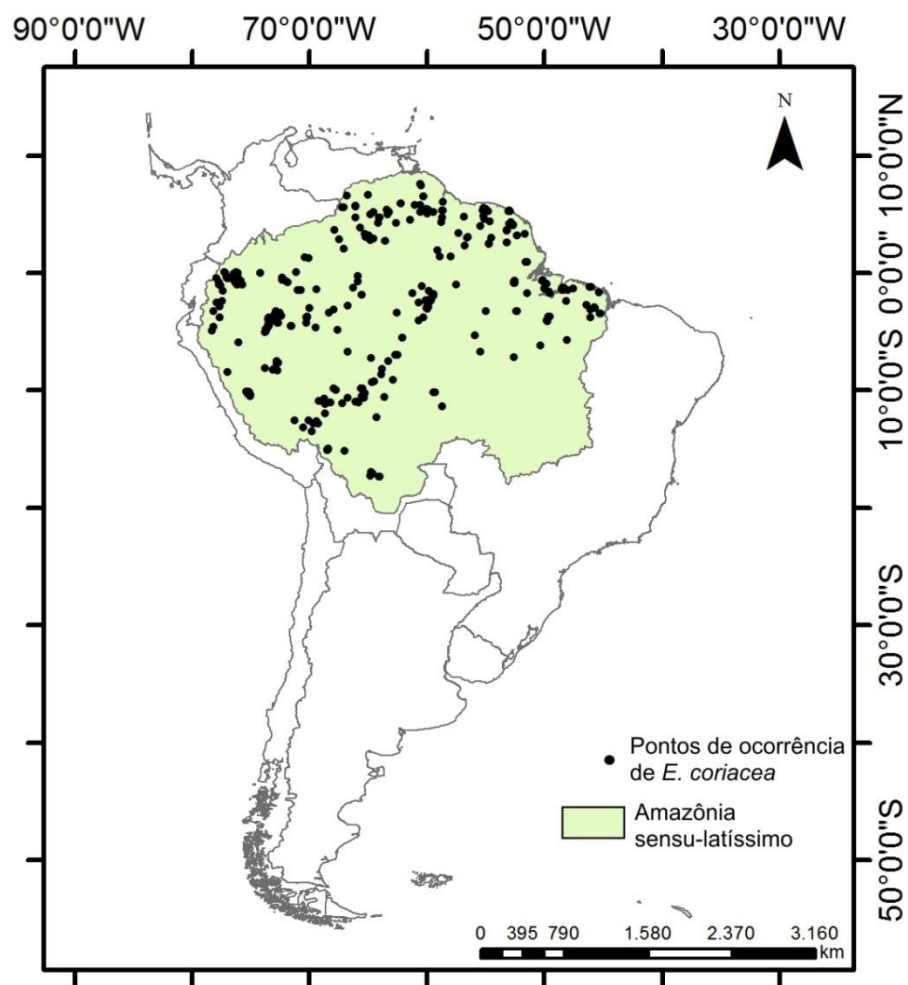


Figura 2. Pontos de ocorrência de *E. coriacea* distribuídas pela Amazônia sensu-latíssimo proposta por Eva et al. (2005). (Fonte: Elaborado pelo autor).

3.4 Dados ambientais

Os dados climáticos utilizados foram de temperatura média anual (bio 1) e de precipitação anual (bio 12), que das 19 variáveis bioclimáticas descritas no *WorldClim*, são as mais incluídas nas modelagens de distribuição de espécies (Tabela 1) (MOD et al., 2016). Foram utilizadas as seguintes fontes de dados climáticos:

- a) WorldClim - Global Climate Data, (Hijmans et al. 2005 <<http://www.worldclim.org/>>). Foram utilizadas as duas versões WorldClim. As duas versões fornecem as 19 variáveis bioclimáticas obtidas através de interpolação de dados de temperatura e precipitação de estações climáticas distribuídas pelo globo. A primeira versão considera dados de 1960 a 1990, a segunda versão dados de 1970 a 2000. Outra diferença da segunda versão é a inclusão de algumas variáveis nos dados (radiação solar, pressão de vapor e velocidade do vento). Ambos dados foram utilizados com resolução de 1 km;
- b) CHELSA - “Climatologies at high resolution for the earth’s land surfaces areas” - (<<http://www.chelsa-climate.org/>>). Fornecem as 19 variáveis bioclimáticas de 1979 a 2013 com resolução espacial de 1 km. O dado é criado com dados de temperatura e precipitação baseado numa redução quantitativa quase mecanicista do modelo de circulação global *ERA-interim* (KARGER et al., 2016);
- c) ecoClimate (<<http://www.ecoclimate.org/>>). Disponibilizam as 19 variáveis bioclimáticas simuladas para o período de 1950 a 1999 a partir dos projetos CMIP5 e PMIP3 com 50 km de resolução, como Modelo de Circulação Global Atmosfera-Oceano (AOGCM) foi utilizado o CCSM;
- d) MODIS - “Moderate Resolution Imaging Spectroradiometer” - Retirado do produto MOD11 – “Land Surface Temperature and Emissivity” (<https://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD_NUMBER=11>) e CHIRPS - “Climate Hazards Group InfraRed Precipitation with Station data” (<<http://chg.ucsb.edu/data/chirps/>>). Foram incluídos através da base de dados com as 19 variáveis bioclimáticas de 5,5 km de

resolução especial, disponibilizadas por Deblauwe et al. (2016) (<https://vdeblauwe.wordpress.com/download/>).

Para avaliação de outras variáveis ambientais foram incluídos em um dos modelos dados de biomassa extraídos do trabalho de Avitabile et al. (2016), onde os autores integraram dois mapas de biomassa existente (BACCINI et al., 2012) e disponibilizam os dados, que levam em conta observações de campo, sensoriamento remoto e dados de estudos de locais específicos e/ou estudos de escala global, focado na região tropical (AVITABILE et al., 2016).

Dados de elevação do terreno foram obtidos a partir de dados do *Shuttle Radar Topography Mission* (SRTM) na base do *United States Geological Survey* (USGS <<http://earthexplorer.usgs.gov/>>), onde estão disponíveis com resolução de 90 metros e 30 metros. Para a modelagem, foi utilizado o dado de altitude do SRTM presente nos dados WorldClim, onde já se encontra disponível com resolução de 1 km.

