UNIVERSIDADE FEDERAL DO PARANÁ

SANTIAGO JOSÉ ELÍAS VELAZCO

# O CERRADO ALÉM DAS FRONTEIRAS: PADRÕES DE DIVERSIDADE E AMEAÇAS PRESENTES E FUTURAS À SUA FLORA NA BOLÍVIA, BRASIL E PARAGUAI

CURITIBA
2018

SANTIAGO JOSÉ ELÍAS VELAZCO

# O CERRADO ALÉM DAS FRONTEIRAS: PADRÕES DE DIVERSIDADE E AMEAÇAS PRESENTES E FUTURAS À SUA FLORA NA BOLÍVIA, BRASIL E PARAGUAI

Tese apresentada ao Curso de Pós-Graduação em Engenharia Florestal do Setor de Ciências Agrárias da Universidade Federal do Paraná, como requisito à obtenção do grau e título de Doutor em Ciências Florestais.

Orientador: Dr. Franklin Galvão
Coorientador: Dr. Paulo De Marco Júnior
Coorientador: Dr. Fabricio Villalobos

CURTIBA
2018

Ficha catalográfica elaborada pela
Biblioteca de Ciências Florestais e da Madeira - UFPR

Velazco, Santiago José Elías
O cerrado além das fronteiras: padrões de diversidade e ameaças presentes e futuras à sua flora na Bolívia, Brasil e Paraguai / Santiago José Elías Velazco. – Curitiba, 2018.
175 f. : il.

Orientador: Prof. Dr. Franklin Galvão
Coorientadores: Prof. Dr. Paulo De Marco Júnior
Prof. Dr. Fabricio Villalobos
Tese (Doutorado) - Universidade Federal do Paraná, Setor de Ciências Agrárias, Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 06/03/2018.
Área de concentração: Conservação da Natureza.

1. Plantas dos cerrados - Bolívia. 2. Plantas dos cerrados - Brasil. 3. Plantas dos cerrados - Paraguai. 4. Mudanças climáticas. 5. Cerrados - Conservação. 6. Teses. I. Galvão, Franklin. II. Marco Júnior, Paulo de. III. Villalobos, Fabricio. IV. Universidade Federal do Paraná, Setor de Ciências Agrárias. V. Título.
CDD – 634.9
CDU – 634.0.182.3(8)

Bibliotecária: Berenice Rodrigues Ferreira – CRB 9/1160

MINISTÉRIO DA EDUCAÇÃO
SETOR SETOR DE CIENCIAS AGRARIAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA FLORESTAL

# TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **SANTIAGO JOSE ELIAS VELAZCO** intitulada: **O CERRADO ALÉM DAS FRONTEIRAS: PADRÕES DE DIVERSIDADE E AMEAÇAS PRESENTES E FUTURAS À SUA FLORA NA BOLÍVIA, BRASIL E PARAGUAI**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 06 de Março de 2018.

FRANKLIN GALVÃO
Presidente da Banca Examinadora

Paulo De Marco Jr.
PAULO DE MARCO JÚNIOR
Avaliador Externo

Marília Borgo
MARÍLIA BORGO
Avaliador Externo

Kelly Geronazzo Martins
KELLY GERONAZZO MARTINS
Avaliador Externo

Marques
MARCIA CRISTINA MENDES MARQUES
Avaliador Externo

As minhas avós Ana Scholz e
Graciolinda Almeida

DEDICO

# AGRADECIMENTOS

A meus orientadores Franklin Galvão, Paulo De Marco Jr e Fabricio Villalobos, por me ensinarem tanto, não só com suas explicações, mas também por serem exemplos de pessoas e de cientistas.

Ao professor Franklin, por ter confiando em mim e por seu constante apoio e estímulo.

Ao professor Paulo, por abrir as portas do seu populoso e receptivo laboratório, pelas muitas conversas e risadas acompanhadas com pães de queijo, bolachas e cafés. Obrigado pela tua imensa humanidade.

Ao TheMetaLand e todos os seus integrantes, por sua parceria, amizade e os ótimos momentos em Goiás.

A meus amigos Pablo Hoffmann, Marilia Borgo e Marcelo Bosco Pinto que tantas vezes me buscaram no tubo Ferrari, acolherar nas suas casas e cuidaram de mim em Campo Largo.

A meus tios Dario y Malena por terem presenteado este computador que tantas análises e modelos criou.

Ao John Wood, por revisar a lista de espécies do Cerrado.

À Marcia Cristina Mendes Marques, Kelly Geronazzo Martins, Marilia Borgo, Paulo De Marco Júnior e Ana Paula Dalla Corte, por lerem este trabalho, seus questionamentos e recomendações.

A minha amda Natalia Alejandra Bedrij, pelo estímulo, ajuda e amor em todos esses anos.

Ao R Core Team e todos as pessoas que trabalham de forma coletiva nos pacotes, sem vocês dificilmente teria alcançado os objetivos desta tese.

À Universidade Federal do Paraná e a o programa de Pós-graduação em Engenharia Florestal, por me oferecer a oportunidade de me aperfeiçoar.

Ao programa de Pós-graduação em Ecologia e Evolução da Universidade Federal de Goiás, por me permitir participar das disciplinas que tanto me serviram.

Ao Instituto de Biologia Subtropical, por me deixar ocupar um espaço para trabalhar nos últimos meses da minha tese.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa concedida.

Ao povo Brasileiro que me acolheu sempre de braços abertos dos quais nunca me desprenderei.

# RESUMO

No primeiro capítulo " *Using worldwide edaphic data to model plant species niches: an assessment at a continental extent* ", foi abordado as seguintes questões principais: (i) Qual é a influência de distintas variáveis preditoras (por exemplo, climática vs. edáfica) em diferentes algoritmos empregados em modelos de nicho ecológico (ENM´s, em inglês)? (ii) Qual é a relação entre o desempenho de diferentes preditores e as características geográficas das espécies? Utilizou-se 125 espécies de plantas distribuídas na região Neotropical para explorar o efeito de usar dados edáficos disponíveis no banco de dados SoilGrids e sua combinação com dados climáticos do CHELSA. Além disso, essas diferentes variáveis foram relacionadas com as características geográficas das espécies e diferentes algoritmos. O uso de diferentes preditores (climáticos, edáficos e ambos) afetou significativamente o desempenho do modelo e a complexidade espacial das previsões. Foi demostrado que o uso simultâneo de variáveis climáticas e edáficas geram ENM´s com acurácia similar ou melhor que as construídas apenas com variáveis climáticas. Além disso, o desempenho de modelos que consideram essas diferentes variáveis preditoras, separadamente ou em conjunto, teve uma relação com as propriedades geográficas das ocorrências das espécies, como o número e a amplitude de distribuição. A grande extensão geográfica, a variabilidade dos ambientes e as características geográficas das diferentes espécies consideradas permitiram demonstrar que os dados edáficos globais agregam informações úteis para ENM´s da planta. Isto é particularmente valioso para estudos de espécies que ocorrem em regiões onde informações mais detalhadas sobre propriedades do solo são pobres ou não existem. No segundo capítulo "*Looking beyond borders: patterns of richness and rarity of the Cerrado flora and its transboundary conservation opportunities*", foi estuado a flora do Cerrado de forma transfronteiriça para (i) prever padrões de riqueza e raridade de espécies para Bolívia, Brasil e Paraguai, (ii) estudar a concentração de espécies raras a nível global e regional, e (iii) avaliar áreas que apresentam oportunidades e conflitos para conservar a riqueza e a raridade do Cerrado nas três nações. Foi utilizado sete técnicas de modelagem de nicho ecológico para modelar 1559 espécies de plantas do Cerrado. Definimos cenários de oportunidade e conflito para a conservação com base na relação entre o grau de alteração da paisagem e a riqueza e concentração da raridade global e local. A maior concentração de riqueza de espécies e raridade global são encontradas na parte central da ecorregião do Cerrado no Brasil. A Bolívia é o segundo país com maior riqueza e raridade, principalmente associado ao Cerrado e à Floresta Seca Chiquitana. No Paraguai, a riqueza e a raridade estavam concentradas na parte norte do Chaco Úmido e Chaco Seco. Oportunidades e conflitos de conservação foram diferentes para cada país devido à interação entre a diversidade e os padrões de uso da terra. Na medida do nosso conhecimento, esta é a primeira tentativa de descrever a diversidade vegetal do bioma Cerrado além dos limites políticos brasileiros. Apesar de grande parte da flora do Cerrado estar concentrada no Brasil, este país também possui a maior proporção de seu território em um cenário de conflito onde a maior alteração de paisagem ocorre em áreas com maior diversidade. No entanto, foram encontradas grandes oportunidades de conservação na Bolívia e no Paraguai, porque a riqueza e a raridade de espécies preditas estendem-se em áreas com baixo grau de alteração, tornando-os países importantes para conservar a flora do Cerrado. O terceiro capítulo "*A dark scenario for Cerrado plant species, effects of future climate, land use and inefficiency of protected areas*" teve como objetivo (i) avaliar o efeito da mudança climática e do uso da terra na distribuição das espécies de plantas do Cerrado para diferentes países onde ocorre, (ii) avaliar a eficiência da atual rede de áreas protegidas para salvaguardar espécies sob diferentes emissões de gases de efeito estufa e uso da terra, e (iii) estimar o grau de ameaça de espécies causada pela eficiência de proteção e perda de habitat. Foi mostrado que as mudanças climáticas e a mudança do uso da terra poderiam causar grandes danos à flora do Cerrado até 2050 e 2080, mesmo em condições otimistas. Infelizmente, a maior intensidade e expansão do uso da terra ocorrerá nas regiões onde a maior riqueza será abrigada. A conservação das espécies será seriamente afetada, uma vez que a rede de áreas protegidas não é tão eficiente para protegê-las em condições atuais ou futuras. O baixo nível de proteção, junto às perdas causadas pelo avanço da fronteira agrícola levará a que a maioria das espécies seja altamente ameaçada. As estratégias de conservação do Cerrado devem ser implementadas tanto a nível nacional e como transfronteiriço devido a que o clima e os efeitos do uso da terra foram diferentes em cada país.

Palavra-chave: Modelos de distribuição espécies. Mudanças climáticas. Padrões de riqueza. Perda de hábitat. Raridade. Variáveis edáficas

# ABSTRACT

**Cerrado beyond borders: diversity patterns and current and future threats it its flora in Bolivia, Brazil and Paraguay**. **Abstract**: In the first chapter "*Using worldwide edaphic data to model plant species niches: an assessment at a continental extent*" we address the following main questions: What is the influence of distinct predictor variables (e.g. climatic vs edaphic) on different ENM algorithms? and what is the relationship between the performance of different predictors and geographic characteristics of species? We used 125 plant species distributed over the Neotropical region to explore the effect on ENMs of using edaphic data available from the SoilGrids database and its combination with climatic data from the CHELSA database. In addition, we related these different predictor variables to geographic characteristics of the target species and different ENM algorithms. The use of different predictors (climatic, edaphic, and both) significantly affected model performance and spatial complexity of the predictions. We showed that the use of global edaphic plus climatic variables generates ENMs with similar or better accuracy compared to those constructed only with climate variables. Moreover, the performance of models considering these different predictors, separately or jointly, was related to geographic properties of species records, such as number and distribution range. The large geographic extent, the variability of environments and the different species' geographical characteristics considered here allowed us to demonstrate that global edaphic data adds useful information for plant ENMs. This is particularly valuable for studies of species that are distributed in regions where more detailed information on soil properties is poor or does not even exist. In the second chapter "*Looking beyond borders: patterns of richness and rarity of the Cerrado flora and its transboundary conservation opportunities*" we studied the flora of Cerrado in a transboundary way to (i) predict species richness and rarity patterns for Bolivia, Brazil, and Paraguay, (ii) study the concentration of rare species at global and regional levels and (iii) evaluate areas presenting opportunities and conflicts to conserve the plant richness and rarity of Cerrado across these three nations. We defined scenarios of opportunity and conflict for conservation based on the relationship between the degree of landscape alteration and the richness and concentration of global and local rarity. Rarity index was based on species geographic range sizes, marginality, and specificity of their niches. The greatest concentration of species richness and global rarity of Cerrado's plant species is found in the central part of the Cerrado ecoregion in Brazil. Bolivia is the second country with the highest richness and rarity, mainly associated with Cerrado and Chiquitano Dry Forest. In Paraguay, richness and rarity were concentrated in the northern part of the Humid and Dry Chaco. Opportunities and conflicts for conservation were different for each country resulting from the interaction between diversity and land-use patterns. To the extent of our knowledge, this is the first attempt to describe the plant diversity of the Cerrado biome beyond Brazilian borders. Despite much of the Cerrado flora being concentrated in Brazil, this country also has the greatest proportion of its territory under a conflict scenario where the major landscape alteration happens in areas with the highest diversity. However, high opportunities for conservation were found in Bolivia and Paraguay where species richness and rarity concentrate in areas with low alteration degree, which makes them important countries to conserve the flora of Cerrado. "*A dark scenario for Cerrado plant species, effects of future climate, land use and inefficiency of protected areas*" We aimed to (i) assess the effect of the climate change and land-use on the distribution of the Cerrado plant species for different countries where it occurs, (ii) evaluate the efficiency of the current protected area network to safeguards species under different greenhouse-gas emissions and land-use and (iii) estimate the vulnerability of species caused by protection efficiency and habitat loss. We demonstrate that climate change and land-use will cause great damage to Cerrado flora by 2050 and 2080, even under optimistic conditions. Unfortunately, the greatest intensity and extent of land-use will have to overcome on the regions where the greatest richness will be harbored. The conservation of the species will be seriously affected since the protected areas network is not as efficient in safeguarding them under current or future conditions. The low level of protection together with the losses caused by the advance of the agricultural frontier will lead to most species being highly vulnerable. Due to the climate and land-use, effects showed different interactions in each country, conservation strategies should be implemented at transboundary and national levels.

Keywords: Climate changes. Edaphic variables. Habitat loss. Rarity. Richness patterns. Species distribution models.

# LISTA DE FIGURAS

## CAPÍTULO 1

## CAPÍTULO 2

## CAPÍTULO 3

# LISTA DE TABELAS

## CAPÍTULO 1

# ANEXOS

## CAPITULO 1

**CAPÍTULO 2**

**CAPÍTULO 3**

# SUMÁRIO

CAPÍTULO 1

**CAPÍTULO 2**

# INTRODUÇÃO GERAL

Embora seja classificado como um domínio fitogeográfico, bioma ou ecorregião (BATALHA, 2011), o Cerrado apresenta predominantemente formações vegetais do tipo savana, isto é, uma estrato herbáceo com componentes arbustivo-arbóreos esparsos. Entretanto, são comumente encontradas formações estépicas, rupestres, florestais e veredas, dentre outras, que levam à coexistência de diferentes formações vegetais numa mesma paisagem. Por estar contornado por várias ecorregiões, como a Caatinga, Mata Atlântica, Chaco Seco, Pantanal, Floresta Deciduais Chiquitanas e Floresta Amazônica, muitos de seus elementos florísticos são compartilhados com os biomas vizinhos (FRANÇOSO; HAIDAR; MACHADO, 2016). A flora do Cerrado destaque-se ao nível global pela sua alta diversidade e grau de endemismo (SILVA; BATES, 2002). No Brasil, o Cerrado abrange aproximadamente 23% do território nacional, sendo o segundo maior bioma (PENNINGTON; LEWIS; RATTER, 2006; RATTER; RIBEIRO; BRIDGEWATER, 1997). Grande parte dessa superfície se encontra na região do Planalto Central, nos estados de Goiânia, Brasília, parte de Minas Gerais, Tocantins e Mato Grosso do Sul (PENNINGTON; LEWIS; RATTER, 2006). Ao Norte e Nordeste está presente nos estados do Piauí, Maranhão, Ceará e Bahia, muitas vezes em áreas de transição com a Caatinga, Amazônia e Mata Atlântica. Ao Sul apresente-se em São Paulo e Paraná, neste último associado às áreas dos Campos Gerais, sendo no país o extremo austral da sua distribuição. Ao Leste encontra-se bem representado no Mato Grosso e, de forma fragmentada, nos estados de Rondônia. Com respeito aos estados da região Norte do Brasil, o Cerrado muitas vezes aparece em forma de disjunções no Amazonas e Pará; assim como em Roraima e Amapá, embora estes últimos tenham uma associação mais forte com as formações savânicas da Venezuela e das Guianas, respetivamente, daí, possivelmente, a sua baixa relação florística com a flora da região core do Cerrado (RATTER; BRIDGEWATER; RIBEIRO, 2003).

Fora do Brasil, é reconhecida a presença de Cerrado na Bolívia e no Paraguai. Apesar de existir áreas contínuas desta vegetação, ambos os países apresentam disjunções desse bioma, o que poderiam ser tratados como formações relictuais, muito possivelmente derivados de épocas pretéritas onde o Cerrado teve uma maior expressividade geográfica (WERNECK, 2011; WERNECK et al., 2012). Na Bolívia, o Cerrado está presente no setor Nordeste, no departamento de Santa Cruz, onde importantes áreas de chapadões são encontrados no Parque Nacional Noel Kempff (KILLEEN; SCHULENBERG 1998), assim como em áreas de serranias na região de Chiquitos (SEGARRA, 2001), área que IBISCH et al. (2003) nomearam como Cerrado Chiquitano. Próximo ao Paraguai encontra-se o Cerrado Chaqueño, por estar em uma área de transição com o Chaco Seco. Existem também cerrado na ecorregião da Floresta Seca Chiquiatana, Pantanal e as Savanas de Beni (BECK, 2015; IBISCH et al., 2003).

No Paraguai, o Cerrado está presente na região Oriental e Ocidental, embora a superfície seja relativamente pequena quando comparado às outras ecorregiões do país (LÓPEZ et al., 2002). No extremo norte da região Oriental, há importantes áreas de Cerrado nos departamentos de Concepción, Amambay e Canendiyú (CÉSPEDES; MERELES 2006). Em direção sul desta região há presença de manchas de cerrado nos departamentos de San Pedro, Cordillera e Paraguarí (MERELES, 2005, 2013). Já na região Ocidental, o cerrado se limita ao extremo norte do departamento de Alto Paraguai, imediatamente debaixo da Bolívia, em uma área de transição com o Chaco Seco (MERELES, 2013). Locais como Cerro León, Cerro Cabrera y Chovoreca têm sido destacados pela sua importância para a conservação do Cerrado nesse país (MERELES et al., 2013).

Apesar de o Cerrado ser reconhecido fora do Brasil, a maior parte dos estudos em grande escala têm sido realizados dentro dos limites políticos deste país, desconsiderado a sua presença no países vizinhos (AGUIAR et al., 2016; AMARAL et al., 2017; DE OLIVEIRA et al., 2015; OLIVEIRA et al., 2009; RATTER; BRIDGEWATER; RIBEIRO, 2003; SIQUEIRA; PETERSON, 2003; STRASSBURG et al., 2017). Contar com toda a amplitude geográfica de um elemento de estudo, seja uma espécie, formação vegetal ou ecorregião, pode ser útil para fins de conservação, porque permite ter uma visão global do fenômeno de estudo e não enviesa resultados dentro de limites artificiais. Porém, combinar estes dados dentro de unidades geopolíticas são importantes, porque geralmente o planejamento e execução de ações de conservação são definidos dentro dos limites dos países e não nos limites ecológicos (DALLIMER; STRANGE, 2015; HUNTER; HUTCHINSON, 1994; KARK et al., 2015; RODRIGUES; GASTON, 2002).

A América Latina destaca-se por suas altas taxas de perda de cobertura natural, que deixaram ecossistemas altamente fragmentados e antropizados (ARMENTERAS et al., 2017; FEHLENBERG et al., 2017; LE POLAIN DE WAROUX et al., 2016; OVERBECK et al., 2015; STRASSBURG et al., 2017; TEJADA et al., 2016). Assim, por exemplo, as formação vegetais abertas (como estepes e savanas) são os tipos de habitat sob a maior ameaça, sendo a pressão do fogo e do pastejo os principais fatores de risco (JARVIS et al., 2010). No caso de formações florestais, a expansão da agricultura e das pastagens para gado são as principais atividades que levam ao desmatamento no continente (ARMENTERAS et al., 2017). No âmbito mais local, as atividades agrícolas e pecuárias são as principais responsáveis pela perda de habitat nas ecorregiões do Pantanal (JUNK; CUNHA, 2005), Florestas Secas Chiquitanas (SALAZAR et al., 2015), Savanas de Beni (REDO; AIDE; CLARK, 2012), Chaco (SALAZAR et al., 2015; VALLEJOS et al., 2015) e no Cerrado brasileiro (OVERBECK et al., 2015; STRASSBURG et al., 2017). No caso concreto do Cerrado, este tem menos de 20% de sua superfície preservada (STRASSBURG et al., 2017) e menos do 7,7% sob áreas protegidas (OLIVEIRA et al., 2017), o que levou ser uma das ecorregiões mais ameaçadas no continente.

Estimar a distribuição das espécies e compreender quais os fatores que as determinam, tem sido um dos temas centrais em ecologia desde séculos (GUISAN; ZIMMERMANN, 2000; PETERSON; SOBERÓN, 2012). No entanto, a distribuição geográfica da maior parte das espécies permanece ainda desconhecida (*Wallacean shortfall*; WHITTAKER et al., 2005). Com o avanço científico em áreas como a estatística, informática, sistemas de informação geográfica e a disponibilidade de grandes bancos de dados e software gratuitos permitiram que tais questões sejam respondidas com técnicas mais sofisticadas (FRANKLIN, 2009). Dentre os métodos, talvez os modelos de nicho ecológico e de distribuição de espécies (conhecido como ENM e SDM nas suas siglas em inglês; PETERSON; SOBERÓN, 2012) sejam os que tiveram maior desenvolvimento e uso nas últimas duas décadas (FRANKLIN, 2009; PETERSON et al., 2011).

Os ENM/SDM são técnicas amplamente usadas na ecologia vegetal. São comumente empregadas para orientar decisões de conservação (GUISAN et al., 2013), como predizer novas populações de espécie raras (WILLIAMS et al., 2009); avaliar o potencial de invasão de espécie (VÁCLAVÍK; MEENTEMEYER, 2009); auxiliar na determinação do status de conservação em listas vermelhas (SYFERT et al., 2014); informar práticas de conservação de taxas ameaçados (SOUSA-SILVA et al., 2014; WAN et al., 2016); ou estimar o efeito de mudanças climáticas na distribuição das espécies (PRITI et al., 2016). Também são usadas para indagar sobre aspectos teóricos, como descrever padrões macroecológicos (DUBUIS et al., 2011), regiões fitogeográficas (AMARAL et al., 2017), ou distribuições pretéritas sob abordagem paleobiológicas (SVENNING et al., 2011).

Estas técnicas são especialmente importantes para estudos transnacionais, como foi o caso desta tese, já que os países onde foi desenvolvido a pesquisa sofrem de grandes lacunas no conhecimento sobre a distribuição das espécies, além de cada país apresentar diferente esforço amostral e de acessibilidade de dados. Desta forma, espera-se que esta pesquisa contribua uma visão mais abrangente do Cerrado, que sirva como base para ações de conservação, ao mesmo tempo que incentive a pesquisa na área da biogeografia da conservação com base nas fronteiras ecológicas, mas levando em consideração a sua interação com as fronteiras geopolíticas.

## APRESENTAÇÃO DOS CAPITULOS

A tese foi organizada em três capítulos, redigidos em formato de artigo. O primeiro capítulo foi uma pesquisa metodológica, enquanto que o segundo e terceiro avaliaram diferentes aspectos: padrões de riqueza e raridade da flora do Cerrado, assim como as ameaças presentes e futuras.

**Capítulo 1**: Aqui, foi avaliado o efeito da incorporação de dados edáficos globais no desempenho de ENMs para espécies de plantas. Para isso, utilizou-se 125 espécies distribuídas ao longo da região neotropical das Américas para estudar o efeito do uso de dados edáficos e sua combinação com dados climáticos na predição de modelos construídos sob diferentes tipos de algoritmos. Além disso, relacionou-se os diferentes conjuntos de variáveis com certas características geográficas das espécies alvo (como a área de ocorrência, número e densidade de ocorrências) e diferentes algoritmos (este capítulo foi publicado na PlosOne https://doi.org/10.1371/journal.pone.0186025).

**Capítulo 2**: Este capítulo teve como intuito estudar, por meio de modelos de nicho ecológico, a flora do Cerrado de forma transfronteiriça com o fim de (i) prever os padrões de riqueza e raridade das espécies de plantas para Bolívia, Brasil e Paraguai, (ii) estudar a concentração de raridade no bioma do Cerrado nos âmbitos global e regional (iii) e avaliar e detectar áreas que apresentam oportunidades e conflitos para conservar a riqueza e a raridade das plantas do Cerrado nessas três nações.

**Capítulo 3**: Aqui buscou-se (i) estimar o efeito da mudança climática e do uso do solo na distribuição das espécies de plantas do Cerrado nos diferentes países onde ocorre, (ii) avaliar a eficiência da atual rede de áreas protegidas para salvaguardar espécies sob diferentes emissões de gases de efeito estufa e cenários de uso do solo, e (iii) estimar o grau de ameaça das espécies decorrente da eficiência de proteção e perda de habitat.

# RESUMO EXPANDIDO

## CAPÍTULO 1

## Using worldwide edaphic data to model plant species niches: an assessment at a continental extent

Os modelos de nicho ecológico e de distribuição de espécies (ENM e SDM, respectivamente) são amplamente utilizados em diferentes campos da ecologia vegetal. Apesar da ampla aplicação de ENM´s em estudos botânicos, os debates em torno de vários aspectos das ENM´s continuam até a data. Um dos mais importantes aspectos relaciona-se com os dados utilizados nos ENM´s. O desempenho dos ENM´s é sensível a várias características dos conjuntos de dados (como as variáveis ambientais e características dos pontos de ocorrências). Em contraste com as variáveis climáticas comumente utilizadas para ENM´s, que descrevem a variação ambiental em escalas regionais (também conhecidas como variáveis macroclimáticas), as variáveis edáficas variam em escalas locais e com grande complexidade. As condições edáficas desempenham um papel importante na determinação do nicho de espécies de plantas terrestres. Infelizmente, as variáveis edáficas ainda não são usadas como variáveis ambientais em ENM de plantas, que continuam a ser limitadas a aquelas de natureza climática. Uma razão para a não consideração das variáveis edáficas nos ENMs de plantas pode estar relacionada à extensão geográfica para a qual estes dados estão disponíveis. Essa disponibilidade é comumente restrita a certos países ou regiões (por exemplo, EUA, China ou União Europeia), enquanto em muitas outras áreas, como é o caso dos países da América Latina, esses dados simplesmente não estão disponíveis.

Neste capítulo foi abordado as seguintes questões principais: (i) Qual é a influência de distintas variáveis preditoras (por exemplo, climática vs. edáfica) em diferentes algoritmos empregados em modelos de nicho ecológico? (ii) Qual é a relação entre o desempenho de variáveis preditoras e as características geográficas das espécies e diferentes algoritmos?

Para avaliar os efeitos dos dados edáficos globais sobre ENM´s e sua relação com diferentes algoritmos, foi adotado um projeto experimental fatorial com dois fatores: Variável ambiental e Algoritmo, com três e quatro níveis, respectivamente, totalizando 12 combinações entre os níveis dos fatores. O primeiro fator (Variável ambiental), compreendeu três modelos diferentes baseados em diferentes conjuntos de variáveis (edáficos, climáticos ou ambos). Para o segundo fator (Algoritmo) empregou-se quatro tipos de técnicas, Máquina de Vetores Suporte, Modelos Lineares Generalizados, Modelos Aditivos Generalizados e Floresta Aleatória. Os 12 tratamentos foram aplicados a 125 espécies de plantas, assim, foram construídos 1500 modelos.

Para cada conjunto de variáveis foram realizadas uma análise de componentes principais (PCA), e posteriormente foram usadas o número de componentes principais que explicassem mais do 95% da variabilidade dos dados. Os modelos foram avaliados por uma validação cruzada com duas partições, onde a presença de cada espécie e suas respectivas pseudo-ausências foram divididas em conjuntos de teste e treino de 50-50%. Foi utilizado o *True Positive Rate* (TPR), *True Skill Statistic* (TSS) e *Area Under Curve* (AUC) como índices de acurácia. A complexidade dos diferentes padrões espaciais das previsões binárias foi avaliada usando o índice de forma (SI). Foi utilizado um ANOVA de medidas repetidas para testar o efeito dos diferentes conjuntos de variáveis e algoritmo com os valores dos parâmetros TPR, TSS, AUC e SI.

Foi demostrado as vantagens de usar dados edáficos globais como variáveis preditoras em modelos de nicho ecológicos para espécies de plantas. Mais especificamente, comprovou-se que os ENM´s construídos com variáveis climáticas combinadas com variáveis edáficas não afetaram negativamente o desempenho de ENM´s, inclusive melhoraram a precisão de alguns algoritmos. Isso aconteceu mesmo quando os ENM´s baseados apenas em variáveis edáficas não forneceram predições mais acuradas. A complexidade espacial dos padrões preditos pelos modelos foi diferente para cada conjunto de variável ambiental. As descobertas destacam a viabilidade e as vantagens de incluir dados globais do solo, juntamente com variáveis climáticas, em ENM´s para obter previsões precisas das distribuições de espécies de plantas.

A grande extensão geográfica, a variabilidade dos ambientes e as características geográficas das espécies consideradas, permitiram demonstrar que os dados edáficos globais adicionam informações úteis para a construção dos ENM´s das plantas. Isto é particularmente valioso para estudos de espécies que ocorrem em regiões onde informações mais detalhadas sobre propriedades do solo são pobres ou não existem.

## CAPÍTULO 2

### "Looking beyond borders: patterns of richness and rarity of the Cerrado flora and its transboundary conservation opportunities",

Conhecer a distribuição geográficos das espécies é um dos principais objetivos da biogeografia de conservação, devido a que eles são usados como dados básicos, por exemplo, para determinar as áreas prioritárias para conservação e o estado de conservação das espécies. Geralmente, tais avaliações são realizadas no nível do país, particularmente a definição de áreas protegidas, onde as decisões políticas e econômicas determinam o resultado dos planos de conservação. Curiosamente, o estado de conservação das espécies é comumente estabelecido em nível "global" (por exemplo, a lista vermelha da IUCN) ou "nacional". Neste contexto, a conservação de espécies e ecossistemas geograficamente distribuídos em vários países representa um importante desafio. A raridade é um traço-chave para determinar a vulnerabilidade das espécies aos impactos ambientais e, consequentemente, seu status de conservação. A raridade pode torna uma espécie propensa à extinção local porque os impactos ambientais geralmente são fenômenos espacialmente. A cobertura natural restante dentro de uma célula pode ser considerada um substituto da intensidade das atividades antrópicas desenvolvidas na paisagem dessa célula. A relação entre o valor de conservação e as atividades antrópicas atuais ou potenciais dentro de uma célula permite identificar cenários de oportunidades ou de conflitos para conservação.

O bioma Cerrado é um importante modelo de estudo devido à sua presença em três países diferentes (ou seja, Bolívia, Brasil e Paraguai). Uma descrição abrangente do Cerrado pode ajudar a sua conservação em níveis regionais e globais. De fato, sob um contexto regional para uma unidade geopolítica singular (ou seja, país), as espécies de Cerrado podem apresentar distribuições espaciais restritas relevantes para a biodiversidade e a conservação de uma nação. Desta forma foi estudado a flora do Cerrado de forma transfronteiriça para (i) prever padrões de riqueza e raridade de espécies na Bolívia, Brasil e Paraguai, (ii) estudar a concentração de espécies raras ao nível global e regional, e (iii) avaliar áreas que apresentam oportunidades e conflitos para conservar a flora do Cerrado nas três nações.

Foram modeladas 1559 espécies de plantas do Cerrado. Definimos cenários de oportunidade e conflito para a conservação com base na relação entre o grau de alteração da paisagem de uma célula e a riqueza e concentração da raridade global e local nela contida. A área de estudo abrangeu os países da Bolívia, Brasil e Paraguai. Incluímos aquelas ecorregiões relacionadas a formações vegetais abertas, como estepes e savanas, e outras que são formações fechadas, mas onde a existência do Cerrado é bem conhecida. Assim, para o Brasil, foi considerado as ecorregiões do Cerrado, Campos Rupestres, Pantanal e *Florestas* de *Babaçu do Maranhão*; para a Bolívia e Paraguai as ecorregiões de Savanas de Beni, Florestas Secas Chiquitanas, Chaco Seco e Chaco Úmido.

Os ENM´s foram construídos usando variáveis climáticas e edáficas. Foram empregados seis algoritmos, Modelos Lineares Generalizados, Modelos Aditivos Generalizados, Máxima Entropia, Floresta Aleatória, Máquina de Vetores Suporte e Processos Gaussianos. Foi utilizada duas abordagens para avaliar os ENM dependendo do número de ocorrências. Para as espécies com 5 a 19 ocorrências, foi empregado um procedimento de Jackknife. Para espécies com ≥ 20 registos, foi implementado uma validação em blocos com duas partições (como um tabuleiro de xadrez) para controlar a autocorrelação espacial entre os dados de treinamento e teste. O *True Skill Statistic* (TSS) foi usado como métrica do desempenho dos modelos.

A raridade foi baseada num índice que considerou o inverso do tamanho da distribuição e características do nicho das espécies como a marginalidade e a especialização. Foi assumido como limite crítico o valor de 0,5, ou seja, uma célula é considerada inviável para a presença de uma espécie tiver uma proporção de paisagem natural remanescente menor que esse valor. A oportunidade de conservação ou conflito de uma célula foi determinada com base na riqueza ou raridade média de toda a região de estudo ou para cada nação, que determinam a raridade global ou regional, respectivamente.

Foi encontrado que a relação entre as células com o maior distúrbio da paisagem e maior riqueza e raridade de espécies (condição de alto conflito) foi diferente para cada país. Por exemplo, o Brasil mostrou a maior proporção de sua terra em condições de alto conflito para a conservação, uma vez que as áreas mais antropizadas se sobrepõem com as regiões previstas com maior concentração de riqueza de espécies, bem como raridades globais e regionais. Por outro lado, foram encontradas grandes oportunidades de conservação na Bolívia e no Paraguai, dado que nesses países as áreas mais alteradas estavam nas regiões onde se predisseram menor riqueza e raridade. Além disso, foi encontrado que as áreas com maior concentração de riqueza de espécies, raridade global e regional concentram-se na parte central da ecorregião do Cerrado no Brasil. A Bolívia foi o segundo país mais rico em espécies, onde a maioria das espécies e raridades foram distribuídas nas ecorregiões do Cerrado e da Floresta Seca Chiquitana. O Paraguai foi o terceiro país mais rico, com o maior número de espécies e raridade principalmente nas ecorregião do Cerrado, na região norte do Chaco Seco e no extremo nordeste do Chaco Úmido. A raridade global e regional teve uma relação positiva com a riqueza de espécies em todos os países.

Na medida do nosso conhecimento, esta é a primeira tentativa de descrever a diversidade vegetal do bioma Cerrado além dos limites políticos. O Brasil abrange grande parte da flora do Cerrado e possui a maior proporção do bioma no seu território, contudo nesse país predomina o cenário de conflito onde a maior alteração da paisagem ocorre em áreas com maior diversidade. No entanto, foram encontradas grandes oportunidades de conservação na Bolívia e no Paraguai, porque a riqueza e a raridade de espécies se estendem em áreas com baixo grau de alteração.

## CAPÍTULO 3

## A dark scenario for Cerrado plant species, effects of future climate, land use and inefficiency of protected areas

A Terra experimentou múltiplas flutuações climáticas que ocorreram naturalmente ao longo dos tempos geológicos, porém, o clima foi recentemente alterado pelas atividades humanas. A mudança climática antropogênica é considerada um dos principais fatores que podem estar alterando a composição de comunidades, e o funcionamento dos ecossistemas na escala global. O aumento da demanda humana de alimentos e energia levou a que extensas áreas com cobertura natural sejam convertidas para sustentar atividades produtivas e, consequentemente, a maior parte da área terrestre do mundo está biologicamente comprometida. Os efeitos das mudanças climáticas e a perda da cobertura natural afetam não apenas a diversidade biológica, mas também o bem-estar humano. O fato de que as espécies e ecossistemas estão sujeitos a alterações induzidas pela mudança climática pode levar a que as áreas protegidas existentes sejam ineficientes na proteção da biodiversidade, além de que a demanda por espaço para futuros usos da terra poderia comprometer a criação de unidades de conservação. Neste contexto tivemos como objetivo (i) avaliar o efeito da mudança climática e do uso da terra na distribuição das espécies de plantas do Cerrado para os países onde ocorre, (ii) avaliar a eficiência da atual rede de áreas protegidas para salvaguardar espécies sob diferentes cenários de mudanças climáticas e uso da terra, e (iii) estimar o grau de ameaça das espécies baseado na eficiência das áreas protegidas e perda de habitat.

Foram selecionadas 1555 espécies de plantas do Cerrado. Os ENM´s foram construídos empregando seis algoritmos, Modelos Lineares Generalizados, Modelos Aditivos Generalizados, Máxima Entropia, Floresta Aleatória, Máquina de Vetores Suporte e Processos Gaussianos. As variáveis ambientais usadas para construir os modelos foram de tipo edáfica e climática. Foi utilizado seis dados do solo relacionados às propriedades físicas, assumindo que podem ser menos suscetíveis às mudanças climáticas do que aqueles relacionados a variáveis químicas. Os dados edáficos foram fornecidos pelo SoilGrids. Como dados climáticos foram utilizadas as 19 variáveis bioclimáticas para as condições atuais e futuras. As condições climáticas atuais foram obtidas pelo WorldClim v2.0. Ambos os bancos de dados somaram 49 variáveis ambientais. Foi realizado uma análise de componentes principais (PCA) nas variáveis ambientais originais com base em uma matriz de correlação. Selecionou-se nove componentes principais, como novos preditores, que, em conjunto, explicaram até 95,20% da variabilidade total das variáveis originais. Uma vez realizado o PCA para as condições atuais, suas componentes principais foram projetadas para a condição climática futura.

Foram empregadas as projeções climáticas do 5º relatório de avaliação do Painel Intergovernamental sobre Mudanças Climáticas como fonte de condições climáticas futuras. Foi avaliado dois cenários de emissão de gases de efeito estufa, RCP4.5 e RCP8.5, que foram assumidos como cenários otimistas e pessimistas, respectivamente. Foram utilizadas projeções para 2050 e 2080. Foram empregados seis Modelos de Circulação Geral (GCM`s) de 28 pré-selecionados.

Para avaliar o efeito da tendência do uso do solo, foi usado a cobertura para o ano 2015 e futuras estimadas até 2050 e 2080. Foram selecionados os modelos MESSAGE-GLOBIOM e EMIND-MAGPIE porque são consistentes com os cenários de emissões de gases de efeito estufa RCP4.5 e RCP8.5, respectivamente. Foi quantificada a perda de habitat usando as classes de uso do solo: culturas anuais $C_3$ e $C_4$, culturas perenes $C_3$ e $C_4$, culturas $C_3$ de fixação de nitrogênio, pastagens e urbanas. Foram empregadas duas abordagens para avaliar o desempenho do modelo. Assim, foi utilizado o procedimento Jackknife para as espécies que tinham dentre 5 a 15 pontos de presenças. Para espécies com ≥ 20 registros, foi implementado uma validação em bloco com duas partições. Foi usado o *True Skill Statistic* (TSS) como métrica da performance dos modelos.

Os modelos finais para a condição atual foram baseados na média aritmética da qualidade ambiental predita pelos melhores algoritmos de uma espécie, ou seja, aqueles modelos com desempenho superior ou igual ao TSS médio dos algoritmos. Os melhores foram projetados para condições ambientais futuras para cada GCM, com os quais foi construído um modelo de consenso entre os algoritmos para determinado GCM. A projeção futura baseou-se numa nova média de valores de qualidade ambiental entre os sete GCM´s. Deste modo, foram construídos 9.336 modelos (Espécies x Algoritmos), com 261.408 projeções (GCM´s x RCP´s x Períodos) que constituíram os 6.224 modelos finais (Espécies x RCP´s X Períodos).

O efeito das mudanças climáticas na distribuição das espécies foi acessado considerando um cenário de não dispersão, isto é, em condições futuras a espécie poderá permanecer somente naquelas áreas em que existe sobreposição entre a distribuição atual e futura. Não foi considerado o uso futuro da terra na rede de áreas protegidas assumindo um cenário onde essas áreas permaneceriam intactas e só seriam afetadas pela mudança climática.

Os resultados apontam que as mudanças climáticas e a expansão do uso da terra reduzirão de forma significativa a distribuição das espécies em 2050 e 2080. As regiões onde a riqueza atual e futura se prevê serem maiores se sobrepõem com as áreas mais antropizadas. A interação entre esses dois fatores (clima e uso do solo) pode causar perdas biológicas substanciais em cada país, comprometendo seriamente os esforços de conservação nacionais e territoriais. Atualmente, a rede de áreas protegidas não é eficiente para proteger as espécies, nem para condições presentes ou futuras, devido à perda de áreas ambientalmente adequadas dentro das unidades de conservação. A nível nacional, os países apresentaram diferentes eficiências de sua rede de áreas protegidas. No entanto, as projeções mostram que os países, que são promissores para a conservação do Cerrado nas condições atuais, podem ser seriamente afetados no futuro. Este baixo grau de proteção e sua susceptibilidade diante as mudanças climáticas, juntamente com um uso da terra, levaria que as espécies sejam seriamente ameaçadas, ainda sob o cenário mais otimista.

# CAPÍTULO 1

## USING WORLDWIDE EDAPHIC DATA TO MODEL PLANT SPECIES NICHES: AN ASSESSMENT AT A CONTINENTAL EXTENT

**Abstract:** Ecological niche modeling (ENM) is a broadly used tool in different fields of plant ecology. Despite the importance of edaphic conditions in determining the niche of terrestrial plant species, edaphic data have rarely been included in ENMs of plant species perhaps because such data are not available for many regions. Recently, edaphic data has been made available at a global scale allowing its potential inclusion and evaluation on ENM performance for plant species. Here, we take advantage of such data and address the following main questions: What is the influence of distinct predictor variables (e.g. climatic vs edaphic) on different ENM algorithms? and what is the relationship between the performance of different predictors and geographic characteristics of species? We used 125 plant species distributed over the Neotropical region to explore the effect on ENMs of using edaphic data available from the SoilGrids database and its combination with climatic data from the CHELSA database. In addition, we related these different predictor variables to geographic characteristics of the target species and different ENM algorithms. The use of different predictors (climatic, edaphic, and both) significantly affected model performance and spatial complexity of the predictions. We showed that the use of global edaphic plus climatic variables generates ENMs with similar or better accuracy compared to those constructed only with climate variables. Moreover, the performance of models considering these different predictors, separately or jointly, was related to geographic properties of species records, such as number and distribution range. The large geographic extent, the variability of environments and the different species' geographical characteristics considered here allowed us to demonstrate that global edaphic data adds useful information for plant ENMs. This is particularly valuable for studies of species that are distributed in regions where more detailed information on soil properties is poor or does not even exist.

**Keywords**: accuracy, algorithms, climate data, geographical extent, number of records, soil data, species distribution modelling

# 1. INTRODUCTION

Ecological niche and species distribution modeling (ENM and SDM, respectively) are widely-used tool in different fields of plant ecology, including the prediction of new populations of rare species (WILLIAMS et al., 2009); or potential distribution of invasive species (VÁCLAVÍK; MEENTEMEYER, 2009); informing conservation practices for threatened taxa (SOUSA-SILVA et al., 2014; WAN et al., 2016); estimating the effect of climatic change on species distributions (PRITI et al., 2016; STILL et al., 2015); describing macroecological patterns (DUBUIS et al., 2011) and studying past species distributions under a paleobiological approach (SVENNING et al., 2011). Despite the broad application of ENMs in botanical studies, debates surrounding several aspects of ENMs continue to date. One of the most important of these aspects relates to the actual data used in ENMs (AUSTIN; VAN NIEL, 2011).

Data used for conducting ENMs can be grouped in two sets: biogeographical data about the distribution (or presence/absence) of species (i.e. occurrence records) and environmental data (i.e. predictor variables) used to predict those distributions (FRANKLIN, 2009). ENM performance is sensitive to several characteristics of these two datasets (BEALE; LENNON, 2012; BEAUREGARD; BLOIS, 2014; BECK et al., 2014; DUPIN et al., 2011; FERNÁNDEZ; HAMILTON; KUEPPERS, 2013; HARRIS et al., 2013; HORTAL et al., 2008; JIMÉNEZ-VALVERDE; LOBO; HORTAL, 2009; LUOTO et al., 2005; MCPHERSON; JETZ; ROGERS, 2004; NEWBOLD, 2010). For example, regarding biogeographical data, ENMs can be affected by different aspects of the species' distributional patterns and their sampling such as: prevalence (considered here as the ratio between the quantity of presence and absences), range size and spatial autocorrelation (LUOTO et al., 2005; MCPHERSON; JETZ; ROGERS, 2004; TSOAR et al., 2007); which in turn are related to the available sample size (JIMÉNEZ-VALVERDE; LOBO; HORTAL, 2009), data biases along road networks or cities (BECK et al., 2014; MEYER et al., 2015), geographical accuracy (NEWBOLD, 2010) and environmental variability captured by the records (HORTAL et al., 2008). All of these aspects can interact with the environmental data selected to fit the ENMs, and affect model accuracy (BEAUREGARD; BLOIS, 2014; DUBUIS et al., 2013) and resultant suitability (HARRIS et al., 2013). Even if occurrence data were bias-free, environmental data can still severely affect ENM performance, especially if inappropriate environmental variables are used as predictors (MOD et al., 2016).

Choosing a particular environmental variable for ENM depends on the modeling purpose and its biological significance to the species under study (AUSTIN, 2007). Obviously, different species may have particular constraints related to their dependency on environmental factors and no single variable is expected to be equally meaningful for all species. For instance, variables related to soil properties are considered to be particularly important in determining the distribution of plant species, but have little direct effect on the distribution of the majority of animal species (AUSTIN, 2002). Considering this plant-soil relationship, predictors can be grouped, following (AUSTIN; SMITH, 1989), in: (i) resource, matter and energy consumed by an organism, such as oxygen, water, macronutrients and micronutrients; (ii) variables that have direct physiological importance, such as pH, cation exchange capacity, aluminum concentration, hydromorphic condition; and (iii) indirect variables that do not have important physiological effects, such as porosity, bulk density, texture (clay, silt and sand fraction) and soil depth.

In contrast with climatic variables commonly used for ENMs, which describe environmental variation at regional scales (a.k.a. “macroclimatic” variables; e.g. CHELSA; KARGER et al., 2016), edaphic variables vary at local scales and with great complexity (HEUVELINK; WEBSTER, 2001). For example, within the same landscape, climatic conditions can be very homogenous throughout while soil properties can vary widely according to different parental material (ANDERSON, 1988), topographic position (CEDDIA et al., 2009) or land-use (MWANJALOLO JACKSON-GILBERT et al., 2015). Indeed, there are several examples in the literature where soil properties control the distribution of plant species or the structure, composition, and physiognomy of a community within an otherwise climatically homogeneous geographical extent. For instance, mangrove distribution is strongly influenced by soil properties such as salinity, acidity, hydromorphy and nutrient supply (REEF; FELLER; LOVELOCK, 2010). Swamp forests, like the *Caxeitais* (dominated by *Tabebuia cassinoides* (Lam.) DC.) of the Brazilian coast, are mainly distributed over organic and hydromorphic soils (RACHWAL; CURCIO, 2001).

The halophyte vegetation from Chile and Europe is restricted to continental salines (MELEČKOVÁ et al., 2014; TEILLIER; BECERRA, 2003). Furthermore, soil scarcity can also determine natural plant formations such as those inhabiting rocky outcrops (GRÖGER; HUBER, 2007). Narrow plant endemics are also frequently associated with specific types of soil, rock, and bedrock (BÁRCENAS-ARGÜELLO; GUTIÉRREZ-CASTORENA; TERRAZAS, 2013). Even certain soil nutrients can determine the distributional transition from one vegetation type to another, such as that between Neotropical seasonal forests and savannas where the concentration of aluminum or potassium define the structure of these vegetation types (LLOYD et al., 2015; RUGGIERO et al., 2002).

Consequently, it is clear that edaphic conditions play an important role in determining the niche of terrestrial plant species (MOD et al., 2016; THUILLER, 2013). Accordingly, several studies have tested the effect of including edaphic variables in ENMs for plant species such as the importance of soil nutritional variables for predicting plant distribution (COUDUN et al., 2006); the improvement of plant ENMs performance when using physical and chemical soil data (BEAUREGARD; BLOIS, 2014; DUBUIS et al., 2013); and the effect of both landscape and edaphic data in predicting future plant distributions under climate change scenarios (AUSTIN; VAN NIEL, 2011; BERTRAND; PEREZ; GÉGOUT, 2012). All of these studies reinforced the idea that plant ENMs could be improved by using a single or a group of edaphic variables. Unfortunately, edaphic variables are still not frequently used as predictors in plant ENMs, which continue to be limited to climatic variables (THUILLER, 2013). One reason for this lack of consideration of edaphic variables in plant ENMs may be related to the geographical extent for which these data are available. Such availability has been usually restricted to certain countries or regions (e.g. USA, China or the European Union), whereas in many other regions, as in many Latin America countries, these data are simply not available. Recently, however, the ISRIC World Soil Information with the SoilGrids database has provided data related to physical, chemical and taxonomical characteristics of soils across the globe (HENGL et al., 2014). Therefore, this database allows the construction of ENMs for plant species inhabiting large regions of the world or species occurring in countries that differ in the quantity and quality of the available edaphic data.

Indeed, despite including detailed soil data, most plant ENM studies have been conducted on extents that are usually smaller than the complete geographic distribution of plant species. Such ENMs may not comprise the full environmental variability that characterizes a species distribution and thus may affect model performance (CARRETERO; SILLERO, 2016; SÁNCHEZ-FERNÁNDEZ; LOBO; HERNÁNDEZ-MANRIQUE, 2010). Here, we evaluate the potential effect of using the SoilGrids global dataset in improving ENMs for plant species. We used 125 species distributed along the Neotropical region of the Americas, where many countries do not have detailed soil data, to explore the effect of using global edaphic data and its combination with climatic data in the prediction of models constructed under commonly used ENM algorithms. In addition, we related the different variable sets to certain geographical characteristics of target species (e.g. occurrence area, number of records and density of records) and different algorithms.

## 2. METHODS

### 2.1 OVERVIEW

To evaluate the effect of adding global edaphic data into ENMs and its relationship with different modeling algorithms, we adopted a factorial experimental design with two factors: Predictor and Algorithm, with three and four levels respectively, totalizing 12 combinations of factor levels. The first factor, Predictor, comprised three different models based on different predictor sets (edaphic, climatic, or both). For the second factor, Algorithm, we used four types of ENM algorithms (FIGURE 1). The 12 treatments were applied to 125 plant species, our experimental units, thus 1500 models were fitted (see below).

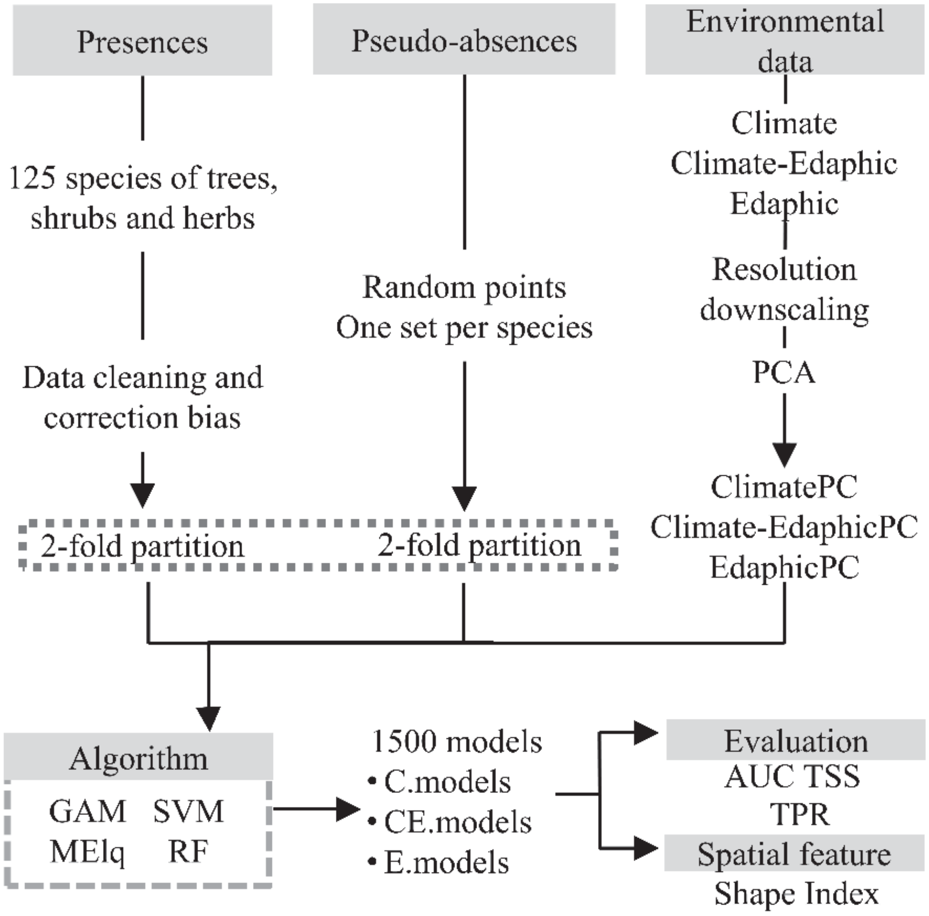


FIGURE 1 - Experimental design for testing the effect of using edaphic variables in ENMs for plant species
SOURCE: the author (2018)

## 2.2 STUDY AREA

Our study area extended from the south of the United States of America to the austral extremes of Chile and Argentina. This area covers a wide variety of climatic conditions, geological formations, and soil types, but many of its constituent countries lack edaphic data. Consequently, our selected plant species (see below) occur in different biomes, from arid regions such as the *Chihuahuan* and *Caatinga* steppe and warm-humid biomes such as the Brazilian Atlantic and *Chaco-Darién* moist forests to the cold regions of the *Nothofagus* forests and *Andean páramos* (see APPENDIX 2 for the complete species list).

## 2.3 ENVIRONMENTAL DATA

We used three sets of environmental variables for building ENMs: climate-only, edaphic-only and both climatic and edaphic variables together (BEAUREGARD; BLOIS, 2014; COUDUN et al., 2006), hereafter called C.models, E.models and CE.models, respectively. Note that all of these predictors were continuous variables. For the C.models, we employed the 19 bioclimatic variables from the recently developed CHELSA v1.1 online database (KARGER et al., 2016). These variables were built based on monthly averages of climate data, mainly temperature and precipitation as collected from meteorological stations, for the 1979-2013 period and interpolated to the global surface (KARGER et al., 2016). E.models were built with 56 variables related to physical and chemical soil properties obtained from the SoilGrids database available from ISRIC-World Soil Information (HENGL et al., 2014), the data were downloaded in June of 2015 (TABLE 1). The SoilGrids database provides global maps of soil classes and some edaphic variables (TABLE 1). In addition, this database has an automated updating system that progressively increases its accuracy when new input data becomes available in the international soil profile databases (HENGL et al., 2014). The CE.models were built combining the climate and edaphic datasets, summing up to 75 variables. Both climatic and edaphic datasets were acquired with a spatial resolution of 30 arc-seconds (≈ 1 $km^2$ cell size) and upscaled to 5 arc-minutes (≈ 10 $km^2$ cell size). This upscaling (resolution change) was based on the aggregation, by taking their average value, of lower resolution cells into higher resolution cells. Later, these datasets were cropped to the extent of the study region ranging from -120° to -30° in longitude and -60° to 35° in latitude.

Different modeling approaches present different sensitivity to collinearity of predictor variables (DORMANN et al., 2013). However, no single methodological procedure has been considered ideal for solving or handling collinearity (DORMANN et al., 2013). Here, we opted to conduct a principal component analysis (PCA) on the original environmental dataset and use the scores of each derived principal components (PCs) as new predictors variables (CRUZ-CÁRDENAS et al., 2014; DUPIN et al., 2011). The PCA is a multivariate technique that produces uncorrelated components from the original data sorted according to the amount of total variance that it explains. We selected a number of PCs that explained more than 95% of the total variance in the original dataset (SÁNCHEZ-FERNÁNDEZ; LOBO; HERNÁNDEZ-MANRIQUE, 2010). The major advantages of this procedure are the correction of multicollinearity among the original variables, the use of almost all information contained in a large dataset that is captured in the PCs, and the reduction of the number of variables used in the models. Accordingly, C.models and E.models were built with the first six PCs and the first 11 for the CE.models (see APPENDIX 3 and 4 for more information about variance explained and variables' coefficients for the selected PCs). The reduced number of new variables (PCs) reveals the high collinearity in our original variable set. In fact, the first two PCs of the PCAs conducted for each variable set explained more than 50% of the variance (in the APPENDIX 1 the relationships of original variables and the first two PCs of each variable set are depicted).

TABLE 1 - Climate and edaphic variables (names and units) used as predictors in plant ecological niche models

| Climate (Source: CHELSA) | Unit | Edaphic (Source: SoilGrids) | Unit |
|---|---|---|---|
| Annual Mean Temperature | °C | Depth to bedrock up to maximum 240 cm | cm |
| Mean Diurnal Range | °C | Predicted probability of occurrence of R horizon | % |
| Isothermality | °C | Mean of bulk density* | kg/m3 |
| Temperature Seasonality | °C | Mean of coarse fragments volumetric* | % |
| Max temperature of warmest week | °C | Mean of soil texture fraction clay* | % |
| Min temperature of coldest week | °C | Mean of soil texture fraction silt* | % |
| Temperature annual range | °C | Mean of soil texture fraction sand* | % |
| Mean temperature of wettest quarter | °C | Mean of cation exchange capacity* | cmolc/kg |
| Mean temperature of driest quarter | °C | Mean of soil organic carbon stock* | Tn/ha |
| Mean temperature of warmest quarter | °C | Mean of soil organic carbon content* | ‰ |
| Mean temperature of coldest quarter | °C | Mean of soil pH in H2O* | |
| Annual precipitation | Mm | | |
| Precipitation of wettest week | Mm | | |
| Precipitation of driest week | Mm | | |
| Precipitation seasonality | C of V | | |

*Data for six depths. SOURCE: the author (2018)

## 2.4 PLANT SPECIES DATA AND CLEANING

We selected 125 terrestrial plant species distributed within the Neotropical region with the purpose of considering the wide variety of environmental conditions in our study region. Data for these taxa was restricted to the species level, thus infraspecific taxa were not considered. Our final species dataset comprised trees (82), shrubs (27), herbs (8) and palm (8) species. We considered only species with more than 20 checked records (described below; see APPENDIX 2). This dataset comprised species inhabiting extreme latitudes such as *Atriplex canescens*, *Prosopis glandulosa* or *Parthenium incanum* in the north, and *Nothofagus antarctica*, *N. pumillo*, and *Mulguraea tridens* in the south. These species also differ in regard to their geographic range sizes, from those with narrow distributions such as *Juglans australis* to those considered as cosmopolites such as *Trema micranta*, *Ipomoea carnea* and *Inga vera.*

We conducted a taxonomic revision for these taxa verifying their accepted names and synonymy using The Plant List Version 1.1.[1] and Tropicos[2], checked by the Taxonomic Name Resolution Service v3.2 (BOYLE et al., 2013) based on APG III (ANAGIOSPERM PHYLOGENY GROUP, 2009). After confirming accepted names and synonymy, we used these names to search occurrence records for these species in the Global Biodiversity Information Facility[3] and the speciesLink database[4].

Occurrence records available in those databases may contain some taxonomic and geographic coordinate errors (GOODWIN et al., 2015). Our first step for data cleaning was the elimination of all records allocated outside the study area and those with repeated geographic coordinates. We also removed those species' records corresponding to invasive or cultivated distributions, thus leaving only those records that pertain to the natural distribution of species. This last step was

---

[1] http://www.theplantlist.org/
[2] http://www.tropicos.org
[3] http://www.gbif.org/
[4] http://splink.cria.org.br/

conducted by using information about species distributions available in the Catalogue of Life[5], Flora del Conosur[6], List of Species of the Brazilian Flora[7], Smithsonian Tropical Research Institute-Scientific Databases[8], PLANTS Database[9] and Tropicos national species list from Bolivia, Panamá, Paraguay, Peru and Ecuador[10]. In order to clean records temporally, we only considered records that were collected between 1979 and 2013, thus corresponding to the temporal span of our climate variables.

It is common that species' occurrence records are biased towards roads, cities or countries (MCCARTHY et al., 2012; MEYER et al., 2015; REDDY; DÁVALOS, 2003). Therefore, these records are not the result of random and homogeneous sampling along the geographic distribution of a species, compromising the accuracy of ENMs (BECK et al., 2014). We used a systematic sampling given its suggested effectiveness to correct geographic bias (FOURCADE et al., 2014) by creating a grid with a resolution of 10 arc min (≈ 20 $km^2$ cell size) and then selecting one occurrence per cell. The number of cleaned records for species ranged from 20 to 1227 (see APPENDIX 2).

## 2.5 MODELING PROCEDURES

The diverse algorithms usually employed to build ENMs have different input requirements (PETERSON et al., 2011), degrees of complexity (RANGEL; LOYOLA, 2012), stability (DUAN et al., 2014) and predictive abilities (ELITH et al., 2006; TSOAR et al., 2007). For these reasons, we also explored how different algorithms respond to distinct sets of environmental variables. We used four methods which are commonly used in ENM and highlighted for their performance: Generalized Additive Models (GAM), Maximum Entropy (ME), Random Forest (RF) and Support Vector Machine (SVM).

---

[5] http://www.catalogueoflife.org/

[6] www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm

[7] http://floradobrasil.jbrj.gov.br/

[8] http://biogeodb.stri.si.edu/bioinformatics/en/

[9] http://plants.usda.gov/java/

[10] http://www.tropicos.org/

GAM is a non-parametric extension of GLM (Generalized Linear Model) that replaces the linear relationship between the dependent and independent variables by the sum of a smooth function (HASTIE; TIBSHIRANI, 1986). Owing to its combination of a link function and a smooth function, the GAM method has the ability to deal with highly non-linear and non-monotonic relationships between the response and explanatory variables (GUISAN; EDWARDS; HASTIE, 2002). These GAMs were fitted using a binomial distribution with all single predictor variables, i.e. without backward or forward selection and interaction. The Newton method was used to optimize the estimation of the smoothing parameter.

ME is a machine learning method based on the principle of maximum entropy (PHILLIPS; ANDERSON; SCHAPIRE, 2006; PHILLIPS; DUDÍK; SCHAPIRE, 2004). This principle is based on minimizing the relative entropy between two probability densities defined in feature space (ELITH et al., 2011). This is a high performance technique (ELITH et al., 2006) and is less sensitive to spatial errors than others algorithms (GRAHAM et al., 2007). This method can be tuned with different features such as linear, quadratic, product, threshold, hinge and binary; the default use of all these features can cause overfitting and affect the models performance (ANDERSON; GONZALEZ, 2011). Thus we used linear and quadratic features (SOUZA; DE MARCO, 2014, hereafter MElq), both of these constrain the approximation of the probability distribution in a way that the variables' mean and variance should be close to its observed values (PHILLIPS; ANDERSON; SCHAPIRE, 2006). We also used 1000 maximum iterations, default regularization values, logistic output format and 10000 maximum background points.