Foi utilizado o Índice de Vegetação da Diferença Normalizada (*Normalized Difference Vegetation Index – NDVI*). Este índice é retirado do produto MODIS-MOD13A3 e fornecem dados espaciais da vegetação, mais especificamente da atividade clorofiliana. Foi feito a média do produto, que contém dados disponíveis de 2000 a 2016, com resolução de 1 km (<<https://modis.gsfc.nasa.gov/data/dataproduct/mod13.php/>>).

Outra variável utilizada foi evapotranspiração, retirada do produto MODIS-MOD16, onde são encontrados quatro diferentes dados: evapotranspiração (ET), potencia de evapotranspiração (PET), fluxo de calor latente (LE) e potencial de fluxo de calor latente (PLE). Para o modelo foi usado o dado de evapotranspiração (ET) (<https://modis.gsfc.nasa.gov/data/dataproduct/mod16.php/>).

Tabela 1. As 19 variáveis bioclimáticas proposta pelo WorldClim (Hijmans et al., 2005).

Variáveis de temperatura	Variáveis de precipitação
bio 1 - Média anual de temperatura	bio 12 - Precipitação anual
bio 2 - Variação da média diurna	bio 13 - Precipitação do mês mais chuvoso
bio 3 - Isotermalidade	bio 14 - Precipitação do mês mais seco
bio 4 - Sazonalidade da temperatura	bio 15 - Sazonalidade da precipitação
bio 5 - Temperatura máxima do mês mais quente	bio 16 - Precipitação do trimestre mais chuvoso
bio 6 - Temperatura mínima do mês mais frio	bio 17 - Precipitação do trimestre mais seco
bio 7 - Variação da temperatura anual	bio 18 - Precipitação do trimestre mais quente
bio 8 - Temperatura média do trimestre mais chuvoso	bio 19 - Precipitação do trimestre mais frio
bio 9 - Temperatura média do trimestre mais seco	
bio 10 - Temperatura média do trimestre mais quente	
bio 11 - Temperatura média do trimestre mais frio	

3.5 Modelagem de distribuição da espécie

Foram usados 347 pontos de ocorrência nos modelos. Os dados finais de ocorrência foram armazenados em uma tabela, formato “csv”, com três colunas. São elas: coluna da espécie, longitude e latitude.

Para a realização dos modelos, os dados ambientais foram transformados do formato “tif” para formato “asc”. Além disso, todos foram padronizados com datum WGS84 e projeção geográfica Lat/Long.

No MaxEnt, usamos *Regularization Multiplier* (RM) = 1. Essa configuração altera a complexidade do modelo para aumentar ou diminuir o sobreajuste (*overfitting*). Adotamos este valor porque quanto menor for o RM, menor será o erro de omissão. Configuramos o número máximo de pontos de fundo (*Max number of background points*) para dez mil. O tipo de funcionamento de réplica (*replicated run type*) usado foi o *crossvalidate*, que é o método equivalente ao *jackknife*, em que o número de replicações é igual ao número de pontos de ocorrência menos um. Como o objetivo principal do trabalho foi comparar as variáveis ambientais utilizadas nas modelagens de distribuição de espécies, o método *jackknife* permitiu que os modelos fossem testados e replicados com base em todos os pontos de ocorrência. Dessa forma, neste projeto usamos 346 réplicas para gerar os modelos. Adotamos 5

mil como o número máximo de interações (*maximum iterations*). No total, foram realizados seis modelos (Tabela 2).

Após os modelos serem gerados no MaxEnt, o mapa contínuo retornado da modelagem foi transformado em um mapa binário, que mostra áreas de adequabilidade para ocorrência e áreas não adequadas para a ocorrência da espécie. Para isso, adotamos como *threshold* a sensibilidade máxima do teste (Tabela 3). Escolhemos a sensibilidade máxima por se tratar de uma espécie com muitos pontos registrados e por não ter certeza do quanto estes dados são precisos. Além disso, a espécie *E. coriacea* pode ser confundida com outras espécies semelhantes de mesmo gênero. Segundo Mori (2010), é difícil que todas as coleções identificadas como *E. coriacea* representem verdadeiramente a espécie.

Tabela 2. Modelos realizados no MaxEnt para comparação das variáveis ambientais e os valores de threshold (sensitividade máxima) para cada modelo.

	Espécie	Fonte	Variáveis	Resolução	Threshold
Modelagem 1	<i>E. coriacea</i>	WorldClim 1.4	bio 1, bio 12	1 km	0.55
Modelagem 2	<i>E. coriacea</i>	WorldClim 2	bio 1, bio 12	1 km	0.55
Modelagem 3	<i>E. coriacea</i>	CHELSA	bio 1, bio 12	1 km	0.61
Modelagem 4	<i>E. coriacea</i>	MODIS/CHIRPS	bio 1, bio 12	5,5 km	0.53
Modelagem 5	<i>E. coriacea</i>	Ecoclimate	bio 1, bio 12	50 km	0.52
Modelagem 6*	<i>E. coriacea</i>	Distintas variáveis	bio 1, bio 12, MODIS - NDVI, MODIS - Evapotranspiração, SRTM, biomassa	1 km	0.52

*As variáveis bio 1 e bio 12 nesta modelagem foram do *WorldClim 1.4*

Comparamos os mapas dos modelos gerados pelo MaxEnt, procurando padrões de semelhança e diferenças espaciais. Calculamos as divergências da predição da distribuição da espécie entre os mapas, isto é, áreas em que um modelo considerou adequado para a distribuição da espécie e outros modelos não consideraram.

Para validar os modelos, usamos o valor de teste *Area Under Curve* (AUC) gerado no MaxEnt e que serve como uma medida de avaliação (LIU et al., 2005). Um valor de AUC igual a 0.5 indica que o modelo não é melhor do que uma

modelagem aleatória. Portanto, quanto mais próximo de 1 o valor de AUC, melhor é o modelo (DE MARCO JUNIOR, 2009). Para um modelo ser considerado bom, o ideal é obter um valor de AUC superior a 0.7 (DALAPICOLLA, 2016). Externamente ao MaxEnt, calculamos a porcentagem de acerto e a porcentagem de omissão de cada modelo. Para saber a porcentagem de acerto utilizamos os pontos de ocorrência da espécie que foram incluídas no MaxEnt. Depois extraímos o valor de cada ponto no mapa binomial que prediz a distribuição da espécie e calculamos a porcentagem de pontos preditos como ocorrência de *E. coriacea*. Sendo assim, a taxa de omissão resultou na porcentagem de pontos da espécie que não foram preditas pelo modelo.

Outra forma de validação foi criar uma matriz de confusão usando os dados de presença utilizados na modelagem e pontos criados de forma aleatória. Foram criados 3.470 pontos aleatórios, isto é, dez vezes mais que a quantidade de pontos utilizados para fazer os modelos. Essa forma de validação permite avaliar se os modelos predizem a distribuição da espécie melhor do que o acaso. Através da matriz fizemos um teste binomial usando os dados da matriz de confusão para calcular o *p-value* de cada modelo.

Tabela 3. Formação de uma matriz de confusão. A (Presença verdadeira) – B (Falsa presença) – C (Falsa ausência) – D (Ausência verdadeira).

	Presença real	Ausência real
Presença prevista	A	B
Ausência prevista	C	D

4 RESULTADOS E DISCUSSÃO

O modelo que resultou na maior área de distribuição predita para *E. coriacea* foi o ajustado a partir dos dados WorldClim versão 1.4. A área de distribuição potencial totalizou uma área de 3.536.678 km² (35,53% da área total da Amazônia sensu-latíssimo) para a ocorrência da espécie. Por outro lado, o modelo que gerou a menor área foi o ajustado com os dados da base Ecoclimate, resultando em um total de 2.917.500 km² (29,31%) adequados para a ocorrência da espécie. A diferença entre esses dois modelos foi de 619.178 km² (6,22%) (Tabela 5, Figura 3).

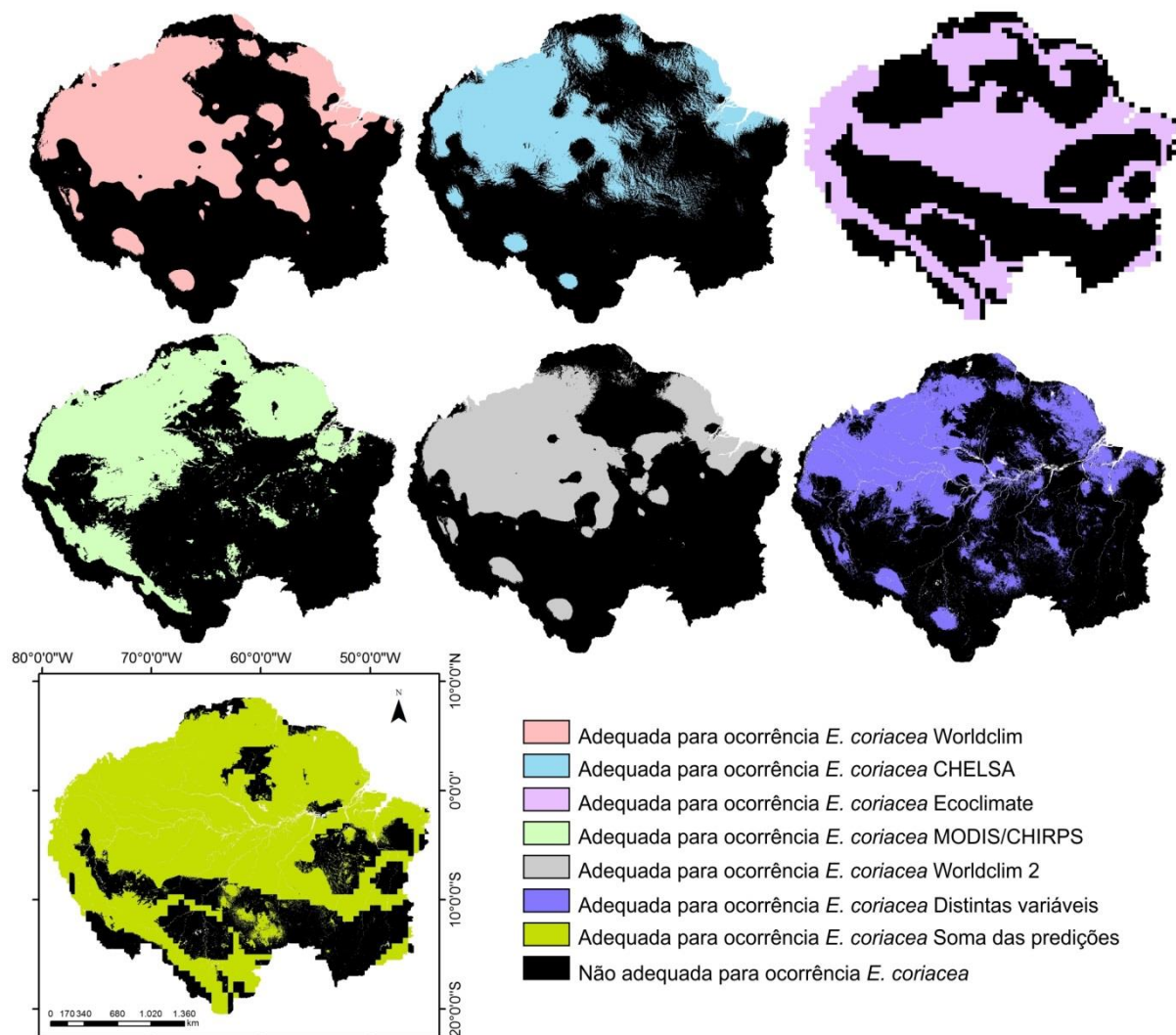


Figura 3. Mapas da modelagem de distribuição de *E. coriacea* na Amazônia sensu-latíssimo para cada modelo estimado pelo algoritmo MaxEnt, usando diferentes variáveis ambientais. (Fonte: Elaborado pelo autor).

Os modelos tiveram diferenças quanto à contribuição das variáveis. É interessante notar que nos modelos com as duas versões do WorldClim, o percentual de contribuição da variável de precipitação anual (bio 12) foi de 99,99%, ou seja, os resultados da área de distribuição predita pelos dados WorldClim foram baseados praticamente em uma única variável. Porém, no modelo com as distintas variáveis, mesmo usando os dados bioclimáticos WorldClim 1.4, a contribuição da variável bio 12 (66,59%) foi menor. Além disso, a variável bio 1 que nos modelos utilizando dados WorldClim havia sido insignificante (0,01%), agora aumenta em porcentagem de contribuição para o modelo gerado (7,77%). Isso indica que uma mesma variável pode contribuir de maneira diferente, dependendo das outras variáveis incluídas no modelo (Tabela 5).