SVM uses linear models to find a decision function, which is a hyperplane determined by non-linear decision boundaries that split samples in different classes within a higher-dimensional space (KAMATH, 2009; SALCEDO-SANZ et al., 2014). The optimal hyperplane is the one that maximizes the buffer between the boundary (i.e. support vectors) and the data (KUHN; JOHNSON, 2013). Mapping of the input data in a high-dimensional feature space is defined by a kernel function (HORNIK; MEYER; KARATZOGLOU, 2006). These models were built based on probability classes, performed with a radial basis kernel (RBF) and with a constant cost value ($C$=1).

RF comprises a family of algorithms that perform classification and regression analyses. RF is a modification of bagging trees, which build a model based on the average of a large collection of non-correlated trees (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). In each node of these trees, a random sample of *m* predictors is chosen as split candidates from the full set of predictors (JAMES et al., 2013). These algorithms have the advantage of not overfitting the data (BREIMAN, 2001) and use the out-of-bag (OOB) sample to construct different variable importance measures (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). To determine the optimal number of variables randomly sampled at each split the RF algorithm was tuned automatically. 500 trees were used at the tuning step, with default values of the step factor and the improvement in OOB error parameter. We considered those models with the minimum OOB error as our final RF models.

Given that we did not have real absences of our species, we created pseudo-absences to fit GAM, SVM and RF models. The prevalence and the method of pseudo-absence allocation can affect ENM performance, which can vary for distinct algorithms (BARBET-MASSIN et al., 2012; HANBERRY; HE; PALIK, 2012; LOBO; JIMÉNEZ-VALVERDE; HORTAL, 2010). To reduce potential noise, we used a prevalence of 1, thus the number of pseudo-absences for each species was equal to its presences. These pseudo-absences were allocated across the study area, which constitutes the biogeographic domain that the modeled species could have used as an accessible area over relevant periods of time (BARVE et al., 2011; SOBERÓN; PETERSON, 2005). We used one soil layer as a raster mask for creating the pseudo-absences given that some cells with climate data may have no soil data (i.e. “empty cells”), such as lakes and some mountain regions.

## 2.6 MODEL EVALUATION

Models were evaluated by a 2-fold cross-validation where the presences of each species and its respective pseudo-absences were partitioned into 50-50% training-testing sets. To control for spatial autocorrelation between training and testing records, we used a checkerboard partitioning method similar to

(MUSCARELLA et al., 2014). This method generates checkerboard grids that partition the records into bins by subdividing the geographic extent equally. For this, a particular grid resolution (i.e. cell size) must be chosen a priori, which does not guarantee a balanced number of records in each bin (MUSCARELLA et al., 2014). Therefore, we adapted the method to select the grid resolution that optimizes representation and balance of records within bins. To do so, we created 30 grids with resolutions varying from 0.5 to 15 degrees, with a gradual increase of 0.5. The optimum grid resolution was the one which (i) represented both training and testing records and (ii) minimized the difference between the number of training and testing records. Finally, to maintain a prevalence of 1, we randomly allocated pseudo-absences within each partition group.

Model performance was assessed by dependent and independent threshold metrics (LIU; WHITE; NEWELL, 2011). We used the True Positive Rate (TPR) and the True Skill Statistic (TSS, ALLOUCHE; TSOAR; KADMON, 2006) as threshold-dependent indices and the Receiver Operating Curve (AUC) as a threshold-independent evaluation. The threshold was the value that maximized the sum of sensitivity and specificity that produced the most accurate predictions (JIMÉNEZ-VALVERDE; LOBO, 2007). The complexity of the different spatial patterns of binary predictions (ENM outputs) was evaluated using the shape index (SI). This index measures the complexity of the predicted patches of pixels (i.e. potentially suitable cells) by considering the relationship between the sums of each patch perimeter ($p_i$) divided by the square root of patch area ($a_i$), $\mathrm{SI} = \sum_1^i \left(0.25\, p_i / \sqrt{a_i}\right)$ (MCGARIGAL; CUSHMAN; ENE, 2012).

## 2.7 DATA ANALYSIS

We used Repeated Measures ANOVAs to test the effect of the Predictor (C.models, E.models and CE.models), Algorithm (e.g. GAM, SVM, etc.) and their interaction on TPR, TSS, AUC and SI indices. We assumed that the Predictor and Algorithm as within-subject factors. To perform this analysis correctly and avoid a high Type II error rate, it was necessary that the data met the sphericity condition: the variances of the differences between combinations of levels do not differ. We used Mauchly's Sphericity Test at 95% confidence to validate the sphericity condition of the covariance matrix. When this condition was rejected, the degrees of freedom were corrected by the Greenhouse-Geisser method and used Type III sums of squares. We performed a post-hoc test using linear contrasts based on linear mixed effect models, considering the Predictor and Algorithm as fixed factors and the species as random factor. These models were used to perform pairwise comparisons of means between different predictors for a single algorithm at 95% confidence level. The p-values were corrected using the false discovery rate procedure.

After evaluating the models, their predicted suitabilities were projected onto the geographical space. For each species, we conducted pair-wise comparisons between the suitabilities of different kinds of models and algorithms by calculating the Kendall rank correlation coefficient ($\tau$) with cells of the entire study area. Values of this coefficient range from -1 (perfect disagreement) to 1 (perfect agreement), with values near zero representing independence between the compared ranks.

We used Pearson correlation ($r$) to explore the relationship between variation captured by records for different predictor sets and species' geographic characteristics, which were, for each species: (i) geographical extent, based on the number of cells within a minimum convex polygon comprising all of a species' records; (ii) number of records and (iii) density of records, which is the ratio between a species' number of records and its geographical extent. For each ENM algorithm, we explored the effect of such species' characteristics and predictors on TSS (LUOTO et al., 2005) by fitting linear mixed-effect models. These characteristics were considered as fixed effects within the mixed-effect models, along with the models with different predictors (C.models, E.models and CE.models), whereas the species were considered as random effects. TSS values were arcsine transformed. We used the variance inflation factor (VIF) to test for collinearity among predictors (species geographic characteristics), their significances were determined by a likelihood ratio test.

Construction of ENMs and statistical analyses were conducted in the R environment v. 3.3.2 (R CORE TEAM, 2017). The *dismo* v. 1.1.1 package (HIJMANS et al., 2016) was used to create pseudo-absences, model prediction and validation, and to fit MElq using Maxent v. 3.3.3. The GAMs, SVMs and RF models were fitted using the *gbm* v. 2.1.1 (RIDGEWAY, 2015), *kernlab* v. 0.9.25 (KARATZOGLOU et al., 2004) and *randomForest* v. 4.6.12 (LIAW; WIENER, 2002) packages, respectively. We used the packages *raster* v. 2.5.8 (HIJMANS, 2015), *SDMTools* v. 1.1.221 (VANDERWAL et al., 2014), and *pcaPP* v. 1.9.61 (PETER FILZMOSER; FRITZ; KALCHER, 2014) to handle raster, calculate the shape index, and the Kendall rank correlation coefficient. To fit the linear mixed effect models, repeated measures ANOVAs and the pairwise mean contrasts, we used the packages *nlme* v. 3.1.128 (PINHEIRO et al., 2016), *lsmeans* 2.26.3 (LENTH, 2016) and *car* v. 2.1-5 (FOX; WEISBERG, 2011), respectively.

## 3. RESULTS

The use of different predictors (climatic, edaphic, and both) significantly affected model performance, as measured by the TSS, TPR and AUC indices. They also affected the spatial complexity of the geographic predictions (SI). Moreover, TSS, TPR, AUC and SI showed different responses regarding the use of the distinct ENM algorithms. The interactions between predictors and algorithms were significant for TSS, AUC and SI (TABLE 2).

TABLE 2 - Results of the repeated measures ANOVA for the TSS, TPR, AUC and SI, considering the algorithm (GAM, MEql, SVM and RF) and predictor (climate, climate-edaphic, edaphic) factors

| Index | Factors | Sum of Squares | Df | Mean Square | F |
|---|---|---|---|---|---|
| TSS | Algorithm | 0.577 | 2.501 | 0.294 | 79.429*** |
| | Predictor | 1.344 | 1.463 | 0.858 | 76.316*** |
| | Algorithm * Predictor | 0.027 | 4.220 | 0.006 | 4.657*** |
| TPR | Algorithm | 0.043 | 2.588 | 0.021 | 13.024*** |
| | Predictor | 0.242 | 1.560 | 0.184 | 44.654*** |
| | Algorithm * Predictor | 0.008 | 5.119 | 0.001 | 1.505$^{ns}$ |
| AUC | Algorithm | 0.261 | 2.035 | 0.107 | 76.929*** |
| | Predictor | 0.379 | 1.457 | 0.249 | 70.262*** |
| | Algorithm * Predictor | 0.014 | 3.541 | 0.004 | 6.013*** |
| SI | Algorithm | 17965.610 | 1.705 | 10534.815 | 176.159*** |
| | Predictor | 77863.060 | 1.487 | 52370.852 | 690.234*** |
| | Algorithm * Predictor | 2286.069 | 4.684 | 488.070 | 47.693*** |

Degrees of freedom were corrected using Greenhouse-Geisser estimate of sphericity. TSS: true skill statistic; TPR: true positive rate, AUC: area under curve; SI: shape index. Significance: *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, $^{ns}$ $P > 0.05$
SOURCE: the author (2018)

According to TSS, TPR, and AUC, C.models and CE.models performed better than E.models, regardless of the ENM algorithm used. Nonetheless, MEIq performed better for the CE.models, regarding TSS, whereas SVM, GAM and RF did not show differences between C.models and CE.models. These results were different for the sensitivity, given that the CE.models showed the best values for SVM. Moreover, no algorithm differed significantly regarding only the C.models (FIGURE 2). Regardless of the predictor set, the SVM and RF algorithms had the highest values of TSS and AUC, followed by MEIq and GAM. Regarding the spatial complexity of predictions, C.models showed the most aggregated and continuous prediction, whereas the E.models had the most spread and complex patterns. The CE.models had an intermediate shape complexity. Independent of the predictor set, RF created the most complex spatial patterns, whereas SVM showed the lowest SI (FIGURE 2).

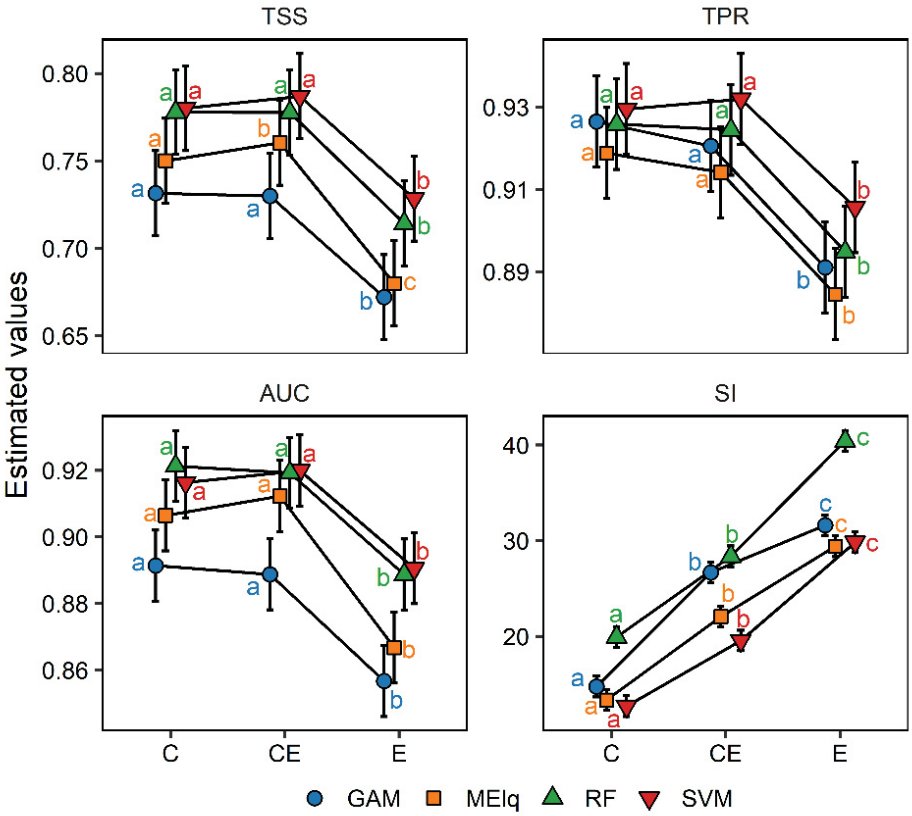


FIGURE 2 - Estimated values and confidence interval (bars) for TSS, AUC, TPR and SI of models fitted with three set of predictors combined with four algorithms. Means with same letter for different predictor and same algorithm denote significant difference using the linear contrast (P < 0.05). TSS: true skill statistic, TPR: true positive rate, AUC: area under curve; SI: shape index, C: models with climate predictors, CE: models with climate and edaphic predictors, E: models with edaphic predictors

SOURCE: the author (2018)

Mean values of Kendal rank correlation of suitabilities always showed positive values for the pair-wise comparison of models with different predictors (FIGURE 3A). The highest values of suitability correlation were for C.models-CE.models and E.models-CE.models for all algorithms. MEIq had the most similar suitability for these paired comparisons. The lowest correlation was between the suitability of C.models and E.models, with mean values smaller than 0.4 for all algorithms. For the C.models-E.models comparison, the highest correlations were for MEIq and RF, whereas in the C.models-CE.models comparison, highest correlation was for MEIq and GAM, and for the E.models-CE.models comparison highest correlations were for MEIq, RF and SVM (FIGURE 3A). These comparisons of suitability between algorithms showed that GAM-MEIq were the most similar, followed by RF-SVM, for any predictor set. The lowest correlation was found between GAM-SVM and between MElq-SVM (FIGURE 3B).

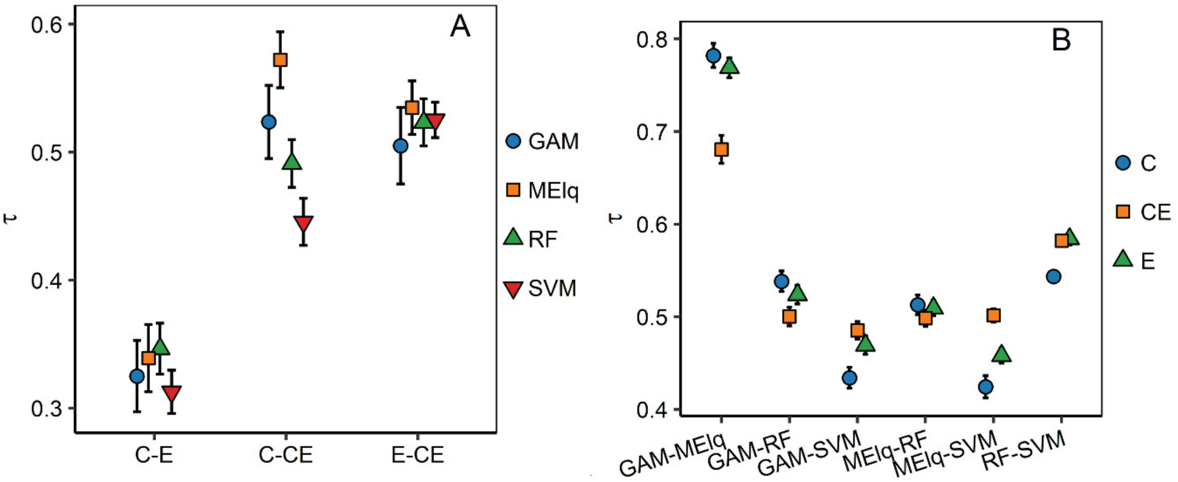


FIGURE 3 - Mean and confidence interval of Kendall rank correlation coefficient ($\tau$) of pair comparison of suitability. (A) Suitability comparison between models with predictors sets for different algorithm. (B) Suitability comparison between algorithms for different predictors sets. C: models with climate predictors, CE: models with climate and edaphic predictors, E: models with edaphic predictors
SOURCE: the author (2018)

The relationship between different geographic characteristics of species revealed that species with wider distributions had also more records sampled ($r$ = 0.790, $p < 0.001$) but they showed lower record density ($r$ = -0.640, $p < 0.001$). However, the relationship between number of records and their density was weak ($r$ = -0.090, $p$ < 0. 334; APPENDIX 5). Widely distributed species presented higher standard deviation for the first principal component of the climate predictors ($r$ = 0.670, $p < 0.001$). These patterns were weaker for the climate-edaphic ($r$ = 0.240, $p$ = 0.007) and edaphic predictors set ($r$ = 0.130, $p$ = 0.146; APPENDIX 6 and 7).

The linear mixed-effect models revealed, for all algorithms, that the species geographical extent negatively affected the TSS, the number of records affected SVM, whereas for this algorithm number of records affected the TSS positively, implying better model performance (APPENDIX 8). In addition, we found that different predictors affected model accuracy but the interaction among predictors and species geographic characteristic differed among the ENM algorithms. Interaction between predictor sets and geographical extent were significant for GAM, MElq and RF, but not for SVM, which had significant interaction between the number of records and predictor sets. (TABLE 3; APPENDIX 8). Species geographic characteristics, predictor sets, and their interaction explained between 54 and 68 % of model performance (TSS) variability (TABLE 3).

As expected, the response of the ENMs to different predictors varied individually for each species. Thus, for some species, the use of edaphic data (E.models and CE.models) considerably improved model accuracy in comparison with those models constructed with climate-only predictors (e.g. *Astronium graveolens*, *Cedrela odorata*, *Ficus insipida, Genipa Americana*, *Guarea glabra* and *Salix humboldtiana*). Conversely, edaphic-only predictors notably decreased model performance for other species (e.g. *Casearia decandra*, *Phytolacca dioica*, *Hevea brasiliensis*, *Matayba eleagnoides, Schinus molle* and *Baccharis crispa*). In addition, there were species that presented similar outputs irrespective of the kind of predictors used for modeling (e.g. *Chuquiraga avellanedae*, *Nothofagus pumilio* and *Persea schiedeana*; FIGURE 4). The effect of different predictor variables on the species' suitability pattern varied among species. For example, for species such as *Salix chilensis* and *Guarea glabra* that have broad distributions, CE.models and E.models showed suitable areas that were more constrained compared to those from C.models. Conversely, *Bulnesia sarmientoy* showed an expansion of the suitable areas for those models that used edaphic data compared to those that did not include these data (FIGURE 5).

Based on SVM models, better performance when using the CE.models was observed for 54 species, whereas for 53 species this was true when using the C.models and five species showed the best model performance when using E.models. Also, there were species whose models had the same maximum accuracy independent of the predictors set used. For example, models with climatic-only or climatic-edaphic predictors performed equally well for 13 species, whereas edaphic-only or climatic-edaphic predictors did the same for three species. Finally, models for only three species showed the same TSS irrespective of the considered predictor variables (FIGURE 5).

TABLE 3 - Summary of linear mixed effect models for four algorithms and the significance of covariates. The model selection was based on the likelihood ratio test. GE: geographical extent; NR: Number of records; DR: density of records; Predictor: models constructed with climate, climate-edaphic or edaphic variables; Df: degree of freedom; LRT χ2: Chi square for the likelihood radio test. R2: marginal determination coefficient calculated for the final models with significance values < 0.05 of their covariates

| **Algorithm** | **Covariates** | **LRT $\chi^2$** | **Df** | **p-value** | **$R^2$** |
|---|---|---|---|---|---|
| GAM | GE | 110.204 | 1 | <0.001 | |
| | NR | 0.081 | 1 | 0.777 | |
| | DR | 0.001 | 1 | 0.969 | |
| | Predictor | 63.803 | 2 | <0.001 | |
| | GE*Predictor | 21.418 | 2 | <0.001 | |
| | NR*Predictor | 5.704 | 2 | 0.058 | |
| | DR*Predictor | 0.628 | 2 | 0.628 | 54.255 |
| MEIq | GE | 110.405 | 1 | <0.001 | |
| | NR | 2.104 | 1 | 0.147 | |
| | DR | 0.003 | 1 | 0.957 | |
| | Predictor | 12.642 | 2 | 0.002 | |
| | GE*Predictor | 10.736 | 2 | 0.005 | |
| | NR*Predictor | 4.654 | 2 | 0.097 | |
| | DR*Predictor | 0.191 | 2 | 0.909 | 68.686 |
| RF | GE | 138.093 | 1 | <0.001 | |
| | NR | 0.424 | 1 | 0.515 | |
| | DR | 0.387 | 1 | 0.534 | |
| | Predictor | 139.853 | 2 | <0.001 | |
| | GE*Predictor | 2.678 | 2 | 0.262 | |
| | NR*Predictor | 3.558 | 2 | 0.169 | |
| | DR*Predictor | 1.876 | 2 | 0.391 | 64.059 |
| SVM | GE | 59.592 | 1 | <0.001 | |
| | NR | 19.461 | 1 | <0.001 | |
| | DR | 0.281 | 1 | 0.596 | |
| | Predictor | 133.931 | 2 | <0.001 | |
| | GE*Predictor | 4.389 | 2 | 0.111 | |
| | NR*Predictor | 19.322 | 2 | <0.001 | |
| | DR*Predictor | 0.233 | 2 | 0.890 | 68.614 |

SOURCE: the author (2018)

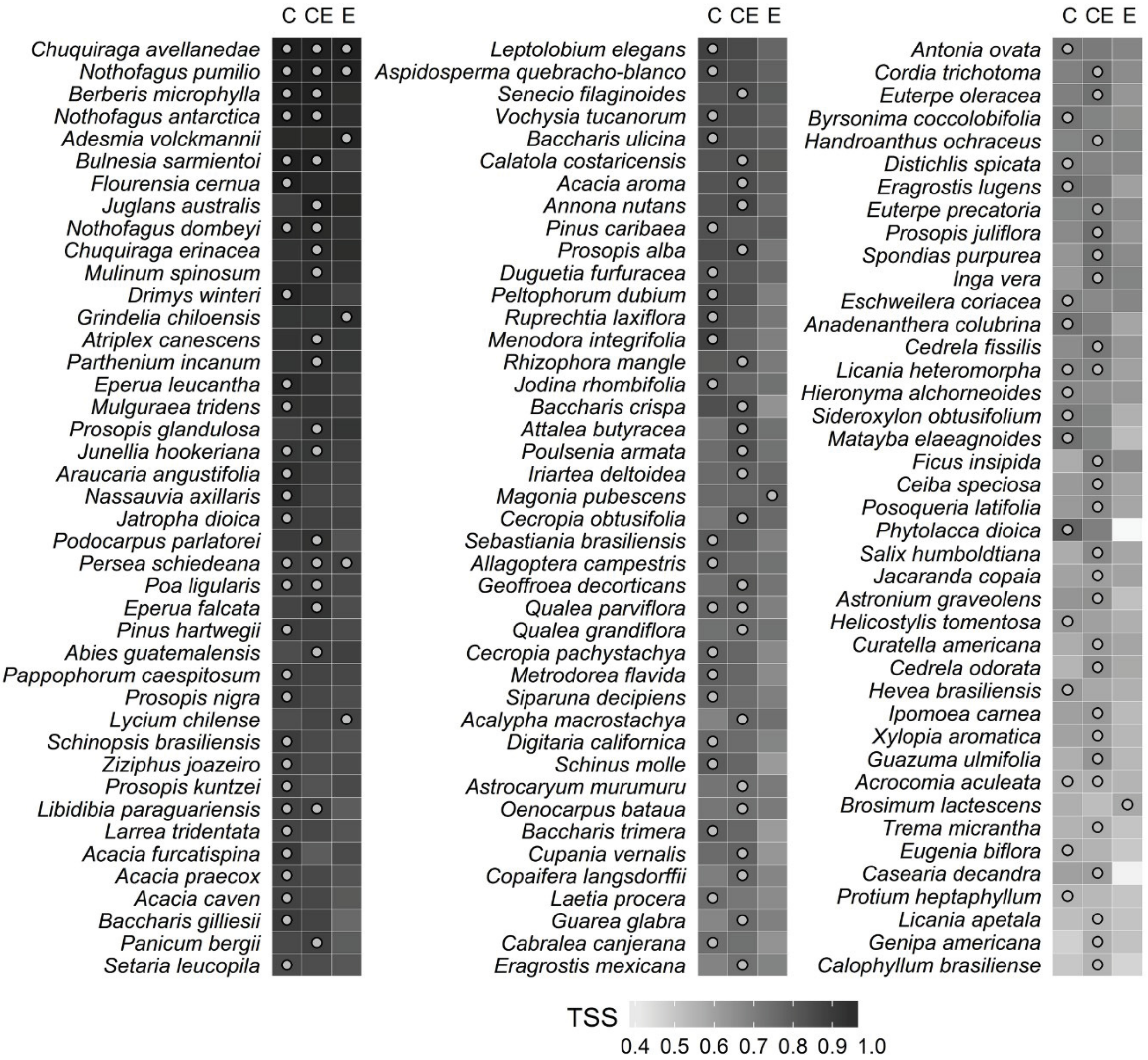


FIGURE 4 - List of modelled species and model accuracy performed with SVM and three predictors sets. Species are sorted by their overall TSS mean, the points represent the highest model accuracy for a particular species. TSS: true skill statistic, C: models with climate predictors, CE: models with climate and edaphic predictors, E: models with edaphic predictors

SOURCE: the author (2018)

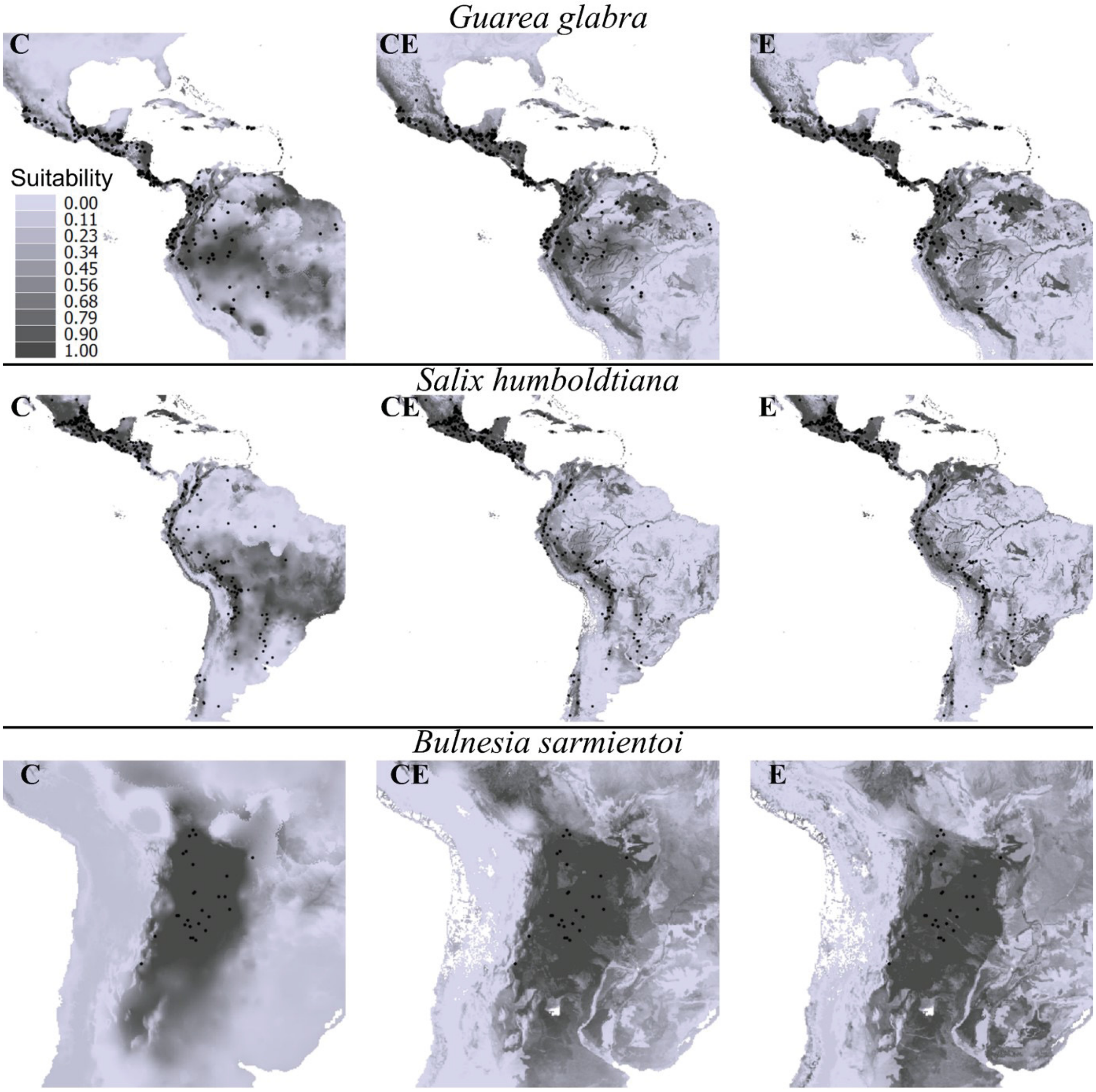


FIGURE 5 - Examples of change in suitability predicted by SVM models for three species derived from the use of three predictors set. C: models with climate predictors, CE: models with climate and edaphic predictors, E: models with edaphic predictors

SOURCE: the author (2018)

## 4. DISCUSSION

We have shown here the advantages of using worldwide edaphic data as predictors in ecological niche models for plant species. More specifically, we showed that ENMs constructed with commonly used climatic variables plus edaphic variables did not affect negatively the performance of ENMs but instead improved the accuracy for some algorithms. This happened even when the ENMs based only on edaphic variables did not provide accurate predictions for any algorithm. Owing to the particular spatial patterning of these different predictor sets, climatic and edaphic, their use also affected the shape and spatial complexity of ENM outputs. In addition, the performance of models considering these predictor sets, separately or jointly, was strongly related to geographic properties of species records, irrespective of the algorithm. Our findings highlight the feasibility and advantages of including global soil data, along with climatic variables, into ENMs to achieve accurate predictions of plant species distributions.

Soils are the consequence of different forming factors such as climate, organisms, topography, parent material, time (JENNY, 1994), among other local factors (SCHAETZL; ANDERSON, 2005). Any particular combination of these factors will give rise to particular processes that can be extremely complex (BREEMEN; BUURMAN, 2002; SCHAETZL; ANDERSON, 2005), involving disintegration, integration, weathering, decomposition, neoformation and transformation (ARNOLD, 2008). These factors and processes acting in soil genesis define the chemical and physical characteristics of soils, which will ultimately determine the underground environment for terrestrial plants. We want to reinforce here the widely accepted idea that soil is one of the most important factors affecting plant ecology and that the common practice of only using climatic predictors represents a weak conceptual basis for the application of ENMs for plant species (see also MOD et al., 2016). Despite potential disadvantages of our testing framework, such as the lack of real species absences and the extension of the edaphic data used, we demonstrated that the combination of climatic and global edaphic predictor variables increases the accuracy of ENMs for several of our studied plant species. In fact, our findings are consistent with previous studies that found similar results for plants at smaller extents and finer resolutions (AUSTIN; VAN NIEL, 2011; BEAUREGARD; BLOIS, 2014; BERTRAND; PEREZ; GÉGOUT, 2012; COUDUN et al., 2006).