A área predita de distribuição da espécie mais divergente das demais foi a resultante das variáveis Ecoclimate. Nesta predição, foram encontradas as maiores áreas de distribuição da espécie exclusiva de um único modelo (1.261.386 km²). As maiores áreas preditas exclusivamente por um único modelo como inadequadas para a distribuição da espécie também foram do modelo gerado a partir de variáveis Ecoclimate (773.220 km²).

As divergências em área do modelo que utilizou dados MODIS/CHIRPS também foram grandes. A área exclusiva deste modelo tido como adequado para a distribuição da espécie foi equivalente a 451.417 km² e a área exclusiva como inadequada equivalente a 239.513 km² (Tabela 6, Figura 4, Figura 5). Não por acaso, os dois modelos que mais se diferenciaram dos demais foram os únicos a utilizarem resoluções espaciais mais baixas (tamanhos de célula maior para as variáveis ambientais) diferentes das variáveis, evidenciando que não só a origem e natureza dos dados espaciais é importante, mas também sua resolução espacial. A modelagem feita com os dados Ecoclimate foi a única das seis modelagens realizadas que não teve a variável bio 12 como a de maior contribuição; a variável de maior contribuição para este modelo foi bio 1 (86,92%) (Tabela 5).

O modelo realizado com dados CHELSA foi o que menos previu a distribuição da espécie em áreas exclusivas (não preditas pelos demais modelos). Foram 47.402 km² preditos como adequados para a ocorrência da espécie, não indicados pelos outros cinco modelos. Nas circunstâncias em que cinco modelos concordaram acerca da distribuição da espécie para uma área e um único modelo predisse como

inadequada, encontramos a menor área na modelagem feita com dados WorldClim 1.4 (6.632 km²) (Tabela 6, Figura 4, Figura 5).

Tabela 4. Resultados de cada modelo feito pelo MaxEnt.

	Fonte	Área predita de distribuição de <i>E. coriacea</i>	Contribuição bio1	Contribuição bio 12	Teste AUC	p-value	Nº de variáveis usadas
Modelagem 1	WorldClim 1.4	3536,678 km ²	0.01	99.99	0.68	p<0,001	2
Modelagem 2	WorldClim 2	3437,670 km ²	0.01	99.99	0.67	p<0,001	2
Modelagem 3	CHELSA	2991,629 km ²	6.48	93.52	0.65	p<0,001	2
Modelagem 4	MODIS/CHIRPS	3073,158 km ²	34.97	65.03	0.68	p<0,001	2
Modelagem 5	Ecoclimate	2917,500 km ²	86.92	13.08	0.67	p<0,001	2
Modelagem 6	Ditintas variáveis	3181,048 km ²	7.77	66.59	0.73	p<0,001	6

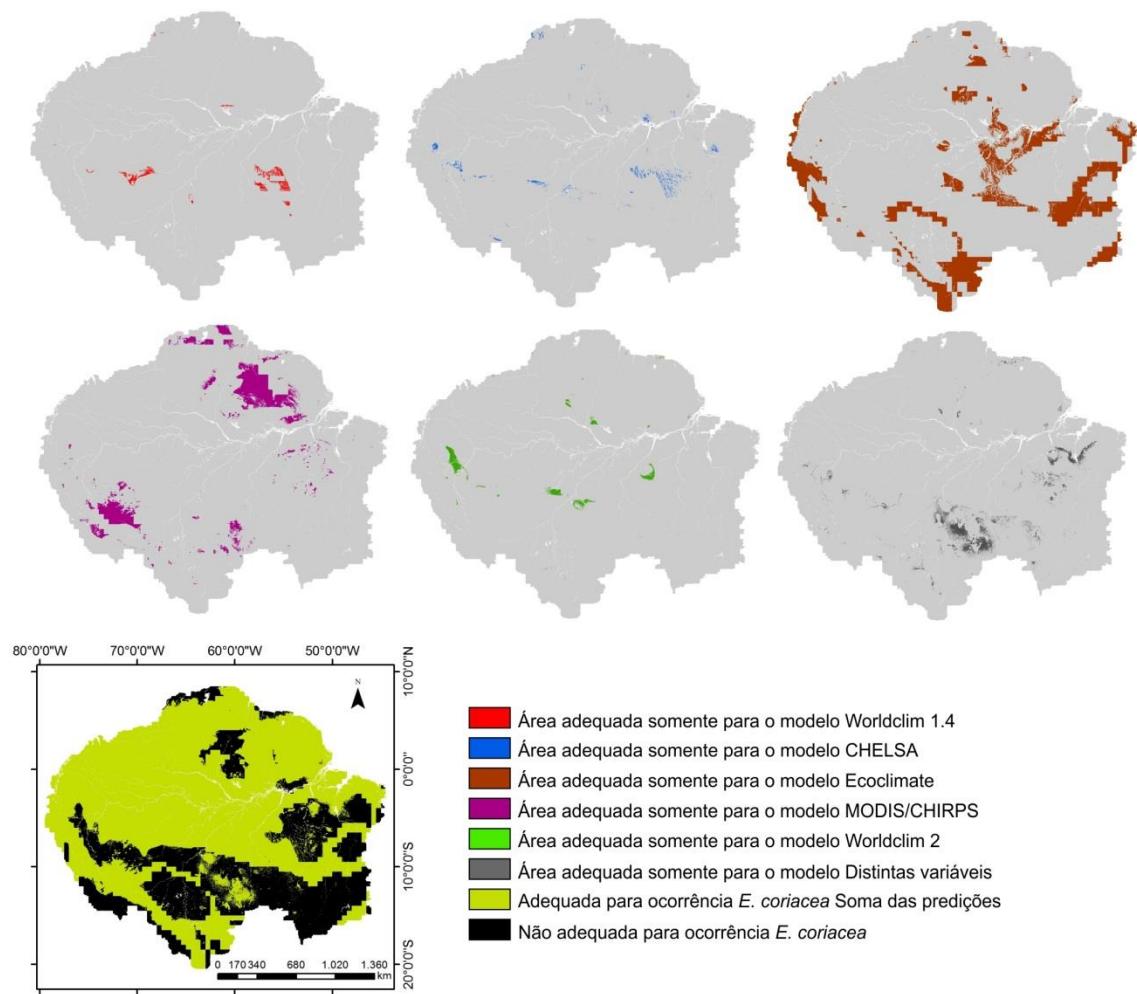


Figura 4. Áreas previstas para a distribuição de *E. coriacea* exclusivas de cada modelo feito no MaxEnt com as diferentes variáveis. (Fonte: Elaborado pelo autor).