The type of organisms under study must guide the selection of predictor variables for ENMs and SDMs. Accordingly, soil properties should be considered when applying ENMs for plant species, whereas these properties may be neglected when modeling animal species (AUSTIN, 2002). However, few studies doing ENM or SDM with plant species have used variables related to soil or, for that matter, variables different than climatic ones (DIEKMANN; MICHAELIS; PANNEK, 2015; THUILLER, 2013). Adding or excluding variables when describing the species environmental niches can affect the form of the resulting hypervolume in multivariate space (APPENDIX 9; HUTCHINSON, 1978). Owing to the reciprocity between the environmental and geographic space (COLWELL; RANGEL, 2009), changes in the multivariate space can affect ENM predictions on geographic space. In our case, this fact may be responsible for the low consensus between the suitability patterns (FIGURE 3A) and the geometry of predicted maps (FIGURE 2) of our modeled species. Regardless of the ENM algorithm used (e.g. GAM vs MElq), using edaphic predictors alone or jointly with climatic variables (E.models and CE.models, respectively) augmented the spatial complexity of the predicted plant distributions. This results from the complex spatial variation of soil properties (HEUVELINK; WEBSTER, 2001) in comparison with climatic variables. Indeed, the spatial complexity of edaphic variables could be responsible for the lower TSS and AUC values predicted by our models based only on edaphic predictors, which are consistent with results obtained for several plant species in Canada with models constructed with the same type of variable and more detailed edaphic data (BEAUREGARD; BLOIS, 2014). Given that different algorithms or variables could produce models with the same accuracy but with different spatial predictions (see FIGURE 5; PLISCOFF et al., 2014), we highlight the importance of acknowledging that model evaluation must be based both on an accuracy metric (e.g. TSS, AUC, etc.) and a preliminary visual examination based on the ecological knowledge of the studied species and their relationships with the selected predictors (AUSTIN, 2007).

It is well known that the performance of an ENM algorithm can vary according to the characteristics of the species niche (QIAO; SOBERÓN; PETERSON, 2015), the training dataset (AGUIRRE-GUTIÉRREZ et al., 2013;

BARBET-MASSIN et al., 2012; DUPIN et al., 2011), and how the algorithm is tuned (ANDERSON; GONZALEZ, 2011; GUILLERA-ARROITA; LAHOZ-MONFORT; ELITH, 2014). One reason that could explain the discrepancy on the accuracy of our different algorithms is the prevalence between presence and absences. For instance, ENMs built under SVM and RF produce better models when the presences/absences ratio is 1, whereas GAM models achieve higher performances with lower presences/absences ratio values (BARBET-MASSIN; THUILLER; JIGUET, 2010). In our case, the fact that we used a number of pseudo-absences equal to that of presences for all algorithms can have negatively affected GAM compared to the other algorithms. Nevertheless, the high positive correlation between GAM and MElq suitabilities is noteworthy. Such correlation may result from the ME tuning with linear and quadratic terms, a feature that reduces the complexity of the ME model making it more similar to GAM. Despite ME being characterized as a high-performance algorithm (ELITH et al., 2006), here this method was outperformed by SVM and RF even when it was conducted with 10,000 background points instead of using the pseudo-absences as in the other algorithms (GUILLERA-ARROITA; LAHOZ-MONFORT; ELITH, 2014). This finding may be particular for ENMs of plant species, for which SVM and RF have been referred as two of the most accurate algorithms for modeling Neotropical plants (LORENA et al., 2011). In addition, we found that SVM and RF were the most accurate methods, although producing different predictions, a fact that reflects their ability to represent complex non-linear relationships. Moreover, each one of these algorithms has important advantages; RF constructs models that avoid overfitting (BREIMAN, 2001) whereas SVM has the ability to construct stable models even with a large set of covariates (LORENA et al., 2011). In fact, both algorithms were the best classifiers with the UC Irvine Machine Learning Repository (FERNÁNDEZ-DELGADO et al., 2014).

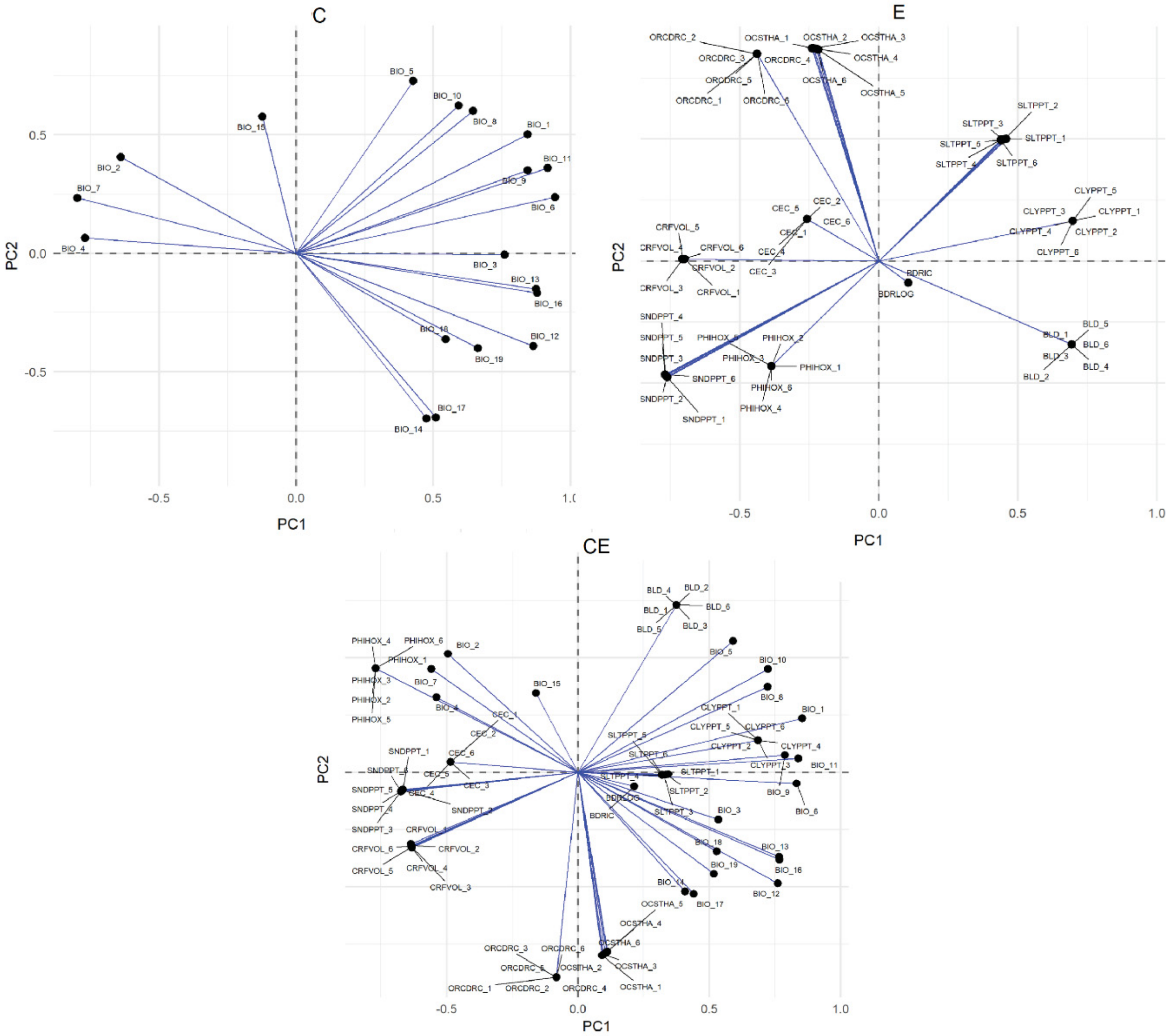
The geographic range size of species can influence the performance of ENMs (LUOTO et al., 2005; MCPHERSON; JETZ; ROGERS, 2004; TSOAR et al., 2007). For instance, small-ranged species could have a limited variability of environmental conditions captured by its presences (SYPHARD; FRANKLIN, 2010) and results from their ENMs may be more marginal (i.e. the difference between the mean environmental condition of the species and the mean of the study areas; HIRZEL et al., 2002) in comparison to other widely distributed species within a particular study region. Our study supports this interpretation given that narrowly distributed species had lower climatic variation represented by their occurrence records. This pattern is more evident in models that considered climatic variables only whereas models including edaphic variables showed wide variability and low correlation between environmental variation and species' range-size (APPENDIX 6). This could explain why some species with the largest geographic ranges showed increased accuracy with the use of the edaphic predictors, whereas for many of the restricted species similar accuracy was observed when using climatic variables only or climatic and edaphic variables together. This finding stresses the dissimilar nature of climatic and edaphic variables and the different way in which those predictors interact with the geographic characteristics of species records.

Our results also showed that species with wide geographic distributions and large numbers of occurrences produce models with lower accuracy, a tendency that is consistent with previous findings (AGUIRRE-GUTIÉRREZ et al., 2013; LUOTO et al., 2005; MCPHERSON; JETZ; ROGERS, 2004; TSOAR et al., 2007), whereas species with a denser aggregation of records showed the opposite trend (APPENDIX 8). On this point we agree with (JIMÉNEZ-VALVERDE; LOBO; HORTAL, 2008), in that the relationship between model accuracy and the geographic range of a species is strongly affected by the extent used to construct the models, which is consequently related with the relative occurrence area and marginality of species. In addition, different extents and resolutions can influence the relative importance of predictors and the predicted suitability (VALE; TARROSO; BRITO, 2014). As the extent increases for an individual model, the environmental difference between predicted and unpredicted cells may increase simply by the broader environmental variability captured by larger extents. This may inflate the metrics designed to estimate model accuracy (LOBO; JIMÉNEZ-VALVERDE; REAL, 2008), especially for species distributed in marginal areas of the environmental space. Other causes for inflating the estimated model accuracy is the relationship between the number of occurrence records and accuracy. In our case, widely distributed species turned out to be more sampled (i.e. had more records) but, as mentioned above, species with more records also showed lower accuracies (APPENDIX 5). Again, one simple explanation for this is that a low discrimination of the environment between predicted and unpredicted cells is expected for species with large distributional area (MCPHERSON; JETZ; ROGERS, 2004).

To the extent of our knowledge, this is the first attempt to evaluate the reliability of using global edaphic information to perform ENMs of plant species over large regions. Nevertheless, ENMs/SDMs experiments have some practical limitations because these approaches are sensible by several factors such as the extent of the area used to construct the models (ANDERSON; RAZA, 2010), the covariates selected and their grain (BEAUREGARD; BLOIS, 2014; VALE; TARROSO; BRITO, 2014), the geographical characteristics and quality of the species records (LUOTO et al., 2005; TSOAR et al., 2007), the pseudo-absences allocation methods (BARBET-MASSIN et al., 2012), the algorithms used and their tuning (GUILLERA-ARROITA; LAHOZ-MONFORT; ELITH, 2014) or even the species selected to perform the experiments (SYPHARD; FRANKLIN, 2010), we acknowledge that there is no “silver bullet” approach that is capable of dealing with all those potential situations (QIAO; SOBERÓN; PETERSON, 2015). Therefore, we suggest that all comparative modeling studies such as ours need to be extrapolated with care. Of course, we are aware that the edaphic data we used may have some deficiencies related to the information on soil sampling and covariates used to generate these data (HENGL et al., 2014) and that it is difficult to make generalizations to other regions of the world. However, the large geographic extent, the variability of environments and the different species geographic characteristics considered here allowed us to show that such global edaphic data adds useful information for plant distribution modeling. This is particularly valuable for studies of species that are distributed in regions where more detailed information on soil properties is poor or does not even exist. Importantly, we do not imply that these global edaphic data must be used in all future studies applying ENMs for plant species, but we do encourage modelers to test some of these edaphic variables and evaluate their model outputs against those conducted with climatic variables only. Recently the SoilGrids was improved by using more accurate technics and with finer-resolution data (HENGL et al., 2015), thus we suggest that future studies consider the effect of different resolutions of soil data when applied to plant ENMs.

# 5. SUPPORTING INFORMATION

APPENDIX 1 - Ordination diagram for the first two axis of three PCAs conducted with three variable set. C: models with climate predictors, CE: models with climate and edaphic predictors, E: models with edaphic predictors



SOURCE: the author (2018)

APPENDIX 2 - List of species modeled, families, habit and number of cleaned record (NR)

| Family | Species | Habit | NR |
|---|---|---|---|
| Pinaceae | *Abies guatemalensis* Rehd. | Tree | 46 |
| Fabaceae | *Acacia aroma* Hook. & Arn. | Tree | 159 |
| Fabaceae | *Acacia caven* (Molina) Molina | Tree | 123 |
| Fabaceae | *Acacia furcatispina* Burkart | Shrub | 35 |
| Fabaceae | *Acacia praecox* Griseb. | Tree | 75 |
| Euphorbiaceae | *Acalypha macrostachya* Jacq. | Tree | 459 |
| Arecaceae | *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. | Palm | 166 |

| Family | Species | Habit | NR |
|---|---|---|---|
| Fabaceae | *Adesmia volckmannii Phil.* | Shrub | 63 |
| Arecaceae | *Allagoptera campestris* (Mart.) Kuntze | Palm | 106 |
| Fabaceae | *Anadenanthera colubrina* (Vell.) Brenan | Tree | 330 |
| Annonaceae | *Annona nutans* (R.E.Fr.) R.E.Fr. | Shrub | 64 |
| Loganiaceae | *Antonia ovata* Pohl | Tree | 113 |
| Araucariaceae | *Araucaria angustifolia* (Bertol.) Kuntze | Tree | 48 |
| Apocynaceae | *Aspidosperma quebracho-blanco* Schltdl. | Tree | 83 |
| Arecaceae | *Astrocaryum murumuru* Mart. | Palm | 66 |
| Anacardiaceae | *Astronium graveolens* Jacq. | Tree | 282 |
| Amaranthaceae | *Atriplex canescens* (Pursh) Nutt. | Shrub | 178 |
| Arecaceae | *Attalea butyracea* (Mutis ex L.f.) We | Palm | 104 |
| Asteraceae | *Baccharis crispa* Spreng. | Shrub | 142 |
| Asteraceae | *Baccharis gilliesii* A.Gray | Shrub | 34 |
| Asteraceae | *Baccharis trimera* DC. | Shrub | 162 |
| Asteraceae | *Baccharis ulicina* Hook. & Arn. | Shrub | 63 |
| Berberidaceae | *Berberis microphylla* Forst. | Shrub | 44 |
| Moraceae | *Brosimum lactescens* (S.Moore) C.C.Berg | Tree | 248 |
| Zygophyllaceae | *Bulnesia sarmientoi* Lorentz ex Griseb. | Tree | 22 |
| Malpighiaceae | *Byrsonima coccolobifolia* Kunth | Shrub | 112 |
| Meliaceae | *Cabralea canjerana* (Vell.) Mart. | Tree | 157 |
| Icacinaceae | *Calatola costaricensis* Standl. | Tree | 201 |
| Calophyllaceae | *Calophyllum brasiliense* Cambess. | Tree | 541 |
| Salicaceae | *Casearia decandra* Jacq. | Tree | 188 |
| Urticaceae | *Cecropia obtusifolia* Bertol. | Tree | 165 |
| Urticaceae | *Cecropia pachystachya* Trécul | Tree | 167 |
| Meliaceae | *Cedrela fissilis* Vell. | Tree | 137 |
| Meliaceae | *Cedrela odorata* L. | Tree | 436 |
| Malvaceae | *Ceiba speciosa* (A. St.-Hil.) Ravenna | Shrub | 80 |
| Asteraceae | *Chuquiraga avellanedae* Lorentz | Shrub | 35 |
| Asteraceae | *Chuquiraga erinacea* D.Don | Shrub | 75 |
| Fabaceae | *Copaifera langsdorffii* Desf. | Tree | 233 |
| Boraginaceae | *Cordia trichotoma* (Vell.) Steud. | Tree | 286 |
| Sapindaceae | *Cupania vernalis* Cambess. | Tree | 97 |
| Dilleniaceae | *Curatella americana* L. | Tree | 467 |
| Poaceae | *Digitaria californica* (Benth.) Henrard | Herb | 324 |
| Poaceae | *Digitaria californica* (L.) Greene | Herb | 357 |
| Winteraceae | *Drimys winteri* J.R.Forst. & G.Forst. | Tree | 53 |
| Annonaceae | *Duguetia furfuracea* (A.St.-Hil.) Saff. | Tree | 390 |
| Fabaceae | *Eperua falcata* Aubl. | Tree | 72 |
| Fabaceae | *Eperua leucantha* Benth. | Tree | 29 |
| Poaceae | *Eragrostis lugens* Nees | Herb | 336 |
| Poaceae | *Eragrostis mexicana* (Hornem.) Link | Herb | 387 |
| Lecythidaceae | *Eschweilera coriacea* (DC.) S.A.Mori | Tree | 324 |
| Myrtaceae | *Eugenia biflora* (L.) DC. | Tree | 200 |
| Arecaceae | *Euterpe oleracea* Mart. | Palm | 72 |
| Arecaceae | *Euterpe precatoria* Mart. | Palm | 393 |
| Moraceae | *Ficus insipida* Willdenow | Tree | 480 |
| Asteraceae | *Flourensia cernua* DC. | Shrub | 55 |
| Rubiaceae | *Genipa americana* L. | Tree | 551 |
| Fabaceae | *Geoffroea decorticans* (Hook. & Arn.) Burkart | Tree | 58 |
| Asteraceae | *Grindelia chiloensis* (Cornel.) Cabrera | Shrub | 43 |
| Meliaceae | *Guarea glabra* Vahl | Tree | 238 |
| Malvaceae | *Guazuma ulmifolia* Lam. | Tree | 1054 |
| Bignoniaceae | *Handroanthus ochraceus* (Cham.) Mattos | Tree | 162 |
| Moraceae | *Helicostylis tomentosa* (Poepp. & Endl.) J.F.Macbr. | Tree | 248 |
| Euphorbiaceae | *Hevea brasiliensis* (Willd. ex A.Juss.) Müll.Arg. | Tree | 85 |
| Phyllanthaceae | *Hieronyma alchorneoides* Allemão | Tree | 347 |
| Fabaceae | *Inga vera* Willd. | Tree | 670 |
| Convolvulaceae | *Ipomoea carnea* Jacq. | Shrub | 317 |
| Arecaceae | *Iriartea deltoidea* Ruiz & Pav. | Palm | 268 |
| Bignoniaceae | *Jacaranda copaia* (Aubl.) D.Don | Tree | 315 |
| Euphorbiaceae | *Jatropha dioica* Sessé | Shrub | 120 |
| Santalaceae | *Jodina rhombifolia* Hook. & Arn. ex Reissek | Tree | 22 |
| Juglandaceae | *Juglans australis* Griseb. | Tree | 29 |
| Verbenaceae | *Junellia hookeriana* (Covas & Schnack) N. O'Leary & P. Peralta | Shrub | 77 |

| Family | Species | Habit | NR |
|---|---|---|---|
| Salicaceae | *Laetia procera* (Poepp.) Eichler | Tree | 167 |
| Zygophyllaceae | *Larrea tridentata* (DC.) Coville | Shrub | 350 |
| Fabaceae | *Leptolobium elegans* Vogel | Tree | 207 |
| Fabaceae | *Libidibia paraguariensis* (D. Parodi) G.P. Lewis | Tree | 105 |
| Chrysobalanaceae | *Licania apetala* (E.Mey.) Fritsch | Tree | 314 |
| Chrysobalanaceae | *Licania heteromorpha* Benth. | Tree | 413 |
| Solanaceae | *Lycium chilense* Bert. | Shrub | 97 |
| Sapindaceae | *Magonia pubescens* A.St.-Hil. | Tree | 79 |
| Sapindaceae | *Matayba elaeagnoides* Radlk. | Tree | 50 |
| Oleaceae | *Menodora integrifolia* Steud. | Shrub | 36 |
| Rutaceae | *Metrodorea flavida* K. Krause | Tree | 91 |
| Verbenaceae | *Mulguraea tridens* (Lag.) N.O'Leary & P.Peralta | Shrub | 32 |
| Apiaceae | *Mulinum spinosum* Pers. | Shrub | 144 |
| Asteraceae | *Nassauvia axillaris* (Lag. ex Spreng.) D.Don | Shrub | 73 |
| Nothofagaceae | *Nothofagus antarctica* (G.Forst.) Oerst. | Tree | 43 |
| Nothofagaceae | *Nothofagus dombeyi* (Mirb.) Oerst. | Tree | 30 |
| Nothofagaceae | *Nothofagus pumilio* (Poepp. & Endl.) Krasser | Tree | 30 |
| Arecaceae | *Oenocarpus bataua* Mart. | Palm | 247 |
| Poaceae | *Panicum bergii* Arechav. | Herb | 124 |
| Poaceae | *Pappophorum caespitosum* R.E.Fr. | Herb | 83 |
| Asteraceae | *Parthenium incanum* Kunth | Shrub | 152 |
| Fabaceae | *Peltophorum dubium* (Spreng.) Taub. | Tree | 124 |
| Lauraceae | *Persea schiedeana* Nees | Tree | 45 |
| Phytolaccaceae | *Phytolacca dioica* L | Tree | 52 |
| Pinaceae | *Pinus caribaea* Morelet | Tree | 61 |
| Pinaceae | *Pinus hartwegii* Lindl. | Tree | 64 |
| Poaceae | *Poa ligularis* Nees ex Steud. | Herb | 115 |
| Podocarpaceae | *Podocarpus parlatorei* Pilg. | Tree | 65 |
| Rubiaceae | *Posoqueria latifolia* (Rudge) Schult. | Tree | 517 |
| Moraceae | *Poulsenia armata* (Miq.) Standl. | Tree | 152 |
| Fabaceae | *Prosopis alba* Griseb. | Tree | 75 |
| Fabaceae | *Prosopis glandulosa* Torr. | Tree | 197 |
| Fabaceae | *Prosopis juliflora* (Sw.) DC. | Tree | 178 |
| Fabaceae | *Prosopis kuntzei* Kuntze | Tree | 36 |
| Fabaceae | *Prosopis nigra* Hieron. | Tree | 54 |
| Burseraceae | *Protium heptaphyllum* (Aubl.) Marchand | Tree | 722 |
| Vochysiaceae | *Qualea grandiflora* Mart. | Tree | 470 |
| Vochysiaceae | *Qualea parviflora* Mart. | Tree | 477 |
| Rhizophoraceae | *Rhizophora mangle* L. | Tree | 307 |
| Polygonaceae | *Ruprechtia laxiflora* Meisn. | Tree | 71 |
| Salicaceae | *Salix humboldtiana* Willd. | Tree | 290 |
| Anacardiaceae | *Schinopsis brasiliensis* Engl. | Tree | 212 |
| Anacardiaceae | *Schinus molle* L. | Tree | 96 |
| Euphorbiaceae | *Sebastiania brasiliensis* Spreng. | Tree | 162 |
| Asteraceae | *Senecio filaginoides* DC. | Shrub | 68 |
| Poaceae | *Setaria leucopila* (Scribn. & Merr.) K.Schum. | Herb | 181 |
| Sapotaceae | *Sideroxylon obtusifolium* (Roem. & Schult.) T.D.Penn. | Tree | 153 |
| Siparunaceae | *Siparuna decipiens* (Tul.) A.DC. | Shrub | 255 |
| Anacardiaceae | *Spondias purpurea* L. | Tree | 319 |
| Cannabaceae | *Trema micrantha* (L.) Blume | Tree | 1227 |
| Vochysiaceae | *Vochysia tucanorum* Mart. | Tree | 299 |
| Annonaceae | *Xylopia aromatica* (Lam.) Mart. | Tree | 462 |
| Rhamnaceae | *Ziziphus joazeiro* Mart. | Tree | 47 |

SOURCE: the author (2018)

APPENDIX 3 - Principal components selected from the PCAs, their eigenvalues, variance explained, and cumulative variance explained for each variable set

| Set of Variables | Principal Components | Eigenvalues | Variance explained for each PC | Cumulative variance explained |
|---|---|---|---|---|
| Climate | 1 | 9.419 | 49.600 | 49.600 |
| | 2 | 4.129 | 21.700 | 71.300 |
| | 3 | 2.481 | 13.100 | 84.400 |
| | 4 | 0.933 | 4.900 | 89.300 |
| | 5 | 0.780 | 4.100 | 93.400 |
| | 6 | 0.423 | 2.200 | 95.600 |
| Edaphic | 1 | 16.284 | 29.078 | 29.078 |
| | 2 | 13.852 | 24.736 | 53.814 |
| | 3 | 11.629 | 20.767 | 74.581 |
| | 4 | 5.935 | 10.598 | 85.179 |
| | 5 | 4.336 | 7.742 | 92.922 |
| | 6 | 1.741 | 3.109 | 96.031 |
| Climate and Edaphic | 1 | 21.959 | 29.300 | 29.300 |
| | 2 | 16.614 | 22.200 | 51.400 |
| | 3 | 12.910 | 17.200 | 68.600 |
| | 4 | 6.770 | 9.000 | 77.700 |
| | 5 | 4.783 | 6.400 | 84.100 |
| | 6 | 3.448 | 4.600 | 88.600 |
| | 7 | 1.790 | 2.400 | 91.000 |
| | 8 | 1.279 | 1.700 | 92.700 |
| | 9 | 1.158 | 1.500 | 94.300 |
| | 10 | 1.053 | 1.400 | 95.700 |
| | 11 | 0.803 | 1.100 | 96.800 |

SOURCE: the author (2018)

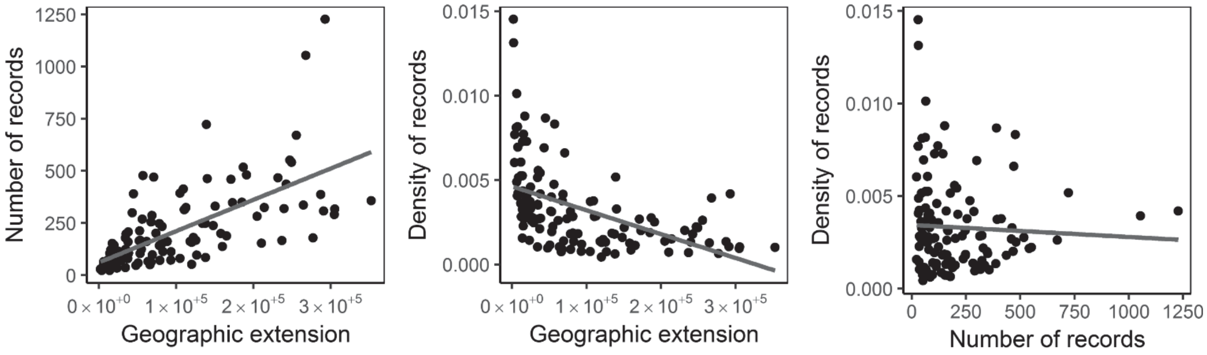
APPENDIX 4 - Coefficients of the Principal components selected from the PCAs performed for each dataset

| Variable | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|---|---|---|---|---|---|---|
| BIO 1 | 0.269 | 0.256 | -0.096 | 0.064 | -0.073 | 0.017 |
| BIO 2 | -0.204 | 0.206 | -0.067 | -0.495 | 0.087 | 0.481 |
| BIO 3 | 0.242 | -0.003 | 0.333 | -0.054 | -0.064 | 0.517 |
| BIO 4 | -0.245 | 0.033 | -0.395 | -0.022 | 0.165 | -0.111 |
| BIO 5 | 0.136 | 0.37 | -0.345 | -0.027 | 0.1 | 0.043 |
| BIO 6 | 0.301 | 0.12 | 0.064 | 0.182 | -0.079 | -0.006 |
| BIO 7 | -0.254 | 0.119 | -0.316 | -0.23 | 0.161 | 0.037 |
| BIO 8 | 0.206 | 0.306 | -0.184 | -0.065 | -0.194 | 0.043 |
| BIO 9 | 0.269 | 0.178 | -0.017 | 0.164 | 0.102 | 0.054 |
| BIO 10 | 0.189 | 0.318 | -0.328 | 0.079 | 0.024 | -0.035 |
| BIO 11 | 0.292 | 0.183 | 0.064 | 0.064 | -0.104 | 0.054 |
| BIO 12 | 0.275 | -0.2 | -0.079 | -0.209 | 0.138 | -0.105 |
| BIO 13 | 0.279 | -0.077 | 0.062 | -0.309 | 0.244 | -0.29 |
| BIO 14 | 0.152 | -0.355 | -0.286 | 0 | -0.106 | 0.355 |
| BIO 15 | -0.039 | 0.293 | 0.384 | -0.386 | 0.119 | 0.097 |
| BIO 16 | 0.28 | -0.086 | 0.059 | -0.306 | 0.24 | -0.276 |
| BIO 17 | 0.162 | -0.353 | -0.285 | -0.005 | -0.079 | 0.319 |

| Variable | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|---|---|---|---|---|---|---|
| BIO 18 | 0.174 | -0.185 | -0.152 | -0.483 | -0.528 | -0.203 |
| BIO 19 | 0.211 | -0.204 | -0.064 | 0.07 | 0.638 | 0.164 |

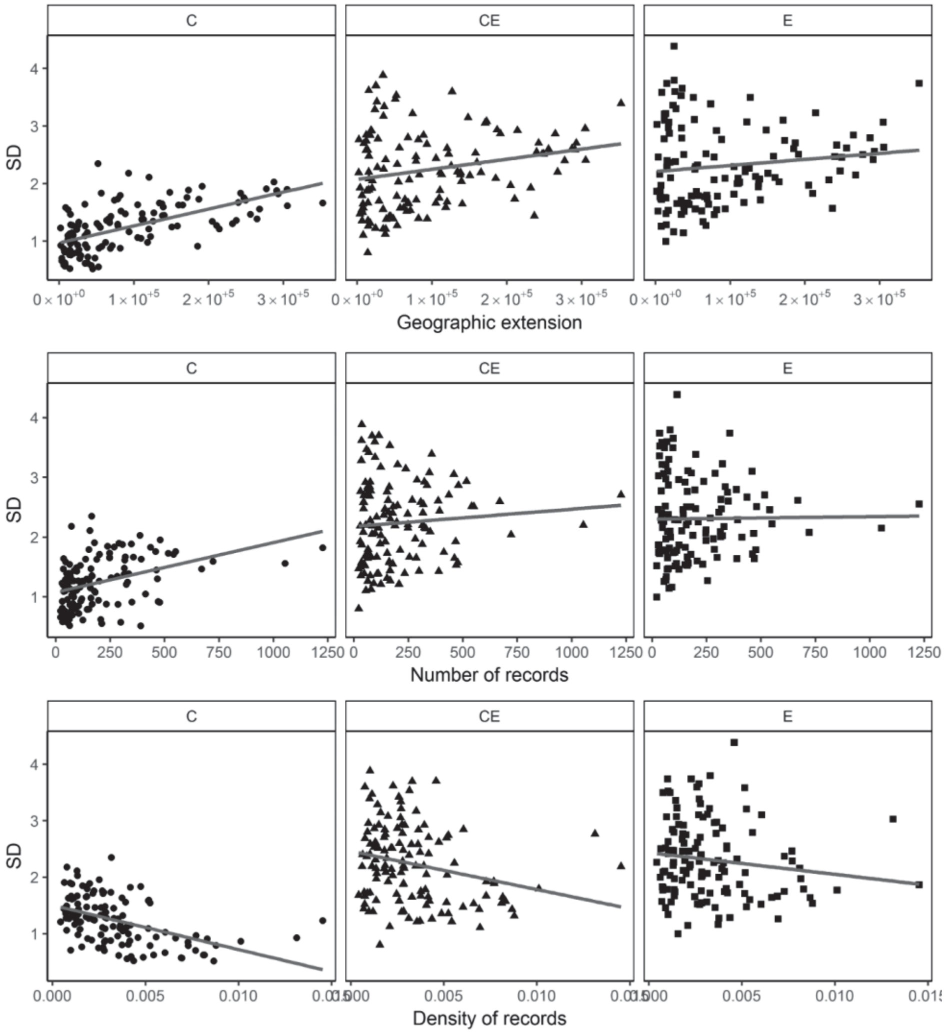
SOURCE: the author (2018)

APPENDIX 5 - Relationship between geographical extent, number of records and density of records for the 125 target species



SOURCE: the author (2018)

APPENDIX 6 - Relationship between the standard deviation for the first principal component, of three predictors set, captured by the records of each species and their relationship with geographical extent, number of records and density of records. C: models with climate predictors, CE: models with climate and edaphic predictors, E: models with edaphic predictors



SOURCE: the author (2018)

APPENDIX 7 - List of species modeled, number of records (NR), geographical extent (GR), density of records (DP) and standard deviation of the first principal component of climatic variables only (C.SD), climatic and edaphic variables (CE.SD) and edaphic variables only (E.SD)

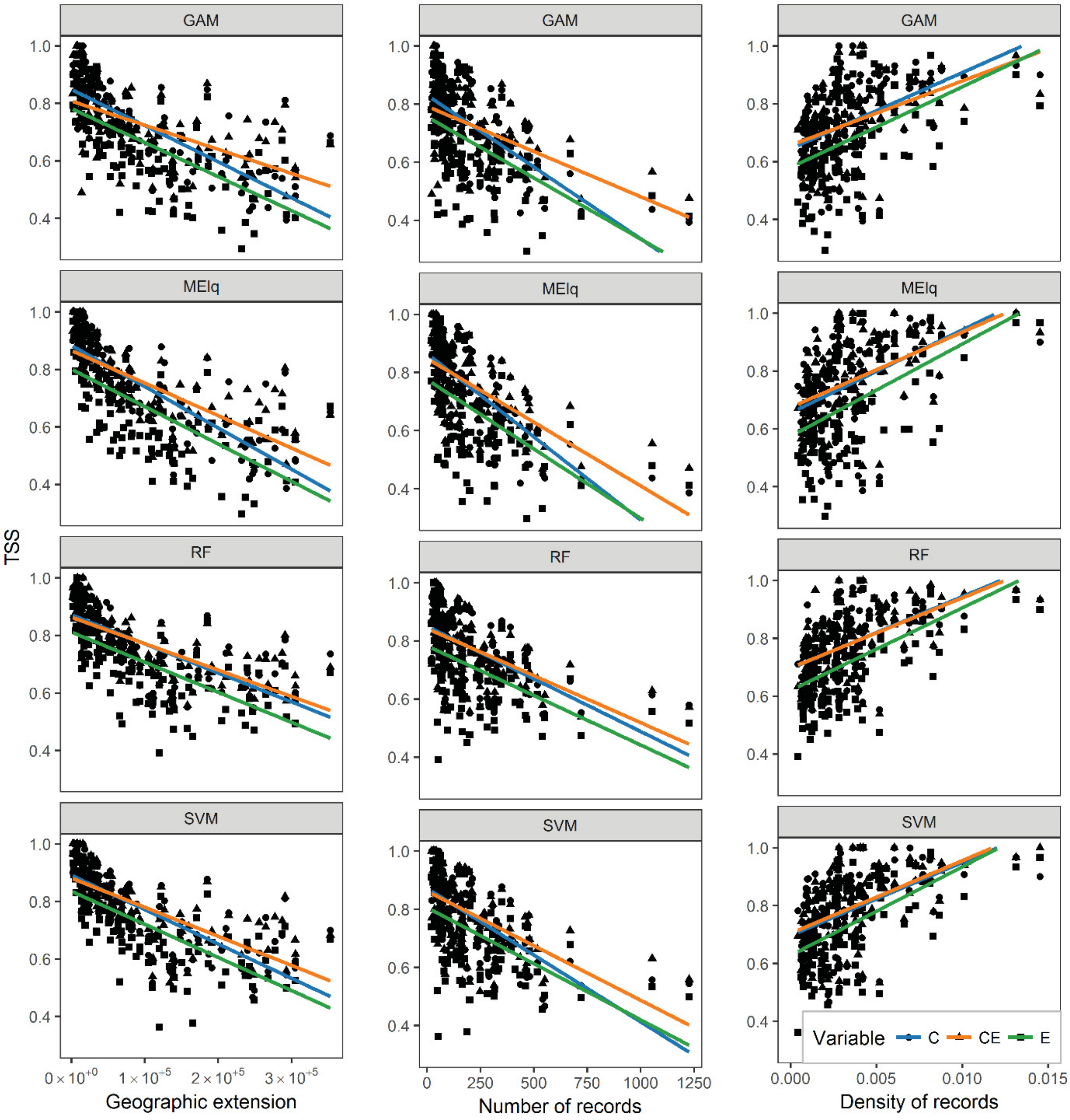
| Species | NR | GE | DR | C.SD | CE.SD | E.SD |
|---|---|---|---|---|---|---|
| *Abies guatemalensis* Rehd. | 46 | 14108 | 0.00326 | 1.129 | 2.581 | 3.214 |
| *Acacia aroma* Hook. & Arn. | 159 | 37983 | 0.00419 | 1.150 | 2.414 | 2.503 |
| *Acacia caven* (Molina) Molina | 123 | 34963 | 0.00352 | 1.263 | 3.171 | 2.962 |
| *Acacia furcatispina* Burkart | 35 | 15379 | 0.00228 | 1.465 | 3.615 | 3.737 |
| *Acacia praecox* Griseb. | 75 | 14282 | 0.00525 | 0.787 | 1.354 | 1.475 |
| *Acalypha macrostachya* Jacq. | 459 | 171035 | 0.00268 | 1.758 | 3.052 | 3.100 |
| *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. | 166 | 236944 | 0.00070 | 1.359 | 1.439 | 1.568 |
| *Adesmia volckmannii Phil.* | 63 | 7715 | 0.00817 | 0.624 | 1.647 | 1.969 |
| *Allagoptera campestris* (Mart.) Kuntze | 106 | 38743 | 0.00274 | 0.724 | 1.381 | 1.517 |
| *Anadenanthera colubrina* (Vell.) Brenan | 330 | 150711 | 0.00219 | 1.247 | 2.141 | 2.176 |
| *Annona nutans* (R.E.Fr.) R.E.Fr. | 64 | 14688 | 0.00436 | 0.522 | 1.262 | 1.282 |
| *Antonia ovata* Pohl | 113 | 83016 | 0.00136 | 1.524 | 2.153 | 1.779 |
| *Araucaria angustifolia* (Bertol.) Kuntze | 48 | 5915 | 0.00811 | 0.625 | 1.543 | 2.160 |
| *Aspidosperma quebracho-blanco* Schltdl. | 83 | 20371 | 0.00407 | 0.998 | 2.048 | 2.098 |
| *Astrocaryum murumuru* Mart. | 66 | 63762 | 0.00104 | 1.361 | 1.724 | 1.338 |
| *Astronium graveolens* Jacq. | 282 | 205073 | 0.00138 | 1.343 | 1.970 | 1.970 |

| Species | NR | GE | DR | C.SD | CE.SD | E.SD |
|---|---|---|---|---|---|---|
| *Atriplex canescens* (Pursh) Nutt. | 178 | 34572 | 0.00515 | 1.094 | 1.551 | 1.709 |
| *Attalea butyracea* (Mutis ex L.f.) We | 104 | 53507 | 0.00194 | 1.270 | 2.335 | 2.291 |
| *Baccharis crispa* Spreng. | 142 | 78551 | 0.00181 | 1.296 | 2.467 | 2.459 |
| *Baccharis gilliesii* A.Gray | 34 | 23677 | 0.00144 | 1.202 | 3.284 | 3.357 |
| *Baccharis trimera* DC. | 162 | 88715 | 0.00183 | 1.473 | 2.840 | 2.764 |
| *Baccharis ulicina* Hook. & Arn. | 63 | 17331 | 0.00364 | 1.144 | 2.837 | 2.972 |
| *Berberis microphylla* Forst. | 44 | 15605 | 0.00282 | 1.288 | 2.477 | 2.397 |
| *Brosimum lactescens* (S.Moore) C.C.Berg | 248 | 134197 | 0.00185 | 1.661 | 2.207 | 2.120 |
| *Bulnesia sarmientoi* Lorentz ex Griseb. | 22 | 3643 | 0.00604 | 0.661 | 1.472 | 1.521 |
| *Byrsonima coccolobifolia* Kunth | 112 | 93649 | 0.00120 | 1.164 | 1.388 | 1.559 |
| *Cabralea canjerana* (Vell.) Mart. | 157 | 120295 | 0.00131 | 2.111 | 2.614 | 2.206 |
| *Calatola costaricensis* Standl. | 201 | 74042 | 0.00271 | 1.710 | 3.213 | 3.382 |
| *Calophyllum brasiliense* Cambess. | 541 | 249427 | 0.00217 | 1.713 | 2.513 | 2.456 |
| *Casearia decandra* Jacq. | 188 | 165441 | 0.00114 | 1.629 | 2.192 | 2.029 |
| *Cecropia obtusifolia* Bertol. | 165 | 52289 | 0.00316 | 2.349 | 3.532 | 3.110 |
| *Cecropia pachystachya* Trécul | 167 | 53447 | 0.00312 | 0.879 | 1.270 | 1.558 |
| *Cedrela fissilis* Vell. | 137 | 159984 | 0.00086 | 1.630 | 2.365 | 2.104 |
| *Cedrela odorata* L. | 436 | 242688 | 0.00180 | 1.832 | 2.704 | 2.667 |
| *Ceiba speciosa* (A. St.-Hil.) Ravenna | 80 | 105164 | 0.00076 | 1.062 | 1.662 | 1.816 |
| *Chuquiraga avellanedae* Lorentz | 35 | 8281 | 0.00423 | 0.594 | 1.576 | 1.722 |
| *Chuquiraga erinacea* D.Don | 75 | 26860 | 0.00279 | 0.854 | 2.280 | 2.379 |
| *Copaifera langsdorffii* Desf. | 233 | 81916 | 0.00284 | 0.878 | 1.394 | 1.544 |
| *Cordia trichotoma* (Vell.) Steud. | 286 | 68706 | 0.00416 | 0.874 | 1.652 | 1.795 |
| *Cupania vernalis* Cambess. | 97 | 69468 | 0.00140 | 1.430 | 2.066 | 1.738 |
| *Curatella americana* L. | 467 | 231996 | 0.00201 | 1.302 | 1.924 | 2.062 |
| *Digitaria californica* (Benth.) Henrard | 324 | 214364 | 0.00151 | 1.210 | 2.923 | 3.225 |
| *Digitaria californica* (L.) Greene | 357 | 352845 | 0.00101 | 1.664 | 3.390 | 3.738 |
| *Drimys winteri* J.R.Forst. & G.Forst. | 53 | 18215 | 0.00291 | 1.274 | 2.917 | 3.063 |
| *Duguetia furfuracea* (A.St.-Hil.) Saff. | 390 | 44976 | 0.00867 | 0.517 | 1.428 | 1.637 |
| *Eperua falcata* Aubl. | 72 | 29121 | 0.00247 | 0.750 | 1.698 | 1.653 |
| *Eperua leucantha* Benth. | 29 | 3770 | 0.00769 | 0.705 | 1.925 | 2.460 |
| *Eragrostis lugens* Nees | 336 | 264847 | 0.00127 | 1.387 | 2.862 | 2.842 |
| *Eragrostis mexicana* (Hornem.) Link | 387 | 287283 | 0.00135 | 2.028 | 2.607 | 2.794 |
| *Eschweilera coriacea* (DC.) S.A.Mori | 324 | 112878 | 0.00287 | 1.380 | 2.158 | 2.023 |
| *Eugenia biflora* (L.) DC. | 200 | 158191 | 0.00126 | 1.263 | 2.528 | 2.474 |
| *Euterpe oleracea* Mart. | 72 | 93195 | 0.00077 | 2.180 | 2.685 | 2.089 |
| *Euterpe precatoria* Mart. | 393 | 105295 | 0.00373 | 1.633 | 2.439 | 2.291 |
| *Ficus insipida* Willdenow | 480 | 191450 | 0.00251 | 1.953 | 2.865 | 2.508 |
| *Flourensia cernua* DC. | 55 | 7912 | 0.00695 | 0.567 | 1.106 | 1.259 |
| *Genipa americana* L. | 551 | 247187 | 0.00223 | 1.757 | 2.510 | 2.219 |
| *Geoffroea decorticans* (Hook. & Arn.) Burkart | 58 | 50862 | 0.00114 | 1.307 | 3.467 | 3.494 |
| *Grindelia chiloensis* (Cornel.) Cabrera | 43 | 15882 | 0.00271 | 0.759 | 2.082 | 2.279 |
| *Guarea glabra* Vahl | 238 | 148227 | 0.00161 | 1.848 | 3.143 | 2.930 |
| *Guazuma ulmifolia* Lam. | 1054 | 268022 | 0.00393 | 1.559 | 2.200 | 2.149 |
| *Handroanthus ochraceus* (Cham.) Mattos | 162 | 118411 | 0.00137 | 0.981 | 1.588 | 1.668 |
| *Helicostylis tomentosa* (Poepp. & Endl.) J.F.Macbr. | 248 | 137789 | 0.00180 | 1.651 | 2.152 | 1.944 |
| *Hevea brasiliensis* (Willd. ex A.Juss.) Müll.Arg. | 85 | 138271 | 0.00061 | 1.443 | 2.188 | 1.999 |
| *Hieronyma alchorneoides* Allemão | 347 | 173089 | 0.00200 | 1.886 | 3.079 | 2.749 |
| *Inga vera* Willd. | 670 | 255898 | 0.00262 | 1.465 | 2.595 | 2.616 |
| *Ipomoea carnea* Jacq. | 317 | 305444 | 0.00104 | 1.613 | 2.405 | 2.629 |
| *Iriartea deltoidea* Ruiz & Pav. | 268 | 56482 | 0.00474 | 1.833 | 2.620 | 2.317 |
| *Jacaranda copaia* (Aubl.) D.Don | 315 | 111133 | 0.00283 | 1.480 | 2.198 | 2.005 |
| *Jatropha dioica* Sessé | 120 | 15558 | 0.00771 | 0.969 | 1.889 | 2.285 |
| *Jodina rhombifolia* Hook. & Arn. ex Reissek | 22 | 14045 | 0.00157 | 0.765 | 0.800 | 0.995 |
| *Juglans australis* Griseb. | 29 | 1996 | 0.01453 | 1.230 | 2.187 | 1.865 |
| *Junellia hookeriana* (Covas & Schnack) N. O'Leary & P. Peralta | 77 | 27702 | 0.00278 | 0.697 | 2.776 | 3.299 |
| *Laetia procera* (Poepp.) Eichler | 167 | 125701 | 0.00133 | 1.349 | 2.306 | 2.261 |
| *Larrea tridentata* (DC.) Coville | 350 | 185224 | 0.00189 | 0.915 | 2.089 | 2.602 |
| *Leptolobium elegans* Vogel | 207 | 38145 | 0.00543 | 0.621 | 1.211 | 1.600 |
| *Libidibia paraguariensis* (D. Parodi) G.P. Lewis | 105 | 14427 | 0.00728 | 0.849 | 1.667 | 1.747 |
| *Licania apetala* (E.Mey.) Fritsch | 314 | 110648 | 0.00284 | 1.378 | 1.881 | 2.091 |

| Species | NR | GE | DR | C.SD | CE.SD | E.SD |
|---|---|---|---|---|---|---|
| *Licania heteromorpha* Benth. | 413 | 109586 | 0.00377 | 1.049 | 1.747 | 1.999 |
| *Lycium chilense* Bert. | 97 | 36039 | 0.00269 | 1.112 | 3.336 | 3.652 |
| *Magonia pubescens* A.St.-Hil. | 79 | 37568 | 0.00210 | 0.886 | 1.124 | 1.144 |
| *Matayba elaeagnoides* Radlk. | 50 | 66233 | 0.00075 | 1.389 | 1.520 | 1.772 |
| *Menodora integrifolia* Steud. | 36 | 34525 | 0.00104 | 1.640 | 3.882 | 3.524 |
| *Metrodorea flavida* K. Krause | 91 | 24427 | 0.00373 | 0.801 | 1.223 | 1.160 |
| *Mulguraea tridens* (Lag.) N.O'Leary & P.Peralta | 32 | 6554 | 0.00488 | 0.583 | 1.356 | 1.870 |
| *Mulinum spinosum* Pers. | 144 | 19762 | 0.00729 | 0.789 | 1.921 | 2.368 |
| *Nassauvia axillaris* (Lag. ex Spreng.) D.Don | 73 | 12052 | 0.00606 | 1.027 | 2.843 | 3.100 |
| *Nothofagus antarctica* (G.Forst.) Oerst. | 43 | 11866 | 0.00362 | 1.098 | 2.579 | 2.773 |
| *Nothofagus dombeyi* (Mirb.) Oerst. | 30 | 2284 | 0.01313 | 0.928 | 2.763 | 3.026 |
| *Nothofagus pumilio* (Poepp. & Endl.) Krasser | 30 | 7369 | 0.00407 | 0.812 | 1.501 | 1.742 |
| *Oenocarpus bataua* Mart. | 247 | 78603 | 0.00314 | 1.732 | 2.387 | 2.064 |
| *Panicum bergii* Arechav. | 124 | 55094 | 0.00225 | 1.238 | 2.029 | 1.864 |
| *Pappophorum caespitosum* R.E.Fr. | 83 | 25162 | 0.00330 | 1.269 | 3.695 | 3.793 |
| *Parthenium incanum* Kunth | 152 | 17290 | 0.00879 | 0.794 | 1.312 | 1.535 |
| *Peltophorum dubium* (Spreng.) Taub. | 124 | 34822 | 0.00356 | 0.615 | 1.212 | 1.462 |
| *Persea schiedeana* Nees | 45 | 8740 | 0.00515 | 1.586 | 2.695 | 3.580 |
| *Phytolacca dioica* L | 52 | 120178 | 0.00043 | 1.210 | 1.682 | 2.262 |
| *Pinus caribaea* Morelet | 61 | 67763 | 0.00090 | 1.393 | 2.799 | 2.516 |
| *Pinus hartwegii* Lindl. | 64 | 12191 | 0.00525 | 1.543 | 2.740 | 3.206 |
| *Poa ligularis* Nees ex Steud. | 115 | 25040 | 0.00459 | 0.932 | 3.700 | 4.383 |
| *Podocarpus parlatorei* Pilg. | 65 | 6422 | 0.01012 | 0.865 | 1.793 | 1.769 |
| *Posoqueria latifolia* (Rudge) Schult. | 517 | 187022 | 0.00276 | 1.731 | 2.931 | 2.705 |
| *Poulsenia armata* (Miq.) Standl. | 152 | 81823 | 0.00186 | 1.794 | 2.760 | 2.919 |
| *Prosopis alba* Griseb. | 75 | 18926 | 0.00396 | 1.108 | 2.922 | 2.870 |
| *Prosopis glandulosa* Torr. | 197 | 35386 | 0.00557 | 0.941 | 2.250 | 2.788 |
| *Prosopis juliflora* (Sw.) DC. | 178 | 277420 | 0.00064 | 1.907 | 2.399 | 2.503 |
| *Prosopis kuntzei* Kuntze | 36 | 12759 | 0.00282 | 1.089 | 1.872 | 1.756 |
| *Prosopis nigra* Hieron. | 54 | 15425 | 0.00350 | 0.956 | 2.200 | 2.258 |
| *Protium heptaphyllum* (Aubl.) Marchand | 722 | 139373 | 0.00518 | 1.593 | 2.039 | 2.079 |
| *Qualea grandiflora* Mart. | 470 | 71126 | 0.00661 | 0.923 | 1.532 | 1.640 |
| *Qualea parviflora* Mart. | 477 | 57324 | 0.00832 | 0.907 | 1.573 | 1.787 |
| *Rhizophora mangle* L. | 307 | 291148 | 0.00105 | 1.877 | 2.587 | 2.414 |
| *Ruprechtia laxiflora* Meisn. | 71 | 78996 | 0.00090 | 1.266 | 1.717 | 1.459 |
| *Salix humboldtiana* Willd. | 290 | 304953 | 0.00095 | 1.899 | 2.956 | 3.066 |
| *Schinopsis brasiliensis* Engl. | 212 | 52919 | 0.00401 | 0.554 | 1.780 | 2.251 |
| *Schinus molle* L. | 96 | 126841 | 0.00076 | 1.490 | 3.597 | 3.495 |
| *Sebastiania brasiliensis* Spreng. | 162 | 62163 | 0.00261 | 0.979 | 1.577 | 1.523 |
| *Senecio filaginoides* DC. | 68 | 26260 | 0.00259 | 1.331 | 3.419 | 3.595 |
| *Setaria leucopila* (Scribn. & Merr.) K.Schum. | 181 | 122429 | 0.00148 | 1.078 | 2.510 | 3.070 |
| *Sideroxylon obtusifolium* (Roem. & Schult.) T.D.Penn. | 153 | 210585 | 0.00073 | 1.275 | 1.733 | 1.827 |
| *Siparuna decipiens* (Tul.) A.DC. | 255 | 67669 | 0.00377 | 1.321 | 1.667 | 1.272 |
| *Spondias purpurea* L. | 319 | 239915 | 0.00133 | 1.671 | 2.548 | 2.469 |
| *Trema micrantha* (L.) Blume | 1227 | 293026 | 0.00419 | 1.823 | 2.703 | 2.553 |
| *Vochysia tucanorum* Mart. | 299 | 43243 | 0.00691 | 0.575 | 1.432 | 1.660 |
| *Xylopia aromatica* (Lam.) Mart. | 462 | 140568 | 0.00329 | 1.450 | 1.647 | 1.716 |
| *Ziziphus joazeiro* Mart. | 47 | 48086 | 0.00098 | 0.708 | 1.391 | 1.784 |

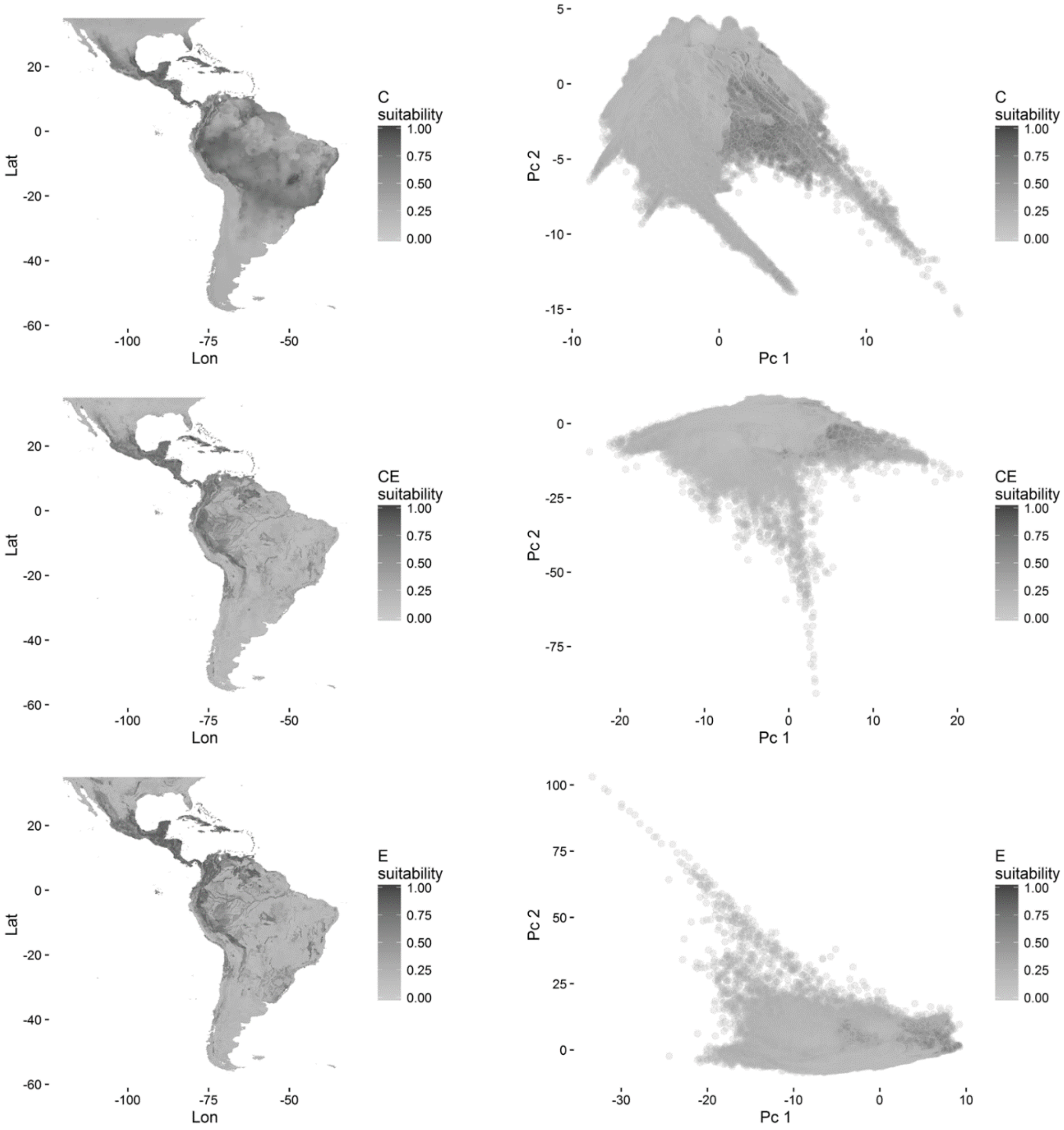
SOURCE: the author (2018)

APPENDIX 8 - Effect of geographical extent, number of records and density of records on the TSS for GAM, MElq, RF and SVM conducted with three predictors sets. TSS: this index was transformed to arcsine; C: models with climate predictors; CE: models with climate and edaphic predictors; E: models with edaphic predictors



SOURCE: the author (2018)

APPENDIX 9 - Predicted suitability by the SVM method for Cedrella odorata and its relationship between the geographical and environmental space for three predictors sets. The right panel shows the first two principal components of the PCA conducted for each variable set. C: models with climate predictors, CE: models with climate and edaphic predictors, E: models with edaphic predictors



SOURCE: the author (2018)

## CAPÍTULO 2

# LOOKING BEYOND BORDERS: PATTERNS OF RICHNESS AND RARITY OF THE CERRADO FLORA AND ITS TRANSBOUNDARY CONSERVATION OPPORTUNITIES

**Abstract**: Knowing the geographic ranges of species is one of the main goals of conservation biogeography. Despite the existence of transboundary projects for biodiversity conservation, conservation planning is defined mainly by local governments and carry out into geopolitical units and not in ecological boundaries. The relationship between the richness or rarity of a region and its remaining natural cover permit identifying scenarios of opportunities or conflicts for conservation. We studied the flora of Cerrado in a transboundary way to (i) predict species richness and rarity patterns for Bolivia, Brazil, and Paraguay, (ii) study the concentration of rare species at global and regional levels and (iii) evaluate areas presenting opportunities and conflicts to conserve the plant richness and rarity of Cerrado across these three nations. We used seven ecological niche modeling techniques to model the distribution of 1559 plant species of Cerrado. We defined scenarios of opportunity and conflict for conservation based on the relationship between the degree of landscape alteration and the richness and concentration of global and local rarity. Rarity index was based on species geographic range sizes, marginality, and specificity of their niches. The greatest concentration of species richness and global rarity of Cerrado's plant species is found in the central part of the Cerrado ecoregion in Brazil. Bolivia is the second country with the highest richness and rarity, mainly associated with Cerrado and Chiquitano Dry Forest. In Paraguay, richness and rarity were concentrated in the northern part of the Humid and Dry Chaco. Opportunities and conflicts for conservation were different for each country resulting from the interaction between diversity and land-use patterns. To the extent of our knowledge, this is the first attempt to describe the plant diversity of the Cerrado biome beyond Brazilian borders. Despite much of the Cerrado flora being concentrated in Brazil, this country also has the greatest proportion of its territory under a conflict scenario where the major landscape alteration happens in areas with the highest diversity. However, high opportunities for conservation were found in Bolivia and Paraguay where species richness and rarity concentrate in areas with low alteration degree, which makes them important countries to conserve the flora of Cerrado.

**Keywords**: anthropized landscapes, country boundaries, neotropical savannas, plant species conservation, rarity index, species distribution models

# 1. INTRODUCTION

Knowing the geographic ranges of species is one of the main goals of conservation biogeography, owing mainly to their use in assessing priority areas for conservation (MARÉCHAUX; RODRIGUES; CHARPENTIER, 2017; MOILANEN; KUJALA; LEATHWICK, 2009) and determining the conservation status of species (IUCN, 2017). Usually, such assessments are conducted at the country level, particularly the definition of protected areas, where political and economic decisions determine the final outcome of conservation plans (BARQUET; LUJALA; RØD, 2014; LIM, 2016; MUBOKO, 2017). Interestingly, conservation status of species is commonly established at both “global” (e.g. IUCN Red List) and “national” levels (GRAMMONT; CUARÓN, 2006). In this context, conservation of species and ecosystems that are geographically distributed over many countries represents an important challenge (CRAIN; WHITE; STEINBERG, 2011; KARK et al., 2015). Despite the existence of transboundary projects for biodiversity conservation, planning is defined mainly by national governments and carried out within geopolitical units and not within ecological boundaries (DALLIMER; STRANGE, 2015; HUNTER; HUTCHINSON, 1994). Therefore, national-level strategies can have important consequences on the conservation output since countries present political, economic, and cultural idiosyncrasies (DALLIMER; STRANGE, 2015; HUNTER; HUTCHINSON, 1994; KARK et al., 2015; RODRIGUES; GASTON, 2002). For instance, such national-level strategies can define the contribution of species for local conservation values (e.g. a region within a country can be valuable for conservation because it harbors a species which is locally rare but widely distributed); determine different conservation status for species between countries or administrative divisions (WELLS et al., 2010) or can influence the selection of priority areas for conservation (RODRIGUES; GASTON, 2002).

Rarity is a key trait to determine species vulnerability to environmental impacts and, consequently, its conservation status (GASTON, 1994). Different variables have been used to define species’ rarity, including abundance, habitat specificity, occupancy area and others (GUERIN; LOWE, 2015; KUNIN; GASTON, 1997; RABINOWITZ, 1981). Among these, the distributional range is the most common metric used to measure species’ rarity (GUERIN; LOWE, 2015). Rarity, in a distributional sense, makes a species prone to local extinction because environmental impacts are usually spatially-restricted phenomena (IŞIK, 2011). For instance, the expansion of agriculture from the south of Amazonia make species restricted to that region especially threatened simply because they are in the wrong place at the wrong time (TER STEEGE et al., 2015). Because habitat loss or fragmentation, spatially restricted species may be lost during these events (IŞIK, 2011). Given that a species geographic distribution is scale-dependent (GASTON, 2003), its interaction with geopolitical boundaries can result in different forms of rarity. For example, there are species that are naturally widely distributed but locally rare within some specific geopolitical boundaries.