Tabela 5. Comparação dos modelos gerados pelo MaxEnt com previsões exclusivas de um único modelo como adequado e inadequado para a distribuição de *E. coriacea*.

Predições exclusivas	Total (km²)
nenhuma das modelagens	2.905.621
somente Worldclim (adequado)	55.002
somente Worldclim 2 (adequado)	65.835
somente CHELSA (adequado)	47.402
somente MODIS/CHIRPS (adequado)	451.417
Somente Ecoclimate (adequado)	1.261.386
Somente com distintas variáveis (adequado)	174.536
Somente com distintas variáveis (inadequado)	112.210
Somente Ecoclimate (inadequado)	773.220
Somente MODIS/CHIRPS (inadequado)	239.513
Somente CHELSA (inadequado)	145.642
Somente Worldclim 2 (inadequado)	37.772
Somente Worldclim (inadequado)	6.632
Area total da previsão somando todos os modelos	7.048.621
Área total	9.954.242

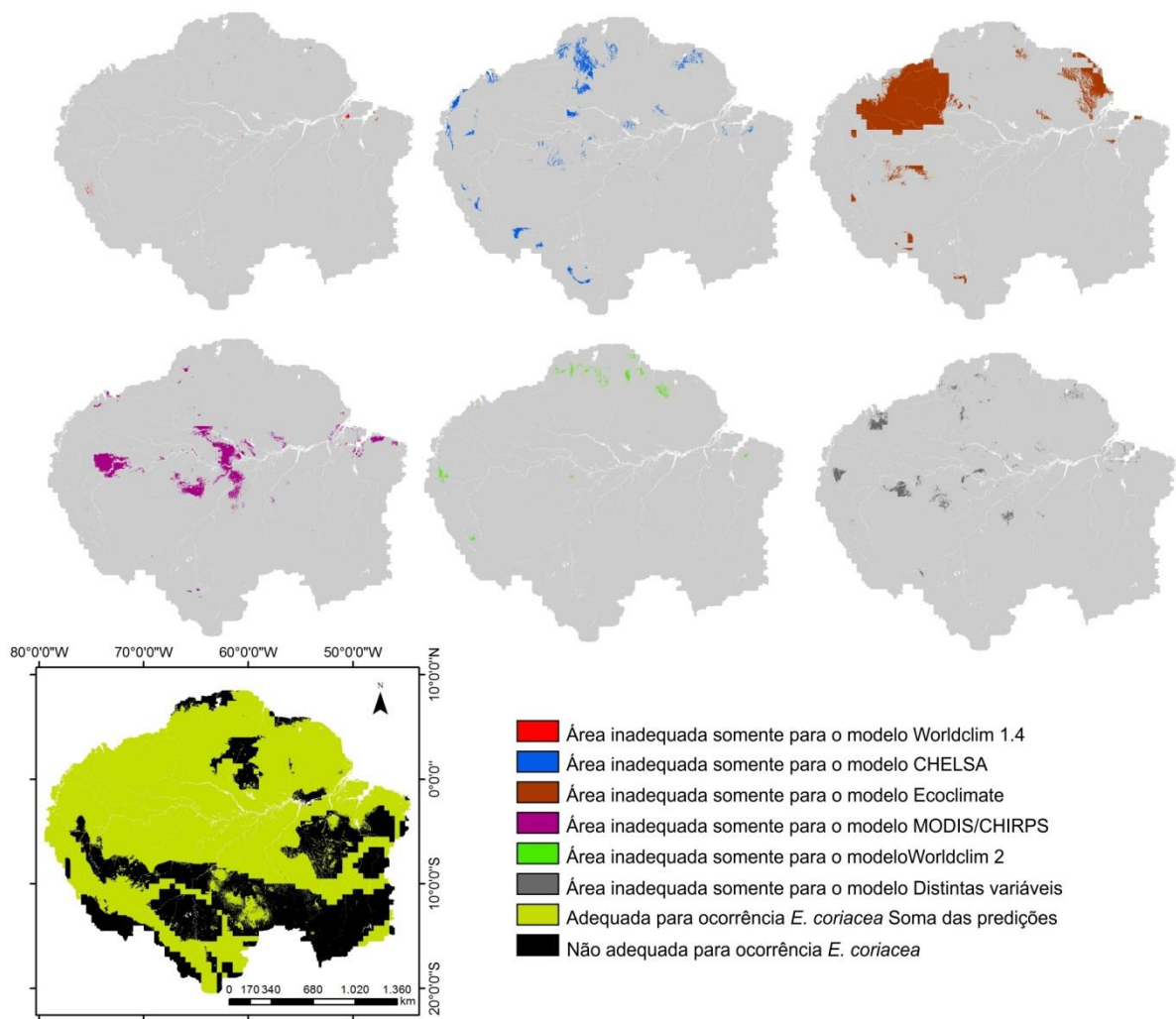


Figura 5. Áreas não previstas para a distribuição de *E. coriacea* exclusivas de cada modelo feito no MaxEnt com as diferentes variáveis. (Fonte: Elaborado pelo autor).

Ao somarmos os mapas da distribuição de *E. coriacea* gerados pelo MaxEnt, tivemos um total de 7.048.621 km² (70,81% da área total da Amazônia sensu-latíssimo) como adequada para a distribuição da espécie. Porém, desta soma total das predições, somente 979.629 km² (9,84%) foram preditos por todas as modelagens. Isso mostra que de toda a área gerada pelos modelos como adequado para a distribuição da espécie, apenas 9,84% da área é comum entre todos os modelos gerados (Figura 6).