Species are not randomly distributed across space mainly because different abiotic and biotic variables determine their distributions (GASTON, 2003). Moreover, species interact with each other arranging into communities and can pertain to associations such as a biome. The presence of a single biome in several countries may indicate the possibility of transboundary opportunities to conserve it. In this sense, the Cerrado biome is an important model of study because of its presence in three different countries (i.e. Bolivia, Brazil, and Paraguay). Essentially located on the Central Plateau of Brazil, the Cerrado is the second largest biome of this country (RATTER; BRIDGEWATER; RIBEIRO, 2003). In Bolivia, this biome is present in the northeastern region in areas that are classified as Cerrado ecoregion, but also in the Dry Chaco, Chiquitano Dry Forest, Beni Savanna and Pantanal (BECK, 2015; IBISCH et al., 2003). In Paraguay, Cerrado distributes in the northern extreme of Eastern Paraguay and the northern extreme of the Paraguayan Dry Chaco (MERELES, 2013; MERELES et al., 2013). High diversity and degree of endemism put the Cerrado biome as one of the most diverse Neotropical biomes (SILVA; BATES, 2002). Agriculture and other economic activities are responsible for the loss of habitat in the Cerrado, which makes it highly threatened (OVERBECK et al., 2015; STRASSBURG et al., 2017). The fact that this biome is present in several countries that may suffer different patterns of land-use change, highlights the importance of the three nations to conserve the Cerrado’s flora.

Conservation values conferred to nature can be conceived by different viewpoints depending on the relationship between nature and humans (MACE, 2014). They can have an intrinsic, utilitarian, or even a non-use value and can be related to the spatial extent and level of biological organization (PEARSON, 2016). When considering spatially explicit conservation based on a grid system, cells within this grid are considered as planning units (MARGULES; PRESSEY, 2000) and the presence of species or ecosystems within such units are commonly used to determine their conservation value (PEARSON, 2016). The remaining natural cover within a cell is a surrogate of the intensity of anthropic activities developed in that cell’s landscape. In this context, the relationship between the conservation value and the anthropic activities within a cell allows identifying scenarios of opportunities or conflicts for conservation (FIGURE 1). Large-scale environmental variables (e.g. climatic conditions) can determine how suitable a grid cell is for a particular species, nevertheless, at smaller scales the persistence of a species will be constrained by the remaining habitat within that cell (MENDES; DE MARCO, 2017). Landscape characteristics, such as the degree of fragmentation, connectivity, edge density and others, could affect the persistence of a species population over time (FAHRIG, 2001; SWIFT; HANNON, 2010). Accordingly, there is a critical point of habitat loss, called critical or extinction threshold, where a species may experience local extinction (GIBBS, 1998; SWIFT; HANNON, 2010). The remaining natural cover could determine the presence of a species within a cell, which will limit the conflicts and opportunities for conservation (see vertical dashed line in FIGURE 1). For example, areas harboring a certain amount of preserved landscape (above a defined critical threshold) and simultaneously with a richness or rarity above the overall mean of the study region, can be considered as having high opportunities for conservation (upper right panel of FIGURE 1). Conversely, areas with high richness or rarity overlapping with strongly altered landscape (i.e. below critical threshold) can be assumed as having high degree of conflict, since the conservation interests and anthropic activities compete for the same space (SHACKELFORD et al., 2015). In addition, areas with richness or rarity below the region’s mean could be considered of low conservation interest or low conflict with land-use, according to whether they are above or below the critical threshold, respectively (right and left bottom panel FIGURE 1).

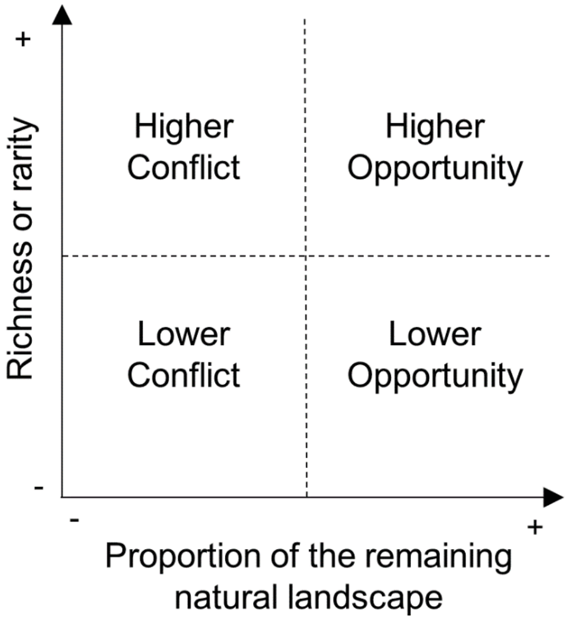


FIGURE 1 - Conceptual framework used to define conflicts and opportunities for conservation based on the relationship between the richness or rarity of a region and the remaining natural landscape. Vertical dashed line denotes a critical threshold value from which we assumed that the anthropic activities do not affect the persistence of the species in a cell. Horizontal dashed line depicts the mean richness or rarity of a region
SOURCE: the author (2018)

A comprehensive description and consideration of the Cerrado biome can assist its conservation at both regional and global levels. Indeed, under a regional context for a singular geopolitical unit (i.e. country), Cerrado species may present restricted spatial distributions being of relevance to the biodiversity and conservation of that country. Complementarily, under a global context, countries may present different patterns of habitat loss and areas with remaining vegetation may help conserve the biome in a more integrated manner. Here, we studied the flora of Cerrado in a transboundary context to (i) predict species richness and rarity patterns for Bolivia, Brazil, and Paraguay, (ii) study the concentration of rare species within the Cerrado biome at global (transboundary) and regional (country) levels and (iii) evaluate and detect areas presenting opportunities and conflicts to conserve the plant richness and rarity of Cerrado across these three nations.

## 2. METHODS

### 2.1 STUDY AREA

The Cerrado contains different phytophysiognomies that vary from open grasslands to forests, therefore our objectives were defined using as basic unit the Cerrado as a Biome independent of vegetation formation. In addition, to evaluate the patterns of plant richness and rarity, our study area comprised the three nations where this biome has been recorded: Bolivia, Brazil, and Paraguay. These countries disagree regarding the methodology utilized to define their ecoregions, thus disagreements about Cerrado limits are also present. To overcome this issue we used the Terrestrial Ecoregions of the World (OLSON et al., 2001). We decided to include those ecoregions related to open formations such as steppes and savannas, and others that are not open formations but where the existence of Cerrado is well known. Thus, we considered the Cerrado, Campos Rupestres Montane Savanna, Pantanal and Maranhão Babaçu Forests ecoregions for Brazil; and the Beni Savanna, Chiquitano Dry Forest, Dry Chaco and Humid Chaco for Bolivia and Paraguay.

### 2.2 SPECIES SELECTION

We modeled the geographic distribution of plant species (species distribution modeling, SDM) inhabiting the Cerrado of three nations using an ecological niche modeling (ENM) approach. These species comprise different life forms such as trees, shrubs, subshrubs, and herbs. Given that the Cerrado is inhabited by plants that constitute different kind of phytophysiognomies; several of its species could be predominant in other neighboring biomes (e.g. Atlantic Forest, Amazon Forest; FRANÇOSO et al., 2016). Therefore it is difficult to determine which species are well distributed within the Cerrado or only occurs in a marginal way. For this reason, we created a general species list of the Cerrado for the three countries and then selected the taxa to be modeled based on two criteria. First, we compiled species lists from the national plants catalogs sourced by Brazilian Flora 2020[11], Bolivia Catalogue[12] and Paraguay Checklist[13] (see APPENDIX 1 for more information about search criteria used for each country to list Cerrado´s plant species). We only considered native, endemic and non-endemic species that are terrestrial or rupicolous, whereas those species listed as strictly aquatic or epiphytic were not considered. Regarding Cerrado’s endemic species, we used only those from Bolivia and Brazil because the Paraguay list did not provide this information. Infra-specific taxa were not considered.

As the lists came from different sources, species may have been listed with different names thus presenting synonyms. To correct for this, once the species lists of each country were merged, we homogenized and updated the scientific names using the Taxonomic Name Resolution Service v4.0 (TNRS; BOYLE et al., 2013) based on APG III. After deleting duplicated names, the final species list comprised a total of 6097 scientific names. To be modeled, species must (i) have more than five occurrence records and (ii) more than 40% of their occurrences located in the Brazilian Cerrado (this criterion was not used for those species endemic to Bolivia).

We used the Global Biodiversity Information Facility[14] (GBIF) as source of records of the species list. We applied a preliminary cleaning to the records of each species, which consisted of removing occurrences that were outside the continent and had repeated coordinates. To count the proportion of species records that fell within the Brazilian Cerrado, we created a layer of Cerrado distribution with a 10 km resolution overlapping the maps of Cerrado boundary taken from Olson et al (OLSON et al., 2001) and IBGE (2012). After the selection of taxa from the species list, we added to this list species suggested by other researchers or cited in the literature (FRANÇOSO; HAIDAR; MACHADO, 2016; MAMANI et al., 2011; RATTER; BRIDGEWATER; RIBEIRO, 2003).

---

[11] http://floradobrasil.jbrj.gov.br/
[12] http://www.tropicos.org/Project/BC
[13] http://www.tropicos.org/Project/Paraguay
[14] http://www.gbif.org/

## 2.3 SPECIES RECORDS AND DATA CLEANING

Based on the selected species from the GBIF database, we also used other sources of occurrence data such as speciesLink[15]; ICMBio[16] ; Plant of Bolivia[17] , and Tropicos. The species name of each record was corrected and updated using the Taxonomic Name Resolution Service v.4.0 (BOYLE et al., 2013), The data cleaning comprised the elimination of records that (i) were located outside the continent, (ii) had repeated geographic coordinates, (iii) had coordinates without decimals or with half degree decimal (i.e. 0.5), (iv) were located in the countries centroid and (v) had no specification of species or genus. Next, we removed those records corresponding to invaded or cultivated areas, thus we left only those records that pertain to the natural distribution of the species. This step was conducted by (i) using information about species distributions available in the Brazilian Flora 2020, Flora del Conosur[18], and national species list from Tropicos; and (ii) searching in the occurrences metadata information that indicated that the collected specimen was cultivated (i.e. cultivated in botanical gardens, parks, universities, nurseries, etc.). Also, we checked the location where the records were georeferenced and the information on location contained in the voucher specimen. This procedure was carried out at provincial/state level for all the countries comprised by the records total geographic extent (i.e. Caribbean, North, Central and South America). For this, we used data from the Global Administrative Areas database v. 2.8[19] and a geographic information system (QGIS v. 2.18.7).

For species with <20 cleaned records, we added occurrences georeferenced at the municipal level provided by speciesLink. We only considered those records located in municipalities with a variation coefficient ≤15% for any of the 11 variables used for constructing the models (see below). The variation coefficients of each municipality were calculated based on the cells of each variable within each municipality. Most of the municipalities from Brazil had variation coefficients <15% allowing the use of most records (APPENDIX 2). We used a systematic sampling

---

[15] http://splink.cria.org.br/
[16] https://biodiversidade.icmbio.gov.br/
[17] http://herbaria.plants.ox.ac.uk/bol/boliviajriwood
[18] www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm
[19] http://gadm.org/

with a grain twice the resolution of environmental variables to correct the geographic bias of species records. Only species with more than five cleaned records were modeled. Thus, the final database comprised 132,512 records for 1559 species (see APPENDIX 3 for further information about the modeled species).

## 2.4 ENVIRONMENTAL DATA

Species ENMs were constructed using climate and edaphic variables because both are important to describe plants niches and could generate models with better accuracy compared to those constructed only with climate variables (VELAZCO et al., 2017). Climate data comprised the 19 bioclimatic variables from the CHELSA v1.1 database (KARGER et al., 2016). Also, we used the annual aridity and potential evapotranspiration from CGIAR-CSI[20] , both datasets with a resolution of 30 arc-seconds. We used SoilGrids (HENGL et al., 2017) as a source of edaphic data. We selected nine variables related to physical and chemical soil properties for seven depths with 0.75 arc-seconds resolution (APPENDIX 4). All variables were upscaled to 5 arc-minutes and cropped to the extent from -120° to -30° in longitude and -60° to 35° in latitude. To avoid multicollinearity of the variables in the ENMs, we conducted a principal component analysis on the original environmental variables based on a correlation matrix, then we used the scores of each derived principal components (PCs) as new variables. We selected 11 PCs which explained up to 95.20% of the variance in the original environmental variables (APPENDIX 5 and 6).

## 2.5 MODELING PROCEDURES

We used six ENM approaches: Generalized Linear Models (GLM), Generalized Additive Models (GAM), Maximum Entropy (ME), Random Forest (RF), Support Vector Machine (SVM), and Gaussian Processes (GP; see APPENDIX 7 for

---

[20] http://www.cgiar-csi.org/

further information about how these algorithms were tuned). We created pseudo-absences to fit GLM, GAM, SVM, RF and GP models. The ratio between presences and pseudo-absences, and the method of allocation, can distinctively affect the algorithm performance (BARBET-MASSIN et al., 2012). Thus, we used presence/pseudo-absence ratio equal to one for SVM, RF and GP approach, whereas for GLM and GAM we used 10,000 pseudo-absences. The pseudo-absences were allocated using environmental restriction. These were based on adjusting and predicting a bioclim model, and then distribute the pseudo-absences in the areas with the lowest suitability values. The areas used to adjust the models (i.e. M component of the BAM diagram; SOBERÓN, 2007) were delimited by the geographic extent of all species records.

## 2.6 MODEL EVALUATION, ENSEMBLE FORECAST OVERPREDICTION CORRECTION

We used two approaches for evaluating species ENMs depending on the number of records. For those species with 5-19 occurrences, we used a jackknife procedure, where each partial model is constructed with *n* – 1 records. For species with ≥ 20 occurrences, we implemented a block fold-validation (ROBERTS et al., 2017) with two partitions (like a checkerboard) in order to control for spatial autocorrelation between training and testing data. To find a grid resolution for each species, we generated 20 grids with resolutions varying from 0.5 to 10 degrees, with a gradual increase of 0.5. The best grid was the one which the training and testing data presented (i) low spatial autocorrelation, (ii) maximum environmental similarity and (iii) minimal difference in the number of records. These three parameters were measured by Moran's I index, Multivariate environmental similarity surfaces, and the standard deviation, respectively. Finally, pseudo-absences were created within each partition group as explained above. We used the True Skill Statistic (TSS) as a metric of model performance.

The final model of each species was constructed using an ensemble forecast procedure based on the arithmetic average of the suitability predicted by the best algorithms, i.e. for each species we used the models with a performance greater than or equal to the algorithms' average TSS. Binary models were obtained using the threshold that maximizes the sum of the sensitivity and specificity of each species.

When an ENM is projected onto the entire study region, it is common to obtain areas of high environmental suitability that fall outside the known distribution domain of a species. The elimination of these areas transforms a species potential distribution closer to its actual distribution, indeed allowing to go from an ENM to a SDM (PETERSON; SOBERÓN, 2012). To correct the ENMs overprediction we used an approach that retained suitable patches for each species according to two criteria: (i) patches with at least one presence record were selected and (ii) for suitable patches without presences, we chose those that were a distance apart from patches with presence records. Such distance was lower or equal than the distance at the first quartile of the pair-wise distance between suitable patches with and without presences.

## 2.7 GLOBAL AND REGIONAL RARITY AND RICHNESS MAPS

Rarity measures can be based on many species attributes (RABINOWITZ, 1981), which are commonly dichotomized into categories and assigned a weight of rarity to each combination class (KATTAN, 1992). Here, we proposed a continuous rarity index based on the inverse of range size (*G*), the overall marginality (*M*) and the global specialization (*S*) of each species. The geographic range size is a commonly used measure to calculate rarity (GUERIN; LOWE, 2015), where geographically restricted species are considered rarer than those more widely distributed. *M* and *S* are two components associated with the Grinnellian niche (SOBERÓN, 2007) by describing the preferred environmental conditions of a species.

*G* was calculated based on species range predicted by the binary models, it is defined as: $G = 1 - (g_i - g_{min})/(g_{max} - g_{min})$; where $g$ is a vector of range sizes of all target species and $g_i$ is the range of the $i^{th}$ species. The range size of a species was determined by the number of cells encompassed by its range. $M$ is the environmental distance between the optimum environmental condition of a species and the mean environmental conditions within the study region, whereas $S$ is defined as the ratio between the environmental variance of the study region and that of the species (HIRZEL et al., 2002). Based on Hirzel's formula and notation, $M = \sqrt{\sum_{e=1}^{V} m_e^2}/1.96$; where $m_e$ is the marginality of the $e^{th}$ environmental variable $V$; and $S = \sqrt{\sum_{e=1}^{V} \lambda_e}/V$; where $\lambda_e$ represent the eigenvalue for $e^{th}$ factor $V$. $M$ and $S$ parameters were standardized between 0-1 using the next formula $X_s = (X - X_{min})/(X_{max} - X_{min})$, where $X_s$ is the standardized values of a $X$ vector (i.e. $M$ or $S$). The rarity index $(R)$ was calculated as $R = (G + M + S)/3$. Thus, greater values of $G$, $M$ and $S$ yield greater magnitudes of $R$. As the parameters that compound $R$ are dependent on the spatial extent (i.e. study region) used to calculate them, this index can be derived in two different ways: (i) comprising the total geographic extent of the study area and (ii) including only the extent of the study area within each country; hereafter global and regional rarity respectively. Such index presents several main advantages: (i) it adds explicit information related to characteristics of species niches, such as marginality and specificity; (ii) its dependence on the extent used for calculation, allowing computing it for different spatial extents (global and regional) and, regarding its regional level version, (iii) it permits comparisons among countries regardless of their species richness patterns.

Species richness maps were constructed by stacking individual SDMs. Rarity maps were created using the following formula, $R_c = \sum_i^T P_i \times R_i$, where $R_c$ represents the cumulative rarity of all species present in a single cell, $P_i$ being the presence and $R_i$ the rarity of the $i^{th}$ species present in a cell, and $T$ is the total species richness of that cell. Rarity maps were standardized between 0-1 using the same formulae described above.

## 2.8 LAND COVER AND OPPORTUNITY FOR CONSERVATION

The land cover classification was obtained from the Climate Change Initiative of the European Space Agency[21]. We used the data for the year 2015 with a 300 m resolution. We reclassified the 37 land cover categories in natural and transformed cover, whereas water bodies and permanent snow and ice were not considered (see APPENDIX 8 for further information about this classification). Then, we calculated the proportion of the remaining natural landscape in each cell using the formula $L = N/(N+T)$, where $L$ is the proportion of natural landscape remaining within a cell of 10 km$^2$ (i.e. the resolution used to construct the SDMs); whereas $N$ and $T$ are the number of cells (of 300 m$^2$) with natural and transformed cover respectively within the cell of 10 km$^2$.

We used the conceptual framework presented in FIGURE 1 to study conflicts and opportunities for conservation of the Cerrado flora. We assumed 0.5 as the critical threshold, i.e. a cell is considered as unfeasible to the presence of a species if its remaining natural landscape ($L_{10}$) was ≤0.5. We selected this threshold because it has been proven that the chance of population extinction could increase when the habitat loss exceeds c. 50% of a given landscape's surface (GIBBS, 1998; MENDES; DE MARCO, 2017; MUYLAERT; STEVENS; RIBEIRO, 2016; SCHMIDT; ROLAND, 2006). The high or low conservation opportunity or conflict of a cell was determined based on the mean richness and both kinds of rarity, global and regional. For this, we summed all the cell values of richness or rarity within a target area and divided by the number of cells. For each conservation category, we also calculated the proportion that is within protected areas (see APPENDIX 9 for more information about protected areas network).

## 2.9 DATA ANALYSIS

---

[21] http://maps.elie.ucl.ac.be/CCI/viewer/index.php

Data processing, construction of ENMs and statistical analyses were conducted in the R environment v.3.4.1 (R CORE TEAM, 2017). We used the packages *maptools*, *raster*, *rgdal*, and *sp*, to handle spatial data; *rgbif* to download records from GBIF; *adehabitatHS* for calculating the *S* and *M* parameters, *flora* for getting data from Brazilian Flora 2020. The *dismo* package was used to create pseudo-absences, predict and validate models. The different ENMs approach were fitted using *GRaF*, *kernlab*; *maxnet*; *mgcv*; *randomForest*; and *stats* packages.

## 3. RESULTS

### 3.1 RICHNESS AND RARITY OF CERRADO PLANT SPECIES

We modeled 1559 plant species distributed in 104 families of Angiosperms and Pteridophytes. The families with the largest number of species were Asteraceae (314), Fabaceae (143) and Poaceae (103), which together comprised more than 35% of our modeled species. Regarding plant life forms, species comprised herbs (590), shrubs (455), subshrubs (275), trees (151), and vines (88). Models had a satisfactory performance with mean TSS for all species of 0.90 ± 0.08 (see APPENDIX 10). The species richness map revealed that the major concentration of plant richness (c. 1550 species), is harbored in the central area of the Cerrado domain in Brazil, mainly in the states of Goiás, Minas Gerais, northern São Paulo and west of Bahia. This richness decreases markedly towards the northwest of our study area in the Maranhão Babaçu Forest and in the Pantanal at the southwest. The second richest country is Bolivia (c. 1070), where the predicted species richness was higher in the northern and northeastern parts of the municipality of Santa Cruz, in the Cerrado and Chiquitano Dry Forest. Considerable richness was also predicted in the upper portion of the Beni Savanna. The country with the lowest predicted plant richness was Paraguay (c. 850), where species concentrated on the north and east of the Humid Chaco and on the north of the Dry Chaco (FIGURE 2).

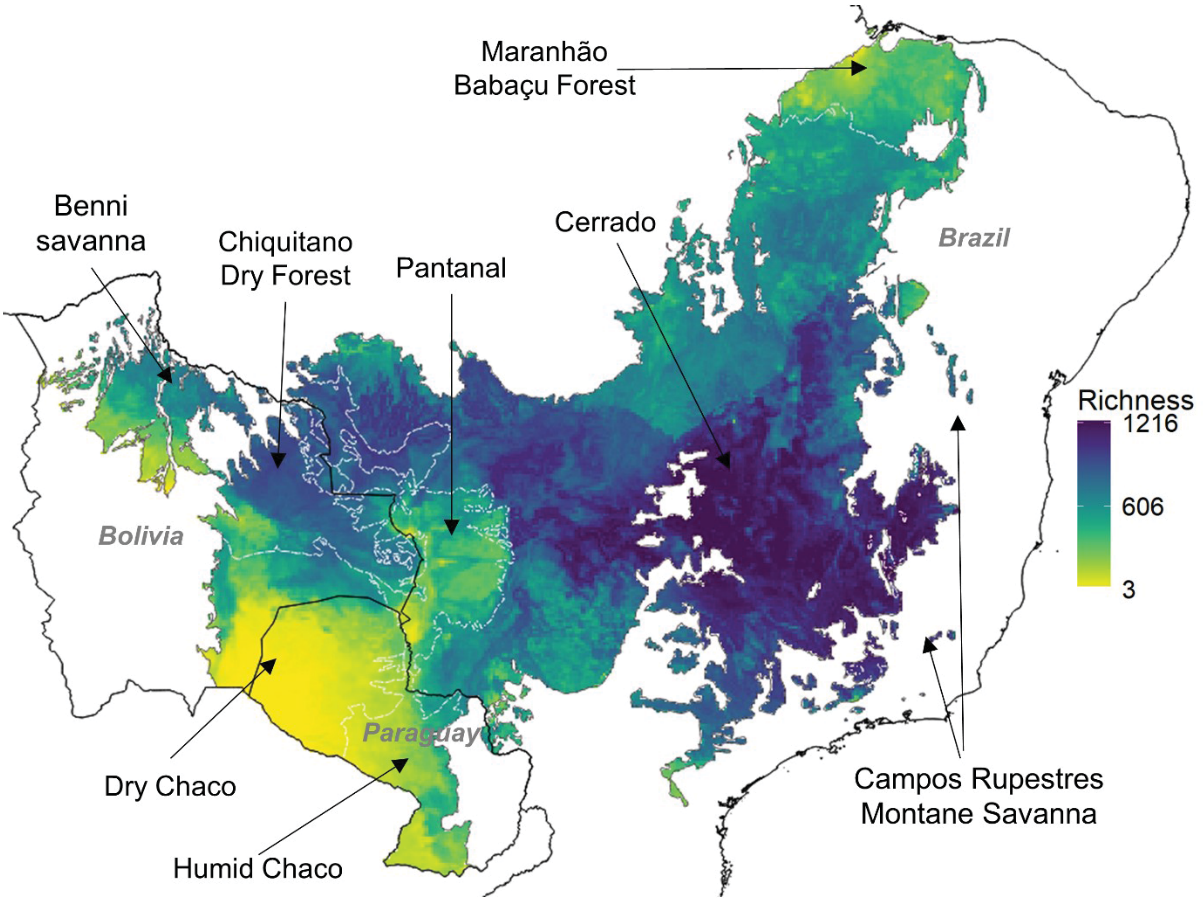


FIGURE 2 - Richness pattern of the Cerrado plant species in Bolivia, Brazil, and Paraguay
SOURCE: the author (2018)

Patterns of global and regional rarity highlighted different important areas within the Cerrado biome across nations. The area with the highest concentration of global rarity was consistent with regions that are also usually the richest in species, such as the central-eastern portion of Brazilian Cerrado, from which there is a decline in rarity to the north, west and south margins of the biome. In Bolivia and Paraguay, an abrupt reduction of global rarity is observed in the Pantanal, Dry Chaco and southwest of the Beni Savannas. With respect to regional rarity, in Brazil, this presents a similar pattern to global rarity. However, in Bolivia the highest values of regional rarity are found in the northeastern region of the Chiquitano Dry Forest, with a decline towards the Dry Chaco region. In Paraguay, the greatest regional rarity is observed in the northern portion of Humid Chaco and Dry Chaco, as well as in the Cerrado ecoregion in eastern Paraguay (FIGURE 3a and b). Overall, global and regional rarity had a positive exponential relationship with species richness in the three countries (FIGURE 3c and d; see APPENDIX 11 for more detailed information about global rarity index and its component for each species).

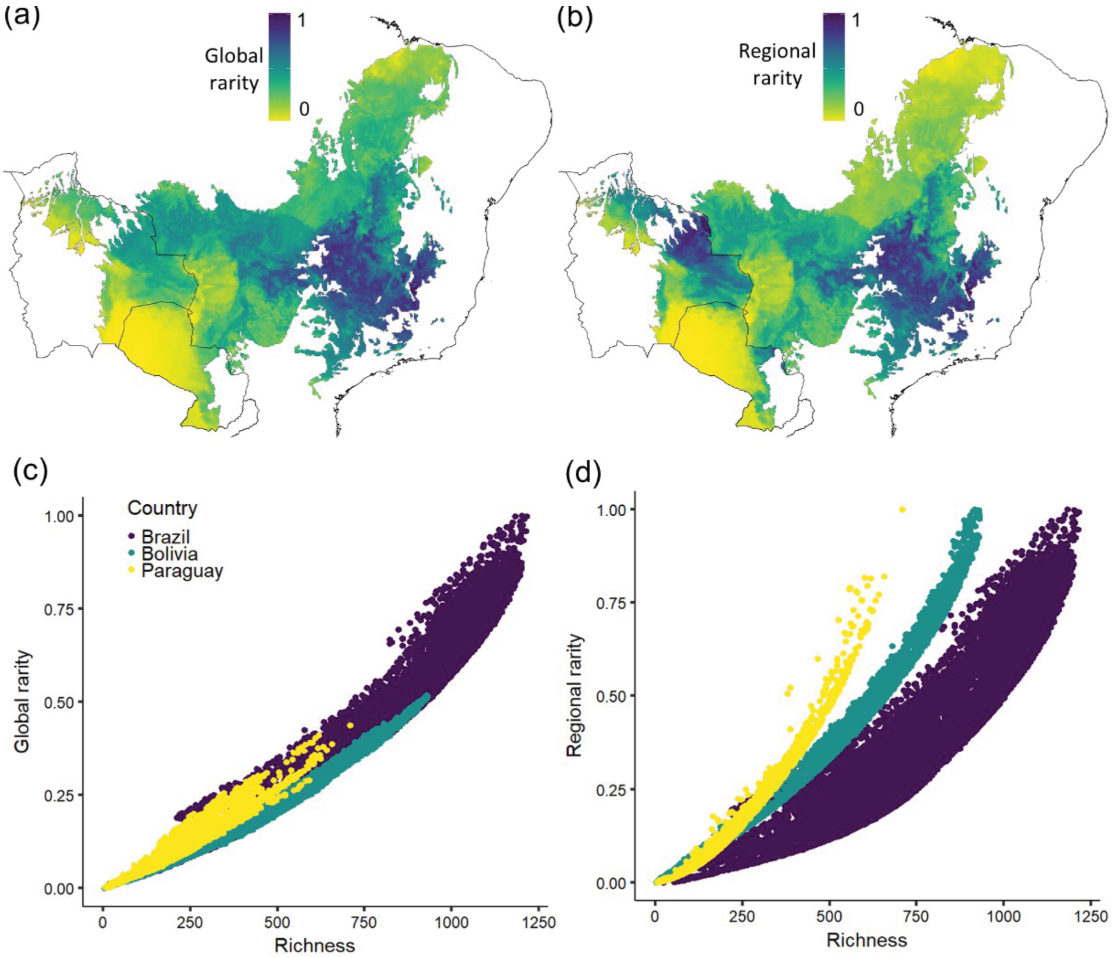


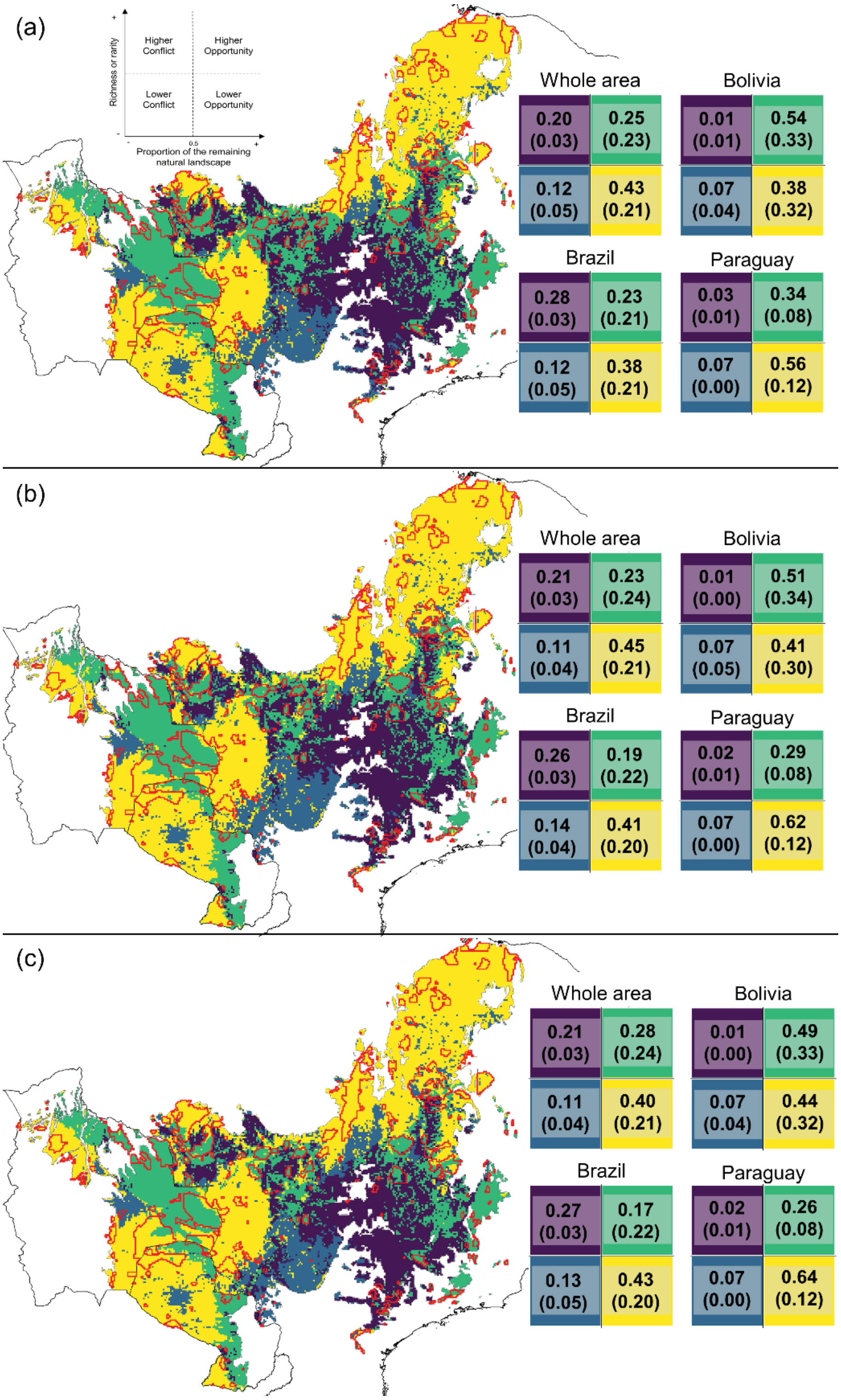
FIGURE 3 - Global and regional rarity of Cerrado's plant species for Bolivia, Brazil y Paraguay (panel a and b respectively). The relationship between species richness and global and regional rarity (panel c and d). Points in panels c and d correspond to cells in the maps
SOURCE: the author (2018)

## 3.2 LAND COVER AND OPPORTUNITY FOR CONSERVATION

The largest extensions of disturbed areas (i.e. cells with <0.5 of remaining natural cover) were found in the central region of the Cerrado ecoregion in Brazil, where the greatest predicted species richness, global and regional rarities were also found (APPENDIX 12). Thus, approximately 27% of the Brazilian territory is under conflict, i.e. areas that could have been priority for conservation but currently are intensively anthropized (FIGURE 4a, b, and c; and APPENDIX 12). Regarding the areas of greatest conservation opportunity for richness and rarity, they represent between 17-22% of the Brazilian territory mainly concentrated in Minas Gerais and northern Goiás states. The relationship between opportunity for conservation and disturbed areas are different for Bolivia and Paraguay because the most disturbed areas are consistent with less rarity and richness. In the case of Bolivia, the most anthropized areas are in the central-eastern sector and some regions in the Beni Savanna. Between 49-53% of this country presents a high proportion of richness and rarity and with low alteration of the natural cover. In Paraguay, the areas with the greatest cover loss were concentrated in the central sector of the western region. 26 to 34% of this nation are promising to conserve the highest values of richness and both kinds of rarity (FIGURE 4). Approximately 24% of the whole study area with high conservation opportunity is into protected areas, however, this proportion varies within each country, with Bolivia showing the highest value and Paraguay the lowest (see values into parentheses in FIGURE 4).