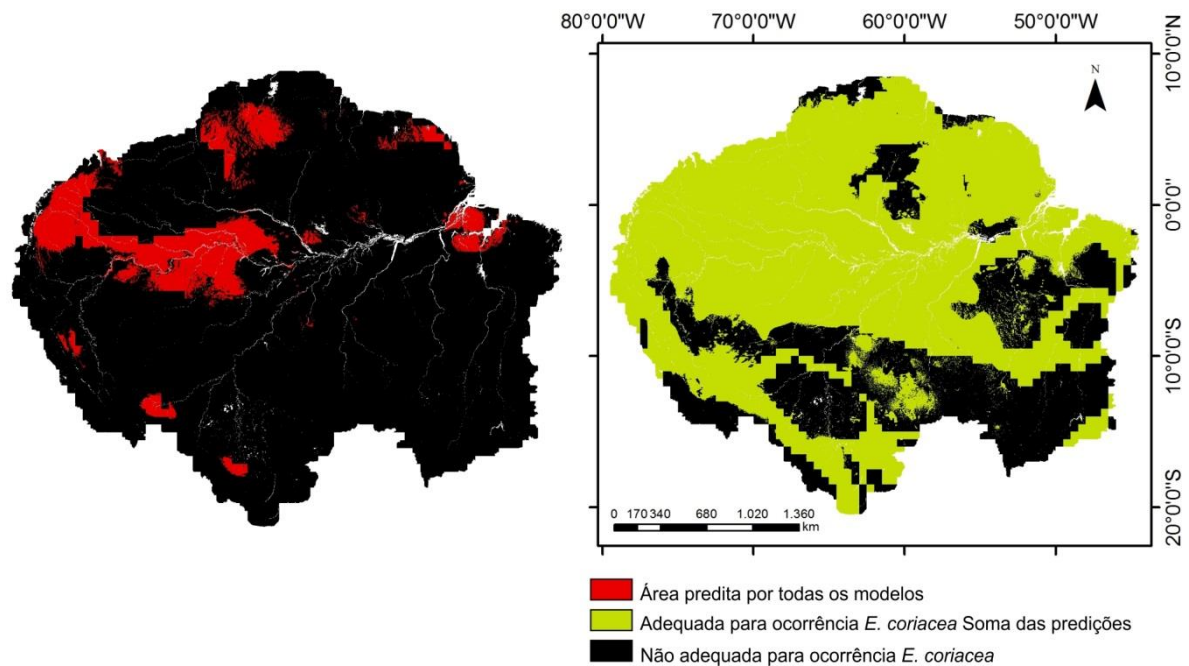


Figura 6. Comparação da predição de todas as modelagens somadas com a área predita em comum por todos os modelos. (Fonte: Elaborado pelo autor).

A área não predita para a distribuição da espécie em nenhum dos modelos foi de 2.911.369 km² (29,19% do total da área da Amazônia sensu-latíssimo). A área predita por apenas um modelo foi de 2.049.948 km² (20,59%). A área predita por dois modelos foi de 826.892 km² (8,31%). A área predita por três modelos foi de 597.407 km² (6%). A área predita por quatro modelos foi de 657.475 km² (6,60%). A área predita por cinco modelos foi de 1.311.337 km² (13,17%) (Figura 7).

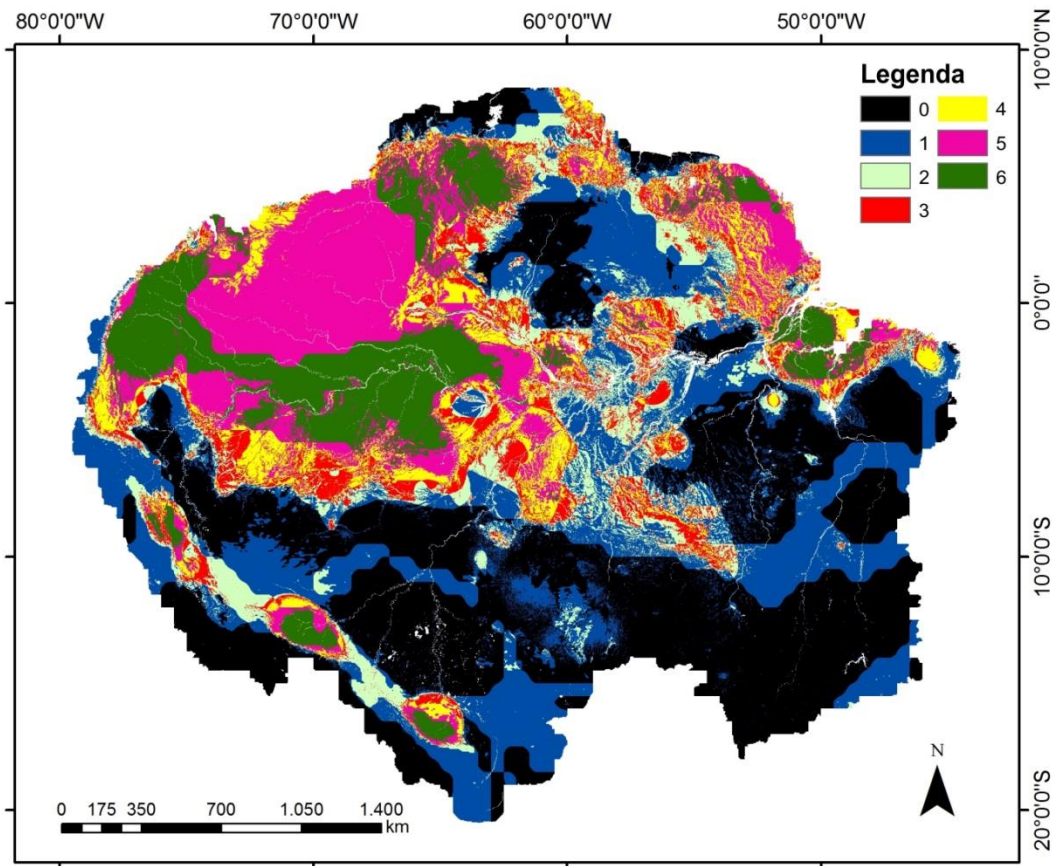


Figura 7. Grau de concordância entre os modelos em relação à área de distribuição da espécie. (Fonte: Elaborado pelo autor).

A maior porcentagem de acerto das predições foi encontrada no modelo com variáveis Ecoclimate (Modelagem 5), chegando a 73,2% de acerto. Mesmo sendo o modelo mais diferente dentre os seis realizados e com a resolução mais grosseira (50 km), foi o que obteve maior porcentagem de acerto. Em seguida foi o modelo com distintas variáveis (Modelagem 6), com 71,6% de acerto. A predição com menor porcentagem de acerto foi o modelo gerado com variáveis CHELSA (Modelagem 3), tendo apenas 55,6% de acerto. O modelo com dados MODIS/CHIRPS (Modelagem 4), com 66,7%, foi a segunda com menor porcentagem de acerto. Isso indica que mesmo variáveis com resolução mais grosseiras podem contribuir com uma boa porcentagem de acerto no modelo. Por outro. Além disso, a inclusão de outras variáveis pode aperfeiçoar a porcentagem de acerto dos modelos. Isso pode acontecer porque ao analisarmos a modelagem realizada com outras variáveis incluídas além de bio 1 e bio 12, notamos um aumento na porcentagem de acerto. Para a modelagem com distintas variáveis, foram utilizadas as variáveis bio 1 e bio

12 provenientes dos dados WorldClim 1.4, que no primeiro modelo tiveram porcentagem de acerto igual a 68,3%. Quando foram incluídas as outras variáveis (Modelagem 6), a porcentagem de acerto passou para 71,6%.