The largest extensions of disturbed areas (i.e. cells with remaining natural cover less than 0.5) were found in the central region of the Cerrado ecoregion in Brazil, where the greatest predicted species richness, global and regional rarities were also found. Thus, approximately 27% of the Brazilian territory is under conflict with areas that may have been a priority for conservation of Cerrado vegetation but currently are intensively anthropized (FIGURE 4a, b, and c). Regarding the areas of greatest conservation opportunity for richness and rarity (i.e. the top right-hand panel in FIGURE 1), they represent between 17-22% of the Brazilian territory mainly concentrated in Minas Gerais and northern Goiás states. The relationship between opportunity for conservation and disturbed areas are different for Bolivia and Paraguay because the areas with the greatest loss of natural cover are found in regions with less rarity and species richness. In the case of Bolivia, the most anthropized areas were found in the central-eastern sector and some regions in the Beni Savanna. Between 49-53% of this country presents a high proportion of richness and rarity and with low alteration of the natural cover. In Paraguay, the areas with the greatest cover loss were concentrated in the central sector of the western region. 26 to 34% of this nation are promising to conserve the highest values of richness and both kind of rarity (FIGURE 4). Given the observed relationships between disturbed areas with the richness and rarity, Brazil presented the largest number of taxa in which more than 50% of their national distribution is under highly altered areas, whereas in Bolivia and Paraguay this number is reduced to two and five species respectively.

Most of the species were mainly distributed within Brazil and around 765 species had >90% of their predicted distribution within this country. Most of species from Bolivia and Paraguay was shared with Brazil. Bolivia is the second country where many species were mainly distributed (FIGURE 5). With respect to loss of distribution (species distribution area which is coincident with areas that lost >50% of their natural cover; FIGURE 5), in Brazil more than 1500 species could have lost more than 80% of their original distribution. Nevertheless, Brazil was still the country where most species had their greatest remaining distribution (FIGURE 5). In the case of Bolivia and Paraguay, species generally lost <10% of their national distribution. Because most of the distribution losses occurred in Brazil, and the fact that many Brazilian species are shared with Bolivia, leads to an increase in the relative national distribution of species in this last country (FIGURE 5; see APPENDIX 11 for more detailed information about the original, remnant and lost range for each species).



(a)

Whole area

| | |
|---|---|
| 0.20 (0.03) | 0.25 (0.23) |
| 0.12 (0.05) | 0.43 (0.21) |

Bolivia

| | |
|---|---|
| 0.01 (0.01) | 0.54 (0.33) |
| 0.07 (0.04) | 0.38 (0.32) |

Brazil

| | |
|---|---|
| 0.28 (0.03) | 0.23 (0.21) |
| 0.12 (0.05) | 0.38 (0.21) |

Paraguay

| | |
|---|---|
| 0.03 (0.01) | 0.34 (0.08) |
| 0.07 (0.00) | 0.56 (0.12) |

(b)

Whole area

| | |
|---|---|
| 0.21 (0.03) | 0.23 (0.24) |
| 0.11 (0.04) | 0.45 (0.21) |

Bolivia

| | |
|---|---|
| 0.01 (0.00) | 0.51 (0.34) |
| 0.07 (0.05) | 0.41 (0.30) |

Brazil

| | |
|---|---|
| 0.26 (0.03) | 0.19 (0.22) |
| 0.14 (0.04) | 0.41 (0.20) |

Paraguay

| | |
|---|---|
| 0.02 (0.01) | 0.29 (0.08) |
| 0.07 (0.00) | 0.62 (0.12) |

(c)

Whole area

| | |
|---|---|
| 0.21 (0.03) | 0.28 (0.24) |
| 0.11 (0.04) | 0.40 (0.21) |

Bolivia

| | |
|---|---|
| 0.01 (0.00) | 0.49 (0.33) |
| 0.07 (0.04) | 0.44 (0.32) |

Brazil

| | |
|---|---|
| 0.27 (0.03) | 0.17 (0.22) |
| 0.13 (0.05) | 0.43 (0.20) |

Paraguay

| | |
|---|---|
| 0.02 (0.01) | 0.26 (0.08) |
| 0.07 (0.00) | 0.64 (0.12) |

FIGURE 4 - Geographic overlap between disturbed areas and the patterns of predicted species richness (a), global (b) and regional (c) rarity. Right-hand panels depict the proportion of each category of the proposed framework for conservation defining high conflict (violet), high opportunity (green), low conflict (blue) and low opportunity (yellow) for different extent. Values in parentheses indicate the proportion of each conservation category into protected areas. Red-colored polygons in the maps represent protected areas

SOURCE: the author (2018)

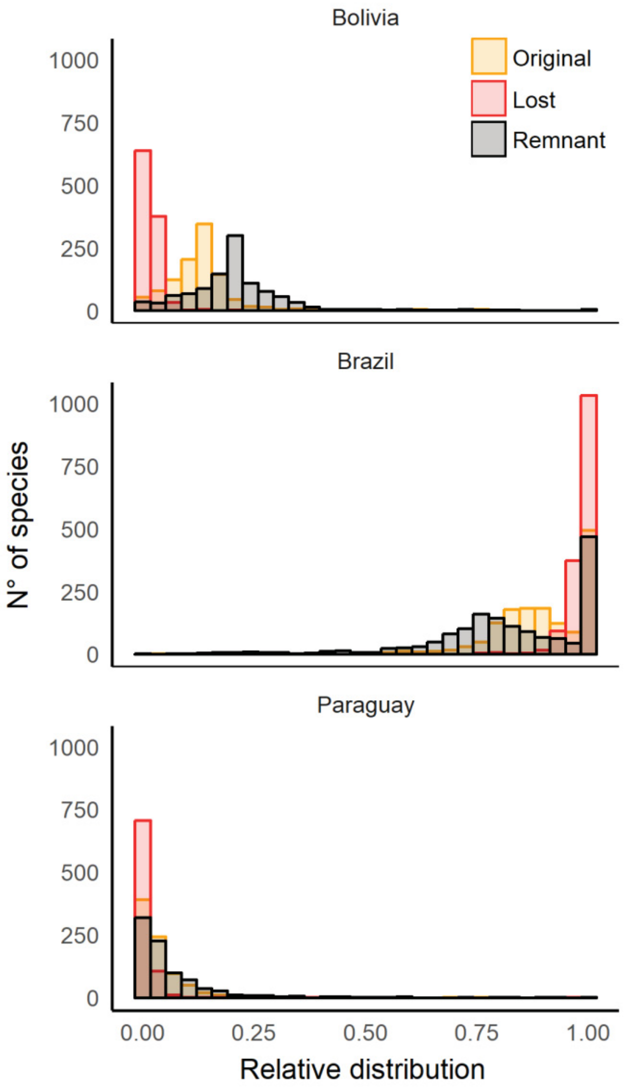


FIGURE 5 -Frequency histogram (number of species) of different relative distribution for Bolivia, Brazil, and Paraguay. The original (yellow), lost (red) and remnant (black) species range were relativized by the proportion of the range size represented within each country

SOURCE: the author (2018)

# 4. DISCUSSION

In this study, we modeled the distribution of 1559 plant species of the Cerrado for three countries, Bolivia, Brazil and Paraguay and described richness and rarity patterns at different geographical levels as well as their relationship with the remnant natural landscape to detect potential conflicts and opportunities for Cerrado conservation. We found that the relationship between cells with the greatest landscape disturbance and those with the greatest species richness and rarity differed among countries. For instance, Brazil showed the highest proportion of its land in high conflict conditions for conservation of Cerrado plants, since the most anthropized areas overlap with regions predicted with the highest concentration of richness as well as global and regional rarities. Conversely, high opportunities for conservation were found in Bolivia and Paraguay, given that the most altered areas were in regions with less predicted species richness and rarity. Also, we found that the areas with the greatest concentration of richness, global and regional rarity are distributed in the central part of the Cerrado ecoregion in Brazil. Bolivia was the second country in species richness, where most species and rarity were distributed in the ecoregions of Cerrado and Chiquitano Dry Forest. Paraguay was the third richest country, with the largest number of species and rarity concentrated in Cerrado ecoregions, in the northern of the Dry Chaco and the extreme northeast of the Humid Chaco.

## 4.1 RICHNESS AND RARITY OF CERRADO PLANT SPECIES

Even though stacked-SDMs do not consider community assembly rules and may overestimate species richness within a region (GUISAN; RAHBEK, 2011), our predicted richness patterns were consistent with the description of several previous studies, including some conducted in Bolivia and Paraguay. Indeed, richness patterns described here, such as the one covering the central-eastern Brazilian Cerrado, were consistent with other studies of Cerrado woody species (AMARAL et al., 2017; BRIDGEWATER; RATTER; RIBEIRO, 2004; FRANÇOSO; HAIDAR; MACHADO, 2016; RATTER; BRIDGEWATER; RIBEIRO, 2003) as well as with other clades such as reptiles and amphibians (DINIZ-FILHO et al., 2009). However, it is important to note that some of these studies highlight as equally species rich, or richer, those areas that are in transition zones with other neighboring biomes (FRANÇOSO; HAIDAR; MACHADO, 2016; RATTER; BRIDGEWATER; RIBEIRO, 2003). This pattern was not observed in our study, possibly because we did not model all the species listed for the Cerrado, precisely because many of them are associated with other biomes and thus may highlight patterns that are not representative of plant species strictly associated with the Cerrado.

The high concentration of predicted richness and rarity in the central-eastern region of the Brazilian Cerrado could be explained by the historical climate stability hypothesis (JANSSON, 2003; WERNECK et al., 2012). According to this hypothesis, areas with high climatic stability, in terms of geological time-scales, would favor greater persistence and speciation, consequently establishing areas of concentration of endemism and richness, which would not occur in areas of more unstable climate (GRAHAM; MORITZ; WILLIAMS, 2006; JABLONSKI; ROY; VALENTINE, 2006; WERNECK et al., 2012). In this sense, areas highlighted here as having the greatest species richness and global rarity were coincident with those characterized as having high climatic stability throughout the Quaternary (COLLEVATTI et al., 2012; SOUZA et al., 2017; WERNECK et al., 2012), mainly in the central Cerrado region, the Espinhaço mountain chain, and even the savannas of Bolivia (WERNECK et al., 2012). The Campo Rupestre Montane Savanna is limited to the presence of rock outcrops isolated to mountains of the Brazilian shield. Such azonal vegetation is controlled by specific edaphic conditions as well as climate and it has been highlighted by its megadiversity, represented by several rare and endangered species (ALVES et al., 2014; ALVES; KOLBEK, 2010; MARTINELLI; MESSINA; SANTOS FILHO, 2014). A large part of the Pantanal is seasonally flooded, which in combination with the topography, shape the mosaic of its vegetation formation (PRANCE; SCHALLER, 1982). Thus, the edaphic conditions in this ecoregion may be responsible for its observed species richness and rarity patterns. The low predicted richness of the Pantanal is consistent with the geographical importance of Cerrado formations within this region, which cover c. 36% (SILVA et al., 2000). However, the Pantanal showed low global rarity that resulted from the presence of widely distributed species and low endemism (POTT et al., 2011; PRANCE; SCHALLER, 1982).

Recently, new regions were proposed as part of the Cerrado ecoregion in Bolivia and Paraguay (MERELES et al., 2013; WERNECK et al., 2012). In fact, Bolivia represents an important source for the conservation of the Cerrado vegetation because it harbors several endemic and threatened plant species (MAMANI et al., 2010, 2011). There are different delimitations and classifications of the Cerrado domain in this country (VILLARROEL; MUNHOZ; PROENÇA, 2016), the majority of areas pointed out in this study as being more species-rich were located in the regions of *Cerrado Chiquitano*, *Cerrado Chaqueño* and *Bosque Seco Chiquitano* according to IBISCH et al (2003) and the Cerrado region proposed by VILLARROEL et al. (2016). It is interesting to note that numerous savannah formations were recently detected in the eastern sub-Andean region (VILLARROEL, 2017), which coincide with the richest areas of the western Dry Chaco. In the case of Paraguay, it ranks third in number of Cerrado plant species. The areas with the greatest richness and rarity were concentrated in Eastern Paraguay. However, a high richness was predicted towards the south-east of this country where the presence of Cerrado formations is well known (MERELES, 2005, 2013). Probably, Cerrado formations could be found in the Alto Paraná Atlantic Forest mainly on sandstone formations, given that areas of vegetation communities similar to those of the Cerrado have been found in the Misiones province of Argentina (VELAZCO et al *in pres*). In western Paraguay, predicted richness and rarity were limited to the extreme north of the region below what is called Cerrado Chaqueño in Bolivia. This region has been described as Cerrado vegetation in transition with the Chaco vegetation (NAVARRO; MOLINA; DE MOLAS, 2006; NAVARRO; MOLINA; VEGA, 2011). However, only recently it has been proposed as part of the Paraguayan Cerrado ecoregion (MERELES, 2013; MERELES et al., 2013). The northern region of western Paraguay also supports the notion that the determinants of different plant formations are associated with clay content, cation exchange and drainage of soil (NAVARRO; MOLINA; DE MOLAS, 2006; NAVARRO; MOLINA; VEGA, 2011). The low richness and rarity predicted towards the southern portion of western Paraguay could be limited by the scarcity of precipitation and presence of loamy-sandy xeric soils (SPICHIGER; CALENGE; BISE, 2004).

## 4.2 GLOBAL AND REGIONAL RARITY

Accounting for both spatial extents of a species rarity, global and regional, may be useful to detect areas for conservation purpose when considering many geopolitical units. Indeed, we found that the pattern of species rarity did not differ in Brazil when the total range of the species or their national distribution was considered. This is because most species had most of their distributions within this country. The extensive area with high rarities (global and regional) suggests that many species that have limited geographic ranges are present in the central Brazilian region, being the central-east of Cerrado and Campo Rupestre Montane Savanna highlighted as the areas with higher concentration of rarity. The significant difference between global and regional rarities in Bolivia and Paraguay compared to that in Brazil demonstrates that many of Cerrado species are marginal to these countries, i.e. have their distributional limits within these countries (FIGURE 4 and 5).

Usually, populations in the edge of a species distribution are avoided when planning conservation strategies (CHANNELL; LOMOLINO, 2000). However, these populations can be essential to conservation (BATEMAN et al., 2015; CHANNELL; LOMOLINO, 2000; LEPPIG; WHITE, 2006). Peripheral populations can serve as refugia from extreme weather events (BATEMAN et al., 2015) or possess genetic distinctiveness (HUNTER; HUTCHINSON, 1994; LEPPIG; WHITE, 2006). Under future climate change, these populations can be the most susceptible to extinction or, alternatively, persist in the periphery and be the starting point for emigrations to regions still inhabited (CHANNELL; LOMOLINO, 2000). Moreover, it is important to consider the value that a nation confers to the presence of a species within its boundaries regardless if it occurs outside of it. Such species can have utilitarian or cultural values for local inhabitants; their regional rarity can lead them to be considered as natural monuments; or can even harbor an ethical responsibility to the fact that they can become extinct from a nation.

## 4.3 LOSS OF NATURAL COVER AND IMPLICATIONS FOR THE CONSERVATION OF CERRADO PLANT SPECIES

When a biome or group of species is distributed in areas where resources of economic interest are found, a dispute over the space occupied by those resources can arise. In South America, the open vegetation formation is the habitat type under the greatest threat by fire and grazing (JARVIS et al., 2010). Based on the relationship between the most disturbed areas and the most attractive ones to be conserved (i.e. highest richness and/or rarity), we found that in the three countries there are different conflict scenarios. Most of the areas with the highest concentration of species and rarity in Brazil is under high anthropic pressure, and therefore it is the country with the highest value of conflict conditions. This was different for Bolivia and Paraguay, as they presented proportionally higher values of conservation opportunities than Brazil (see FIGURE 4). The Cerrado has less than 20% of its area undisturbed (STRASSBURG et al., 2017) and under 7.7% of its extent under protection (OLIVEIRA et al., 2017). Agribusiness is one of the main factors in altering the natural cover (OVERBECK et al., 2015; STRASSBURG et al., 2017), which also affects the Pantanal because cattle ranching, deforestation, and plantation of artificial pastures (JUNK; CUNHA, 2005).

Ecoregions in Bolivia and Paraguay also present strong and rapid land use change. The Chiquitano Dry Forest is one of the Bolivian ecoregions with the highest deforestation rates as a consequence of agriculture, highways, and mining (SALAZAR et al., 2015). Regarding the Beni ecoregion, the strongest vegetation loss has been carried out in the eastern region also by agriculture (REDO; AIDE; CLARK, 2012). In the case of Paraguay, large extensions of the Dry Chaco were transformed to croplands and pastures in the last decades (SALAZAR et al., 2015; VALLEJOS et al., 2015). Despite the negative situation for several of the ecoregions in Bolivia and Paraguay, the large transformed areas still do not overlap with the richest regions of Cerrado species nor those with high global and regional rarity. This may be attributed to two factors, one of them is because the agricultural frontier simply has not yet reached those areas, and the other is related to topographical factor, since several of these areas present rugged topographical conditions unfavorable for mechanization.

We observed that the areas with the higher opportunities for conservation of Cerrado plant species within Bolivia and Paraguay are in regions near to Brazilian boundaries. This could be an artifact caused by having conducted the assessment of conflicts and opportunities for conservation within nations. However, such boundary patterns are also common for the distribution of protected areas, since they are frequently biased to proximity with international borders (BALDI et al., 2017; MOILANEN et al., 2013), as some protected areas in Bolivia and Paraguay. This may be caused by several factors, such as the different management and land-use practice on both sides of the border (SINTHUMULE, 2016); or to assert sovereignty on a territory (MUBOKO, 2017). Despite several studies pointing out the pros and cons of transboundary conservation practice (HUNTER; HUTCHINSON, 1994; KARK et al., 2015; LÓPEZ-HOFFMAN et al., 2010), its implementation requires that borderlands are treated with political, sovereignty, social acceptance and legal arrangements among other issues (BARQUET; LUJALA; RØD, 2014; LIM, 2016). In countries such as the ones studied here, the poor governance in environmental terms, including limited conservation incentives, land tenure conflicts, absence of land-use planning, relaxation of environmental laws and poor law enforcement (LE POLAIN DE WAROUX et al., 2016; OLIVEIRA et al., 2017; STRASSBURG et al., 2017; TEJADA et al., 2016; VALLEJOS et al., 2015), could lead to intergovernmental conservation actions being difficult to implement.

## 5. CONCLUSION

To the extent of our knowledge, this is the first attempt to describe the plant diversity of the Cerrado biome beyond Brazilian political boundaries. We highlight the importance of carrying on practical and theoretical studies of species distribution beyond national limits.

The areas that showed the greatest concentration of predicted species richness and global rarity are found in the central region of the Cerrado ecoregion in Brazil. Our models highlighted Bolivia as the second country with the highest richness of Cerrado plant species with their distributions being mainly associated with Cerrado and Chiquitano Dry Forest ecoregions. In Paraguay, important areas of concentration of richness and rarity were found in the north-west of Humid Chaco and north of Dry Chaco.

Despite much of the Cerrado flora being concentrated in Brazil, this country also has the greatest proportion of its territory under a conflict scenario where the major landscape alteration happens in areas with the highest species richness and rarity concentration. However, Bolivia and Paraguay harbor regions with high opportunity for conservation, because the predicted species richness and rarity are in areas that until now do not present the degree of alteration of Brazil, which makes them important countries to conserve the flora of Cerrado.

## 6. SUPPORTING INFORMATION

### APPENDIX 1 - Search criteria used in the different sites to list Cerrado's plant species for Bolivia, Brazil, and Paraguay

Brazilian Flora 2020

- Listar só nomes aceitos. Angiospermas, forma de vida = Árvore, ocorre no domínio fitogeográfico = Cerrado, só não endêmicas do Brasil, Nativa, Buscar até = espécies.
- Listar só nomes aceitos. Angiospermas, forma de vida = Arbusto, ocorre no domínio fitogeográfico = Cerrado, só não endêmicas do Brasil, Nativa, Buscar até = espécies.
- Listar só nomes aceitos. Angiospermas, forma de vida = Subarbusto, ocorre no domínio fitogeográfico = Cerrado, só não endêmicas do Brasil, Nativa, Buscar até = espécies.
- Listar só nomes aceitos. Angiospermas, forma de vida = Subarbusto, ocorre no domínio fitogeográfico = Cerrado, só não endêmicas do Brasil, Nativa, Buscar até = espécies.

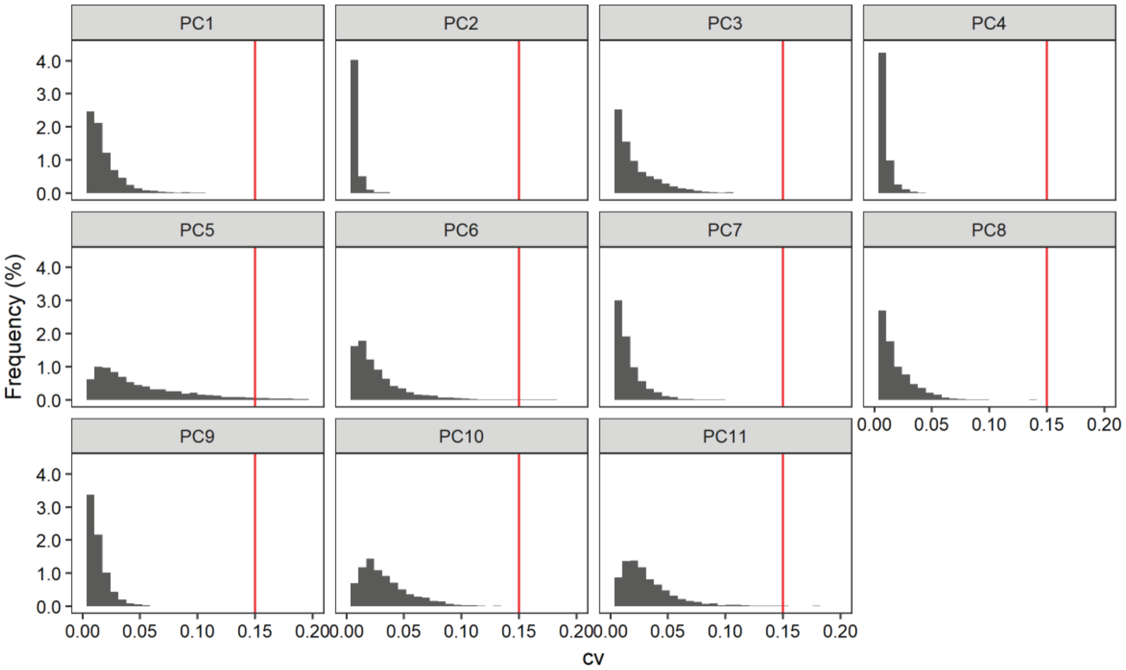
Bolivian Checklist

- Listar nombres de especies nativas con forma de vida como Arboles y Arbolitos que ocurren en zonas de vegetación de campo cerrado
- Listar nombres de especies nativas con forma de vida como Arbustos y Subarbustos que ocurren en zonas de vegetación de campo cerrado
- Listar nombres de especies nativas con forma de vida como hierbas que ocurren en zonas de vegetación de campo cerrado

Paraguay Checklist

- Listar nombres de especies nativas con forma de vida como Arboles que ocurren en zonas de vegetación de Bosque de cerrado o Cerradón, Matorral del cerrado y Matorral rupícola
- Listar nombres de especies nativas con forma de vida como Subarbusto, Arbusto y Arbusto apoyantes que ocurren en zonas de vegetación de Bosque de cerrado o Cerradón, Matorral del cerrado y Matorral rupícola
- Listar nombres de especies nativas con forma de vida como Hierbas erectas y Hierbas postradas que ocurren en zonas de vegetación de Bosque de cerrado o Cerradón matorral del cerrado y Matorral rupícola

APPENDIX 2- Histogram depicting the frequency of Brazilian municipalities for classes of coefficient variation based on the environmental variability within municipalities' boundaries for each principal component



SOURCE: the author (2018)

APPENDIX 3 - Species modeled, their scientific names, family, life form (LF), number of initial occurrences (NIO) and number of cleaned occurrences (NCO; data is available as Excel file)

Data available in: https://figshare.com/s/cda4dc6f8c8c49d78433
SOURCE: the author (2018)

APPENDIX 4 - Climate and edaphic variables used to construct the ecological niche models

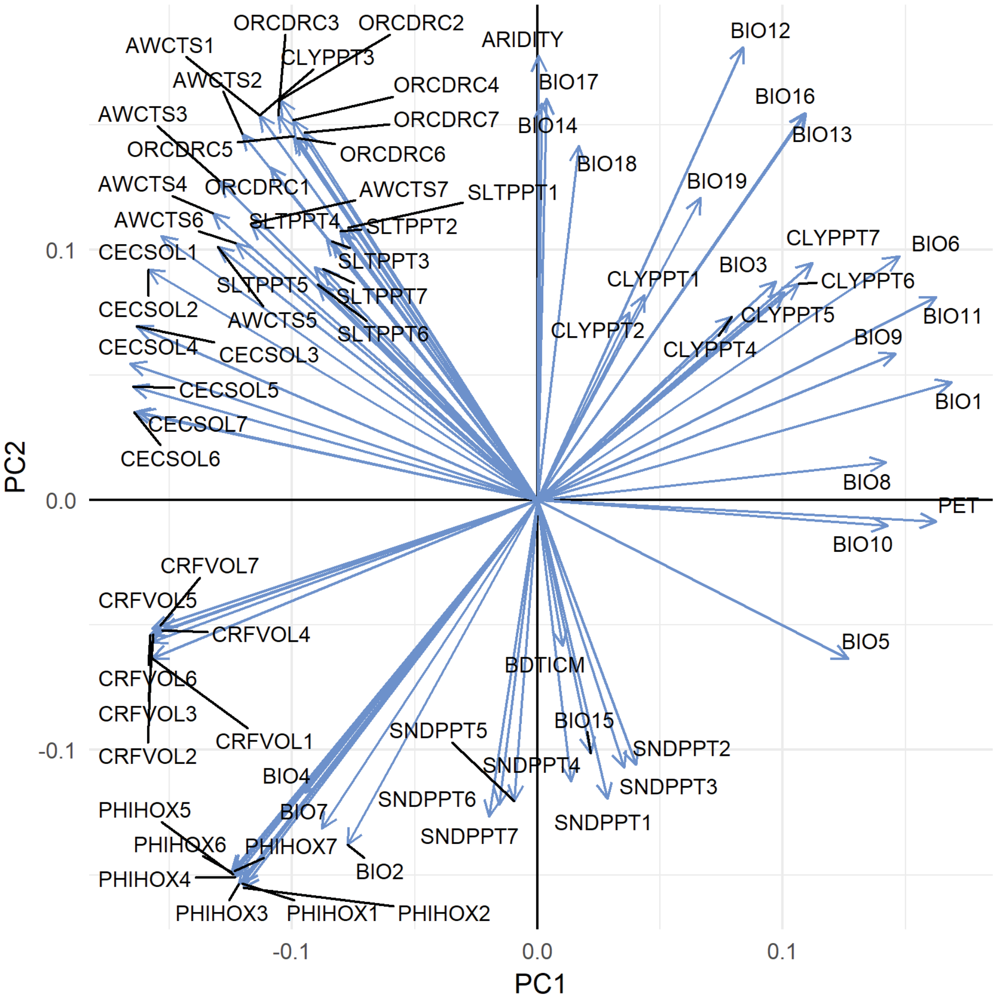
| **Source** | **Variables** | **Units** |
|---|---|---|
| CGIAR-CSI | Annual aridity | - |
| | Annual potential evapotranspiration | mm |
| CHELSA | Annual mean temperature (BIO1) | °C |
| | Mean diurnal range (BIO2) | °C |
| | Isothermality (BIO3) | °C |
| | Temperature Seasonality (BIO4) | % |
| | Max temperature of warmest week (BIO5) | °C |
| | Min temperature of coldest week (BIO6) | °C |
| | Temperature annual range (BIO7) | °C |
| | Mean temperature of wettest quarter (BIO8) | °C |
| | Mean temperature of driest quarter (BIO9) | °C |
| | Mean temperature of warmest quarter (BIO10) | °C |
| | Mean temperature of coldest quarter (BIO11) | °C |

| Source | Variables | Units |
|---|---|---|
| | Annual precipitation (BIO12) | mm |
| | Precipitation of wettest week (BIO13) | mm |
| | Precipitation of driest week (BIO14) | mm |
| | Precipitation seasonality (BIO15) | - |
| | Precipitation of wettest quarter (BIO16) | mm |
| | Precipitation of driest quarter (BIO17) | mm |
| | Precipitation of warmest quarter (BIO18) | mm |
| | Precipitation of coldest quarter (BIO19) | mm |
| SoilGrid | Absolute depth to bedrock (in cm) | cm |
| | Cation exchange capacity of soil * | cmolc/kg |
| | Clay content (0-2 micro meter) mass fraction * | % |
| | Coarse fragments volumetric * | % |
| | Sand content (50-2000 micrometer) mass fraction * | % |
| | Saturated water content (volumetric fraction) for tS * | - |
| | Silt content (2-50 micrometer) mass fraction * | % |
| | Soil organic carbon content (fine earth fraction) * | g/kg |
| | Soil pH x 10 in $H_2O$ * | - |

*Data for seven depths

SOURCE: the author (2018)

APPENDIX 5 - Ordination diagram for the first two axis of three PCAs conducted with environmental variables



SOURCE: the author (2018)

APPENDIX 6 - Principal components selected from the principal component analysis, their eigenvalues, variance and cumulative variance explained

| Principal Components | Eigenvalues | Variance explained for each PC | Cumulative variance explained |
|---|---|---|---|
| 1 | 4.9 | 30.4 | 30.4 |
| 2 | 4.3 | 24.1 | 54.6 |
| 3 | 3.4 | 15.0 | 69.6 |
| 4 | 2.5 | 8.3 | 77.9 |
| 5 | 1.9 | 4.8 | 82.7 |
| 6 | 1.7 | 3.8 | 86.5 |
| 7 | 1.4 | 2.6 | 89.1 |

| | | | |
|---|---|---|---|
| 8 | 1.2 | 2.0 | 91.1 |
| 9 | 1.2 | 1.8 | 92.9 |
| 10 | 1.0 | 1.3 | 94.2 |
| 11 | 1.0 | 1.1 | 95.3 |

SOURCE: the author (2018)

## APPENDIX 7 - Algorithms employed in fitting Ecological Niche Models

GLM is an extension of Linear Models, which allows constructing models with response variables with different distributions (i.e. Poisson, Gamma, etc.). Because the response data assumes values of zero and one, for absences and presences respectively, we used a logistic link function and all single predictor variables without interaction between them. No backward or forward covariates selection was used.