Tabela 6. Porcentagem de acerto e erro de cada modelo gerado pelo MaxEnt.

	Porcentagem de acerto (%)	Porcentagem de erro (%)
WorldClim 1.4	68,3	31,7
WorldClim 2	63,4	36,6
CHELSA	55,6	44,4
Ecoclimate	73,2	26,8
MODIS/CHIRPS	66,7	33,3
Distintas Variáveis	71,6	28,4

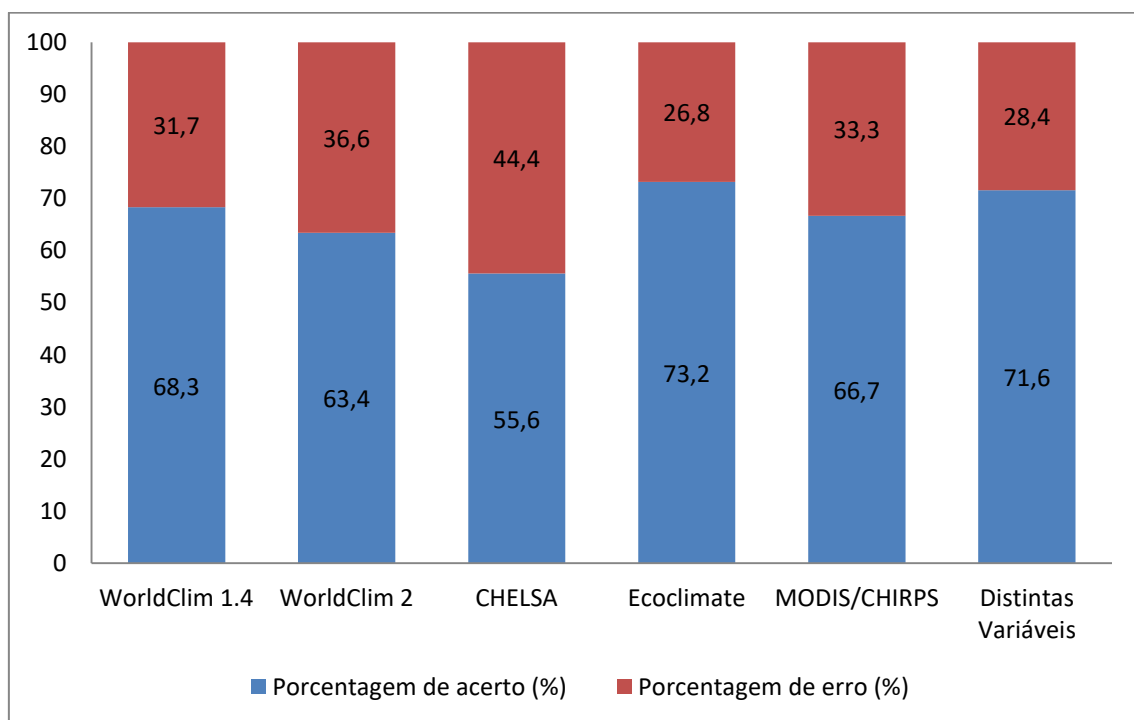


Figura 8. Porcentagem de acerto e erro de cada modelagem gerado pelo MaxEnt. (Fonte: Elaborado pelo autor).

Estatisticamente, encontramos o *p-value* de todos os modelos $<0,05$, e todos os valores de AUC foram superiores a 0.5. Porém, o único superior a 0.7 foi o modelo gerado com mais variáveis ambientais além de bio 1 e bio 12 (AUC = 0.73). Portanto, a inclusão de outras variáveis importantes para as plantas, como índices

de evapotranspiração, biomassa, altitude e atividade clorofiliana podem melhorar a predição dos modelos (Tabela 5).

Como discutido no trabalho de Mod e colaboradores (2016), muitos estudos estão negligenciando outros fatores além de variáveis de temperatura e precipitação. Aqui mostramos como a inclusão de outras variáveis também tidas como fundamentais para as plantas podem acrescentar na confiabilidade do modelo e mudar a projeção de distribuição das espécies.

5 CONCLUSÃO

A seleção das variáveis ambientais a serem utilizadas em uma modelagem de distribuição de espécie é um processo importante e que pode levar a resultados completamente diferentes de predição da distribuição das espécies estudadas. Além disso, as importâncias das variáveis ambientais podem mudar dependendo do conjunto de dados utilizados no modelo. Portanto, a seleção adequada das variáveis ambientais é fundamental para que cada variável possa contribuir com o modelo de distribuição.

Observamos também que as resoluções espaciais das variáveis utilizadas podem mudar drasticamente o resultado de um modelo. Os modelos que mais se diferenciaram em termos de áreas de distribuição preditas foram aqueles que utilizaram resoluções mais grossas. Porém, não necessariamente são os modelos mais imprecisos, pois podem possuir as maiores porcentagens de acertos e estarem estatisticamente confiáveis.

REFERÊNCIAS

- AUSTIN, Mike P.; VAN NIEL, Kimberly P. Impact of landscape predictors on climate change modelling of species distributions: a case study with *Eucalyptus fastigata* in southern New South Wales, Australia. **Journal of Biogeography**, v. 38, n. 1, p. 9-19, 2011.
- AVITABILE, Valerio et al. An integrated pan-tropical biomass map using multiple reference datasets. **Global change biology**, v. 22, n. 4, p. 1406-1420, 2016.
- BACCINI, A. G. S. J. et al. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. **Nature climate change**, v. 2, n. 3, p. 182-185, 2012.
- BEGON, M., TOWNSEND, C. R., HARPER, J. L. **Ecologia: de indivíduos a ecossistemas**. Artmed Editora, 2009.
- DALAPICOLLA, J. 2016. **Tutorial de modelos de distribuição de espécies: guia prático usando o MaxEnt e o ArcGIS 10**. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://blog.ufes.br/lamab/tutoriais>
- DEBLAUWE, Vincent et al. Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. **Global Ecology and Biogeography**, v. 25, n. 4, p. 443-454, 2016.
- DE MARCO JÚNIOR, P.; SIQUEIRA, M. F. Como determinar a distribuição potencial de espécies sob uma abordagem conservacionista. **Megadiversidade**, v. 5, p. 65-76, 2009.
- DORMANN, Carsten F. Promising the future? Global change projections of species distributions. **Basic and applied ecology**, v. 8, n. 5, p. 387-397, 2007.
- ELITH, Jane; KEARNEY, Michael; PHILLIPS, Steven;. The art of modelling range-shifting species. **Methods in ecology and evolution**, v. 1, n. 4, p. 330-342, 2010.
- EVA, Hugh D. et al. **A proposal for defining the geographical boundaries of Amazonia; synthesis of the results from an expert consultation workshop organized by the European Commission in collaboration with the Amazon Cooperation Treaty Organization-JRC Ispra, 7-8 June 2005**. EC, 2005.
- FERNÁNDEZ, M.; HAMILTON, H.; KUEPPERS, L. M. Characterizing uncertainty in species distribution models derived from interpolated weather station data. **Ecosphere**, v. 4, n. 5, p. 1-17, 2013.
- FICK, Stephen E.; HIJMANS, Robert J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, 2017.

FRANKLIN, Steven E. **Remote sensing for sustainable forest management**. CRC Press, 2001.

FUNK, Chris et al. The climate hazards infrared precipitation with stations--a new environmental record for monitoring extremes. **Scientific data**, v. 2, p. 150066, 2015.

GUISAN, Antoine et al. Predicting species distributions for conservation decisions. **Ecology letters**, v. 16, n. 12, p. 1424-1435, 2013.

GUISAN, Antoine; ZIMMERMANN, Niklaus E. Predictive habitat distribution models in ecology. **Ecological modelling**, v. 135, n. 2, p. 147-186, 2000.

HESS, L.L., J.M. MELACK, A.G. AFFONSO, C.C.F. BARBOSA, M. GASTIL-BUHL, AND E.M.L.M. NOVO. LBA-ECO LC-07 Wetland Extent, Vegetation, and Inundation: Lowland Amazon Basin. **ORNL DAAC, Oak Ridge, Tennessee, USA**. 2015.

HIJMANS, R.J., CAMERON, S.E., PARRA, J.L., JONES, P.G. & JARVIS, A. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology** 25: 1965-1978, 2005.

HORTAL, J., DE BELLO, F., DINIZ-FILHO, J. A. F., LEWINSOHN, T. M., LOBO, J. M., & LADLE, R. J. Seven shortfalls that beset large-scale knowledge of biodiversity. **Annual Review of Ecology, Evolution, and Systematics**, 46, 523-549, 2015.

JENSEN, John R. **Remote sensing of the environment: An earth resource perspective 2/e**. Pearson Education India, 2009.

JUNK, W. J. Current state of knowledge regarding South America wetlands and their future under global climate change. **Aquatic Sciences**, v. 75, n. 1, p. 113-131, 2013.

KARGER, Dirk Nikolaus et al. Climatologies at high resolution for the earth's land surface areas. **arXiv preprint arXiv:1607.00217**, 2016.

KÖRNER, Christian. Pflanzen im lebensraum. In: **Strasburger– Lehrbuch der Pflanzenwissenschaften**. Springer Berlin Heidelberg, 2014. p. 811-841.

LIMA-RIBEIRO, Matheus Souza et al. EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. **Biodiversity Informatics**, v. 10, 2015.

LOMOLINO, Mark V. et al. **Biogeography**. Sunderland, MA: Sinauer Associates, 2006.

NEW, Mark et al. A high-resolution data set of surface climate over global land areas. **Climate research**, v. 21, n. 1, p. 1-25, 2002.

MCGILL, Brian J. Matters of scale. **Science**, v. 328, n. 5978, p. 575-576, 2010.

MOD, Heidi K. et al. What we use is not what we know: environmental predictors in plant distribution models. **Journal of Vegetation Science**, v. 27, n. 6, p. 1308-1322, 2016.

MORI, S. A. et al. The Lecythidaceae Pages. The New York Botanical Garden, Bronx, New York. 2010.

MORI, SCOTT A.; BECKER, PETER; KINCAID, DWIGHT. Lecythidaceae of a central Amazonian lowland forest. **Lessons from Amazonia: the ecology and conservation of a fragmented forest. Yale University Press, New Haven, Conn**, p. 54-67, 2001.

P ANDERSON, Robert et al. Novel methods improve prediction of species' distributions from occurrence data. **Ecography**, v. 29, n. 2, p. 129-151, 2006.

PEARSON, Richard G.; DAWSON, Terence P. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?. **Global ecology and biogeography**, v. 12, n. 5, p. 361-371, 2003.

PETERSON, A. T., SOBERÓN, J., PEARSON, R. G., ANDERSON, R. P., MARTÍNEZ-MEYER, E., NAKAMURA, M., & ARAÚJO, M. B. **Ecological Niches and Geographic Distributions (MPB-49)**. Princeton University Press, 2011.

PHILLIPS, Steven J. et al. A brief tutorial on Maxent. **AT&T Research**, 2005.

PHILLIPS, Steven J.; ANDERSON, Robert P.; SCHAPIRE, Robert E. Maximum entropy modeling of species geographic distributions. **Ecological modelling**, v. 190, n. 3, p. 231-259, 2006.

ROGAN, John; CHEN, DongMei. Remote sensing technology for mapping and monitoring land-cover and land-use change. **Progress in planning**, v. 61, n. 4, p. 301-325, 2004.

ROUGHGARDEN, J.; RUNNING, Steven W.; MATSON, Pamela A. What does remote sensing do for ecology?. **Ecology**, v. 72, n. 6, p. 1918-1922, 1991.

SAATCHI, Sassan et al. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. **Remote Sensing of Environment**, v. 112, n. 5, p. 2000-2017, 2008.

SYPHARD, Alexandra D.; FRANKLIN, Janet. Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. **Ecography**, v. 32, n. 6, p. 907-918, 2009.

TER STEEGE, Hans et al. Hyperdominance in the Amazonian tree flora. **Science**, v. 342, n. 6156, p. 1243092, 2013.

WASHINGTON, Richard et al. Congo Basin rainfall climatology: can we believe the climate models?. **Phil. Trans. R. Soc. B**, v. 368, n. 1625, p. 20120296, 2013.

Rio Claro, 18 de Dezembro de 2.017

Orientador: Thiago Sanna Freire Silva

Co-orientador: Bruno Garcia Luize

Aluno: Abner Castro de Carvalho