GAM is a non-parametric extension of GLM that replaces the linear relationship between the dependent and independent variables by the sum of a smooth function (HASTIE; TIBSHIRANI, 1986). These models were fitted using a binomial distribution and with the same parameterization covariate as GLM. The Newton method was used to optimize the estimation of the smoothing parameter.

ME is an algorithm extensively used in ENMs. We used the new approach which now is based on inhomogeneous Poisson process procedure (PHILLIPS et al., 2017). We used linear, quadratic, product, and hinge features, default regularization values, 10000 maximum background points and clog-log output format.

The SVM procedure splits samples in the multidimensional space based on a hyperplane. The optimal hyperplane is the one that maximizes the margin of separation between the classes (i.e. support vectors; KUHN; JOHNSON, 2013). We performed these models with a radial basis kernel, with a constant cost value equal to one, and based on probabilities classes.

RF is a modified technic from bagging trees, which builds a model based on the average of a large collection of non-correlated trees. In each node of these trees, a random sample of predictors is chosen as split candidates from the full set of predictors (JAMES et al., 2013). To determine the optimal number of variables randomly sampled at each split, the RF algorithm was tuned automatically with the ‘tuneRF’ function of the randomForest R package. 500 trees were used at the tuning step, with default values of the step factor and the improvement in out-of-bag error parameter. We considered those models with the minimum out-of-bag error as our final RF models.

GS is a Bayesian approach that consists in a generalization of Gaussian probability distribution based on stochastic process. These models were constructed using Laplace approximation (GOLDING; PURSE, 2016).

APPENDIX 8 - Land cover legends, their respective raster values (RV) and new classification used to construct the layer of remaining natural landscape

| Land cover legend | RV | Classification |
|---|---|---|
| No data | 0 | No considered |
| Cropland, rainfed | 10 | Converted |
| Herbaceous cover | 11 | Converted |
| Tree or shrub cover | 12 | Converted |
| Cropland, irrigated or post-flooding | 20 | Converted |
| Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%) | 30 | Converted |
| Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%) | 40 | Converted |
| Tree cover, broadleaved, evergreen, closed to open (>15%) | 50 | Natural cover |
| Tree cover, broadleaved, deciduous, closed to open (>15%) | 60 | Natural cover |
| Tree cover, broadleaved, deciduous, closed (>40%) | 61 | Natural cover |
| Tree cover, broadleaved, deciduous, open (15-40%) | 62 | Natural cover |
| Tree cover, needleleaved, evergreen, closed to open (>15%) | 70 | Converted |
| Tree cover, needleleaved, evergreen, closed (>40%) | 71 | Converted |
| Tree cover, needleleaved, evergreen, open (15-40%) | 72 | Converted |
| Tree cover, needleleaved, deciduous, closed to open (>15%) | 80 | Converted |
| Tree cover, needleleaved, deciduous, closed (>40%) | 81 | Converted |
| Tree cover, needleleaved, deciduous, open (15-40%) | 82 | Converted |
| Tree cover, mixed leaf type (broadleaved and needleleaved) | 90 | Converted |
| Mosaic tree and shrub (>50%) / herbaceous cover (<50%) | 100 | Natural cover |
| Mosaic herbaceous cover (>50%) / tree and shrub (<50%) | 110 | Natural cover |
| Shrubland | 120 | Natural cover |
| Shrubland evergreen | 121 | Natural cover |
| Shrubland deciduous | 122 | Natural cover |
| Grassland | 130 | Natural cover |
| Lichens and mosses | 140 | Natural cover |
| Sparse vegetation (tree, shrub, herbaceous cover) (<15%) | 150 | Natural cover |
| Sparse tree (<15%) | 151 | Natural cover |
| Sparse shrub (<15%) | 152 | Natural cover |
| Sparse herbaceous cover (<15%) | 153 | Natural cover |
| Tree cover, flooded, fresh or brackish water | 160 | Natural cover |
| Tree cover, flooded, saline water | 170 | Natural cover |
| Shrub or herbaceous cover, flooded, fresh/saline/brakish water | 180 | Natural cover |
| Urban areas | 190 | Converted |
| Bare areas | 200 | Converted |
| Consolidated bare areas | 201 | Converted |

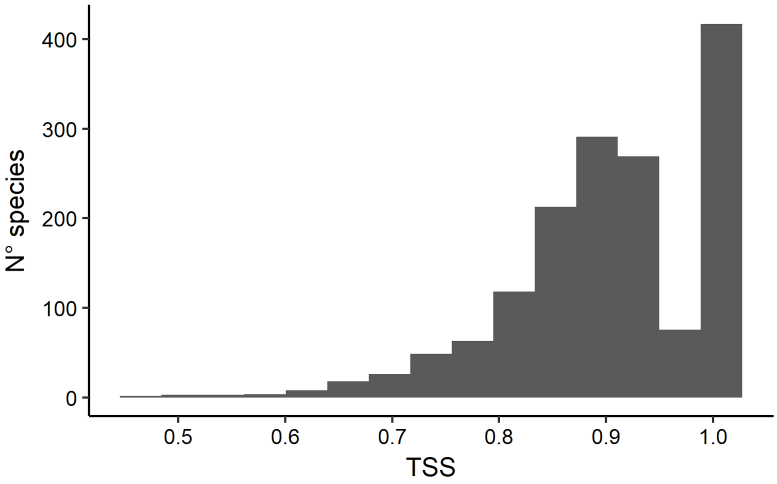
| Land cover legend | RV | Classification |
|---|---|---|
| Unconsolidated bare areas | 202 | Converted |
| Water bodies | 210 | No considered |
| Permanent snow and ice | 220 | No considered |

SOURCE: the author (2018)

## APPENDIX 9 - Protected Areas Network

We collected different protected areas categories for the three countries to construct the protected area network of the study area. Protected areas from Bolivia were sourced by the Servicio Nacional de Areas Protegidas updated to 2015 (available in http://geo.gob.bo), from Paraguay sourced by the Secretaría del Amabiente updated to 2007, from Brazil were sourced by the Ministério do Meio Ambiente updated to 2017, and Fundação Nacional do Índio updated to 2013, these data were sourced by Laboratório de Processamento de Imagens e Geoprocessamento (available in http://maps.lapig.iesa.ufg.br/lapig.html). We construct a raster layer using this protected area dataset by rasterizing it to the resolution of environmental data (i.e. 5 arc-min). We only considered under protection those grid cells that was overlapped by a protected area in a proportion ≥ 10%.

APPENDIX 10 - Model performance of 1559 plant species of Cerrado measured by TSS (True Skill Statistic). Each species value was based on the average of TSS of the best algorithm used to construct ensembled models

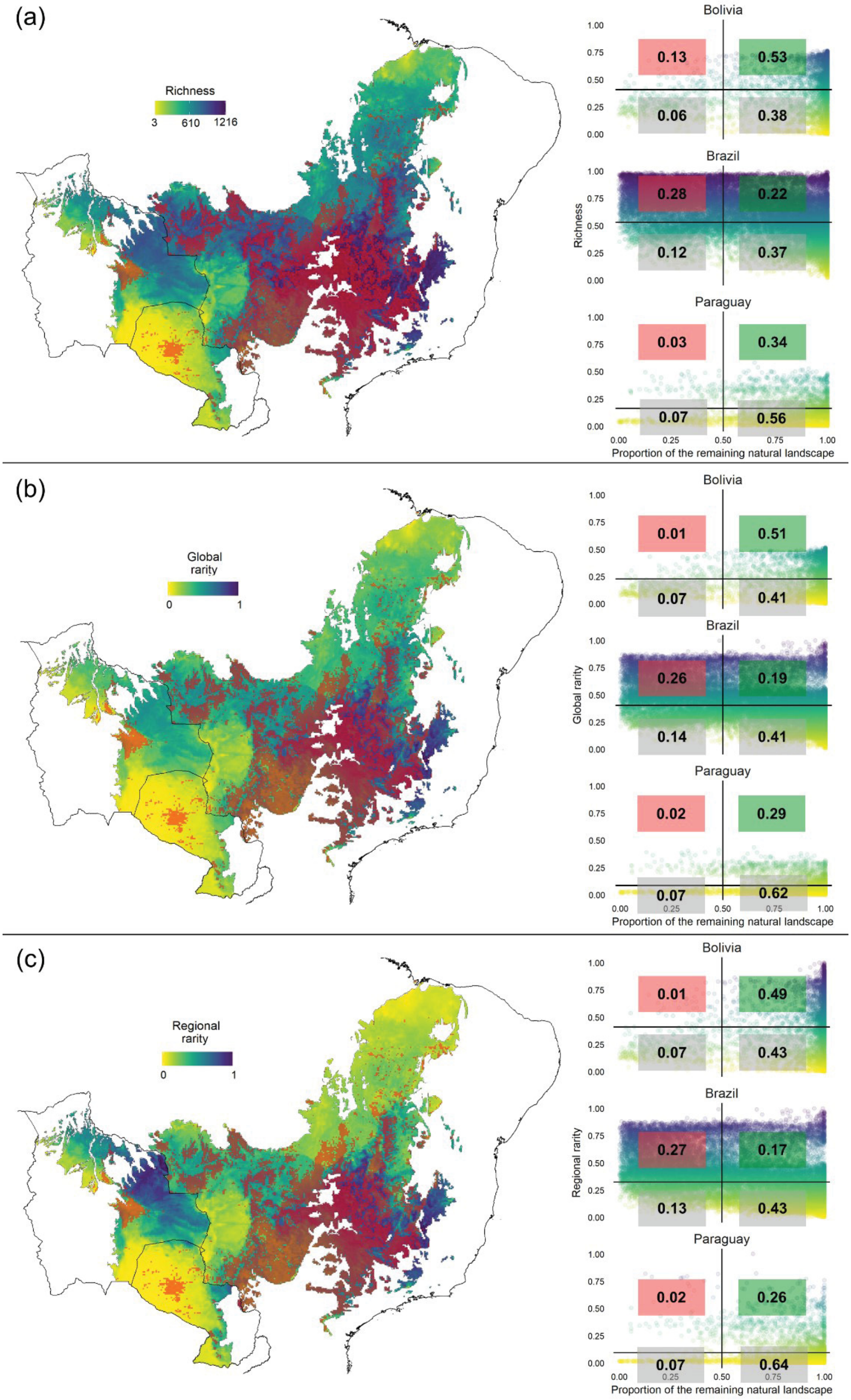


SOURCE: the author (2018)

APPENDIX 11 - Species modeled, global rarity index and its components with standardized values: inverse range size, marginality and specialization; absolute and relative species range regarding the original, remnant and lost area (data is available as Excel file)

Data available in: https://figshare.com/s/d497aee97e8d5bbf910e

APPENDIX 12 - Geographic overlap between disturbed areas and the patterns of predicted species richness (a), global (b) and regional (c) rarity. Transparent red-colored cells in the maps represent disturbed areas (cells with less than 0.50 of natural cover). Right-hand panels depict the observed values within the proposed framework for defining opportunities and conflicts for conservation. In these panels, each point and its respective color is equivalent to a cell and its richness or rarity value in the corresponding left-hand map



SOURCE: the author (2018)

## References

GOLDING, N.; PURSE, B. V. Fast and flexible Bayesian species distribution modelling using Gaussian processes. Methods in Ecology and Evolution, p. n/a-n/a, 2016.

HASTIE, T.; TIBSHIRANI, R. Generalized Additive Models. Statistical Science, v. 1, n. 3, p. 297–318, 1986.

JAMES, G. et al. An Introduction to Statistical Learning. New York, NY: Springer New York, 2013.

KUHN, M.; JOHNSON, K. Applied Predictive Modeling. New York, NY: Springer New York, 2013.

PHILLIPS, S. J. et al. Opening the black box: an open-source release of Maxent. Ecography, v. 40, p. 1–7, 2017.

## CAPÍTULO 3

# A DARK SCENARIO FOR CERRADO PLANT SPECIES, EFFECTS OF FUTURE CLIMATE, LAND-USE AND INEFFICIENCY OF PROTECTED AREAS

**Abstract**: The anthropogenic climate change and land-use change are considered two of the main factors that are altering biodiversity at the global scale. Coping with a new climate and the expansion of productive land will be a challenge for the biological conservation worldwide. Because countries present different patterns of exploitation of their resources, it is necessary to evaluate their impacts on biodiversity in a global and local level. We aimed to (i) assess the effect of the climate change and land-use on the distribution of the Cerrado plant species for different countries where it occurs, (ii) evaluate the efficiency of the current protected area network to safeguards species under different greenhouse-gas emissions and land-use and (iii) estimate the vulnerability of species caused by protection efficiency and habitat loss. We demonstrate that climate change and land-use will cause great damage to Cerrado flora by 2050 and 2080, even under optimistic conditions. Unfortunately, the greatest intensity and extent of land-use will have to overcome on the regions where the greatest richness will be harbored. The conservation of the species will be seriously affected since the protected areas network is not as efficient in safeguarding them under current or future conditions. The low level of protection together with the losses caused by the advance of the agricultural frontier will lead to most species being highly vulnerable. Due to the climate and land-use, effects showed different interactions in each country, conservation strategies should be implemented at transboundary and national levels.

**Keywords:** neotropical savannas, protected area network, species distribution models, vulnerability

# 1. INTRODUCTION

The earth has experimented several climatic fluctuations that occurred naturally throughout geological time, however, climate fluctuations have been recently altered and accelerated by human activities (DIFFENBAUGH; FIELD, 2013). Greenhouse gas (GHG) emissions are the highest in history, putting ecosystems, societies and the economic sector at risk (IPCC, 2014). This anthropogenic climate change is considered one of the main factors altering community composition and ecosystem functioning at the global scale (PECL et al., 2017). However, climate change is not solely responsible for the damage caused to biodiversity. Habitat loss due to anthropic land-use is another important factor with an effect on biodiversity (NEWBOLD et al., 2016). The increase of human demand for food and energy has led to the conversion of large areas with natural cover to productivity land (ALEXANDRATOS; BRUINSMA, 2012) and, consequently, most of the world's land area is now biologically compromised (FOLEY et al., 2005; NEWBOLD et al., 2016). The effects of both climate change and loss of natural cover compromise not only biological diversity but also human well-being (HAUTIER et al., 2015; PECL et al., 2017). Coping with a new climate and the expansion of productive land will be a challenge for conserving our planet's biodiversity.

From an anthropocentric point of view, there is no doubt about all the benefits that biodiversity brings to human beings and therefore the importance of conserving it (PEARSON, 2016). Nonetheless, conservation actions may be compromised or limited by economic interests, therefore exist a territorial dispute between both (MARGULES; PRESSEY, 2000). In order to halt the loss of biodiversity international conservation targets and agreements have been established, such as the United Nation’s Sustainable Development Goals, which, among other objectives, promote the conservation, sustainable use, restoration and protection of the terrestrial ecosystem (UNITED NATIONS, 2015); or the Convention on Biological Diversity[22] which integrates biodiversity conservation, economic and social activities. An appropriate management of protected areas and its strategic expansion are two of the main tools to achieve these international goals (LE SAOUT et al., 2013). However, the factors that determine protected area allocation are not necessarily

---

[22] https://www.cbd.int/sp/targets/#GoalA

based on ecological criteria, hence they suffer strong biases (BALDI et al., 2017; MARGULES; PRESSEY, 2000), which can compromise the efficiency to protect some species (GRAY et al., 2016). The fact that the geographic distribution of species and ecosystems are subject to alterations induced by the climate change can lead to existing protected areas being inefficient in conserving biodiversity (MONZÓN; MOYER-HORNER; PALAMAR, 2011). In addition, the demand of space for future land-uses could jeopardize the expansion of protected areas (POUZOLS et al., 2014).

The geographic distribution of species results mainly from its tolerance to environmental conditions and its ability to reach suitable areas (SOBERÓN, 2007). Consequently, its occurrence in a specific region could be at least partially predicted based on its environmental conditions (PETERSON et al., 2011). Moreover, species distributions are dynamic over time owing either to expansion, shift or contraction of their ranges as a response to climatic fluctuations (GASTON, 2003). Species may respond to climate change in different ways; they can shift their geographical distribution, adapt to new conditions, change their abundance patterns or become extinct (JEZKOVA; WIENS, 2016; LENOIR; SVENNING, 2015). The capacity of species to track environmental changes is limited by several factors such as life cycle, dispersal capacity, biotic interaction or adaptability (FEI et al., 2017; PECL et al., 2017), and can respond idiosyncratically or grouped by phylogenetic or trait characteristics (FEI et al., 2017; FISICHELLI et al., 2014; PUCKO et al., 2011). According to the latter response, communities are also expected to change and consequently their resilience capacity and ecosystem services (PECL et al., 2017). Moreover, climate change interacts with land-use change, which can have synergic effect on diversity (OLIVER; MORECROFT, 2014). Landscape characteristics such as the degree of connectivity, fragmentation, and edge density could affect the persistence of a species population (FAHRIG, 2001; SWIFT; HANNON, 2010). The loss and degradation of habitats influence on abundance and richness of species (CHAUDHARY; PFISTER; HELLWEG, 2016; NEWBOLD et al., 2015), reduce ecosystem productivity (HAUTIER et al., 2015), and decrease viability population (MENDES; DE MARCO, 2017).

Proper management and conservation of biodiversity should take into account the impact of the climate change and trends in land-use (PECL et al., 2017;

POUZOLS et al., 2014). Explicitly considering future threats and species sensitivity to these is crucial towards implementing more effective conservation actions, such as identifying areas for species reallocation (PAYNE; BRO-JØRGENSEN, 2016) or to assist in ex-situ conservation by designing seed-banks (SMITH; LONG; ALBRECHT, 2016). Species vulnerability to climate can be evaluated using a mixture of spatial and demographic data variables (e.g. dispersal ability, population size, vital rates, KEITH et al., 2014; PEARSON et al., 2014). However, in highly diverse areas, these kinds of data are difficult to obtain for large groups of species. In such cases, ecological niche models (ENMs) are practical tools to estimate species ranges (PETERSON; NAVARRO-SIGÜENZA; GORDILLO, 2016), and assess the effect of climate change on species geographic distributions (PETERSON et al., 2011). There is evidence that this ENM approach can be effective for evaluating species vulnerability to climate change, even in its simplest form based on correlating species presences and climate variables without coupling these models with dispersion or population dynamic models (FORDHAM et al., 2017).

South America stands out for hosting ecoregions with high biodiversity and endemism (e.g. Tropical Andes, Atlantic forest or Cerrado) and, at the same time, for having high rates of natural cover loss, which have resulted in highly fragmented and anthropized ecosystems (FEHLENBERG et al., 2017; LE POLAIN DE WAROUX et al., 2016; OVERBECK et al., 2015; STRASSBURG et al., 2017; TEJADA et al., 2016). A clear example of this situation is the Cerrado ecoregion. Outstanding by its high diversity and degree of endemism, the Cerrado is considered as one of the most diverse Neotropical savannas (SILVA; BATES, 2002). Essentially located on the Central Plateau in Brazil, this is second largest ecoregion of this country (RATTER; RIBEIRO; BRIDGEWATER, 1997). Cerrado also expands to two neighboring nations, Bolivia and Paraguay, being present in the Dry Chaco, Chiquitano Dry Forest, Beni Savanna and Pantanal ecoregions (BECK, 2015; IBISCH et al., 2003; MERELES, 2013; VILLARROEL; MUNHOZ; PROENÇA, 2016). In Brazil, the advance of agribusiness on large areas has caused the rapid disappearance of ecosystems and habitats characteristic of the Cerrado, leaving it as a highly threatened and fragmented ecoregion with < 20% of its remaining area undisturbed (STRASSBURG et al., 2017). Furthermore, this ecoregion suffer of low protection because only 7.7% of its surface is under protection (OLIVEIRA et al., 2017). Unfortunately, the other two countries that can preserve part of the Cerrado flora, Bolivia and Paraguay, are also currently vulnerable given the rapid change of their natural cover (REDO; AIDE; CLARK, 2012; SALAZAR et al., 2015, 2015; VALLEJOS et al., 2015).

Because countries present different patterns of exploitation of their resources (ARMENTERAS et al., 2017), it is necessary to evaluate the impact of such activities on biodiversity at global and regional scales. This multiscale assessment combined with the potential effects of future climatic conditions can be relevant to detect which species could be the most vulnerable to climate changes and reveal the regions of highest stability or susceptibility to biodiversity loss. Here, we apply such combined assessment in order to (i) determine the effect of climate and land-use change on the distribution of Cerrado's plant species across the different countries where this biome occurs (Bolivia, Brazil and Paraguay), (ii) evaluate the efficiency of the current protected area network to safeguard species under different GHG emissions and land-use scenarios and (iii) estimate the vulnerability of species caused by protected area efficiency and habitat loss.

## 2. METHOD

### 2.1 STUDY AREA

The study area includes the countries of Bolivia, Brazil, and Paraguay where the Cerrado biome is present. Although Cerrado's boundaries are well defined in Brazil, there are disagreements about its limits in Bolivia and Paraguay. Therefore, we used the WWF's Terrestrial Ecoregions of the World (OLSON et al., 2001) to overcome the methodological differences used to define this biome within each nation. We included ecoregions related to open formations such as steppes and savannas, and others that are not open formations but well known for the existence of Cerrado within them (RIBEIRO; WALTER, 2008). Therefore, the study area comprised the ecoregions of Cerrado, Beni Savanna, Campos Rupestres Montane Savanna, Chiquitano Dry Forest, Dry Chao, Humid Chaco, Maranhão Babaçu Forests, and Pantanal.

## 2.2 SPECIES SELECTION

We modeled plant species that inhabit the Cerrado vegetation domain of Bolivia, Brazil, and Paraguay, which comprised different life forms such as trees, shrubs, subshrubs, and herbs. Because the Cerrado is considered one of the most diverse ecoregions in the world with thousands of species of vascular plants (MENDONÇA et al., 2008), it is difficult to determine which taxa are well distributed within the Cerrado domain or occur in a marginal way, given that several species are predominant in other neighboring biomes. For this reason, we created a general plant species list of the Cerrado for the three countries and then selected the taxa based on two criteria (see below).

The species list was constructed by compiling the national plant's catalogs from Brazilian Flora 2020[23], Bolivia Catalogue[24] and Paraguay Checklist[25]. We only considered native, endemic and non-endemic species that are terrestrial or rupicolous. Strictly aquatic or epiphytic taxa were not included. Regarding Cerrado’s endemism, we used species from Bolivia and Brazil (Paraguay does not have this information). Infra-specific taxa were not used because ENMs were constructed at the species level (see below). Since the lists came from different sources, species with different names could be synonymous. Thus, we homogenized and updated the scientific names by the Taxonomic Name Resolution Service v4.0 -TNRS- (BOYLE et al., 2013) based on APG III. After deleting duplicated names, the species list counted with a total of 6097 valid scientific names.

For species to be modeled, they had to meet the following criteria: (i) have more than five occurrence records and (ii) more than 40% of their occurrences located in the Brazilian Cerrado domain. This last criterion was used because most of the species are distributed in this country and we wanted to focus on the most typical species of the Cerrado. This criterion was not applied to the endemic species from the Bolivian Cerrado. For the species selection procedure, we used species

---

[23] http://floradobrasil.jbrj.gov.br/

[24] http://www.tropicos.org/Project/BC

[25] http://www.tropicos.org/Project/Paraguay

occurrence data from the Global Biodiversity Information Facility[26] (GBIF) because it concentrates the largest amount of species records for the study region. We applied a preliminary cleaning to the records of each species, removing occurrences that were outside the continent and had repeated coordinates. To estimate the proportion of records inside the Brazilian Cerrado, we created a raster of the Cerrado domain with 10 km resolution overlapping the maps of the Cerrado boundaries from Olson et al. (2001) and IBGE (2012). After the selection of species from the general list, we added species suggested by other researchers or cited in the literature (FRANÇOSO; HAIDAR; MACHADO, 2016; MAMANI et al., 2011; RATTER; BRIDGEWATER; RIBEIRO, 2003).

## 2.3 SPECIES RECORDS AND DATA CLEANING

For modeling species distribution, we used as species records source the GBIF; speciesLink[27]; *Instituto Chico Mendes de Conservação da Biodiversidade*[28]; Plant of Bolivia[29], and Tropicos. We checked, corrected and updated the species names of every record using TNRS and the species without information in this webpage were checked using The Plant List Version 1.1[30] and Tropicos. We cleaned the occurrences by removing records (i) located outside the continent, (ii) with repeated geographic coordinates, (iii) with coordinates without decimals or with half degree decimal (i.e. 0.5), (iv) located in the countries centroid and (v) without specification of species or genus. Later, we removed occurrences from invaded or cultivated areas, hence only those records that pertain to the natural distribution of the species were considered. For this, we used information about species distributions available in Brazilian Flora 2020, Flora del Conosur[31], and national species list from Tropicos. We also used the species voucher records to find information indicating that the collected specimen was cultivated (e.g. cultivated in

---

[26] www.gbif.org/
[27] http://splink.cria.org.br/
[28] https://biodiversidade.icmbio.gov.br/
[29] http://herbaria.plants.ox.ac.uk/bol/boliviajriwood
[30] http://www.theplantlist.org/
[31] www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm

botanical gardens, parks, universities or nurseries). We checked and corrected geographical location at provincial/state level for all the countries that comprised the records extent (i.e. Caribbean, North, Central and South America), by comparing the location of the records and the location specified in the specimen vouchers. We used geopolitical data sourced by Global Administrative Areas database v. 2.8[32] and a geographic information system (QGIS v. 2.18.7).

We added occurrences georeferenced at the municipal level provided by the speciesLink for those species with fewer than 20 cleaned occurrences. We only considered those records located in municipalities with a variation coefficient of environmental condition ≤ 15% for any of the 11 variables used for constructing the models (for further information about the selection of municipal georeferenced records see APPENDIX 1). As the species occurrences are commonly geographically biased we used a systematic sampling (FOURCADE et al., 2014) with a grain twice the resolution of environmental variables (10 arc-min). Only species with more than five cleaned records were modeled. Thus, the final database comprised 132,450 records for 1,555 species.

## 2.4 CLIMATIC AND LAND-USE DATA

The environmental variables used to construct the niche models considered both edaphic and climatic information. Usually, climate-only data are used to model plant species distributions (THUILLER, 2013). However, we decided to also use edaphic data because these variables determine an important part of the niche of terrestrial plant species and can improve the performance of ENMs (VELAZCO et al., 2017). Moreover, edaphic data affect the projection of plant species ENMs into future conditions (AUSTIN; VAN NIEL, 2011; BERTRAND; PEREZ; GÉGOUT, 2012). We used six soil variables related to physical properties assuming they may be less susceptible to climate change than those related to chemical variables (e.g. pH, phosphorus, nitrogen, or carbon). Edaphic data were provided by the SoilGrids (HENGL et al., 2017) with 0.75 arc-seconds resolution (c. 250 m) and then upscaled

---

[32] http://gadm.org/

to 5 arc-min (c. 10 km) by taking the average value of lower resolution cells into higher resolution cells. We used the 19 bioclimatic variables as climate data from current and future conditions. Current climatic conditions were obtained from WorldClim v2.0 (FICK; HIJMANS, 2017) with 5 arc-min resolution. Both databases summed up to 49 climatic variables (see the complete list of variables in APPENDIX 2). We performed a principal component analysis (PCA) on the original climatic variables based on a correlation matrix, to overcome multicollinearity problems and reduce the number of predictors variables. The scores of each derived principal component (PC) were used as new predictors. We selected nine PCs which explained up to 95.20% of the total variance from the original climatic variables (see APPENDIX 3 and 4). Once the PCA was performed for the current conditions, the resulting PCs were projected to future climate conditions based on the eigenvectors for current conditions (see below).

We used the climate projection from the 5th assessment report of the Intergovernmental Panel on Climate Change (IPCC) as the source of future climate conditions. We evaluated the effect of climate change using two Representative Concentration Pathways (RCP). They were the medium stabilizing 4.5 W/m$^2$ and very heavy 8.5 W/m$^2$ radiative forcing levels (VAN VUUREN et al., 2011), hereafter RCP4.5 and RCP8.5, respectively. These RCPs were assumed as optimistic and pessimistic scenarios, respectively. We used projections for 2050 (mean for the period from 2041 to 2060) and 2080 (mean for the period from 2071 to 2090).

Uncertainty on the estimation of future species ranges may be due to the use of different ENM algorithms and General Circulations Models -GCMs- (DINIZ-FILHO et al., 2009). As many GCMs are available for the region and in order to avoid their subjective selection, we use an adaptation of the Casajus et al. (2016) approach, which has the advantage of reducing the redundancy between climate scenarios, retain an adequate number of GCMs, and covering the greatest uncertainty between them (CASAJUS et al., 2016). This procedure was performed for both RCPs (4.5 and 8.5) by the year 2050. First, we obtained 28 GCMs from the Global Climate Model database[33] that were downscaled to the 5 arc-min resolution. Then, a PCA was carried out based on the 19 bioclimatic variables for each GCM, using cell values that comprised the study area. The first PC of all PCAs explained more than 50% of the

---

[33] http://ccafs-climate.org/

variation (mean 51.83 ± 0.85%; see APPENDIX 5 the complete list of GCMs and the corresponding explained variance of the first PC). We constructed a matrix using the scores of the first PC for each GCM. In this matrix, each row represented a GCM and each column a cell of the study area. The next step followed Casajus' approach, which consisted in performing a hierarchical cluster based on Euclidean distance matrix and Ward's linkage methods. From this hierarchical cluster, we extracted from 2 to 28 groups and calculated their centroids, which were used as the initial seed of each k-mean cluster analysis. The optimum number of groups was defined as the one that showed the maximum variability captured by the cluster (further information about GCMs selection see CASAJUS et al., 2016). Once the optimal number of cluster for each RCP was determined, we selected a GCM from shared groups formed by both RCPs cluster dendrograms (see APPENDIX 6 for more information about the selection of the Global Circulation Models). Thus, we used seven GCMs: CESM1-BGC, CSIRO-ACCESS-1.3, FIO-ESM, GFDL-ESM2G, GISS-E2-R, IPSL-CM5A-LR, and MOHC-HADGEM2-ES.

In order to evaluate the effect of land-use trends, we used current and future land cover provided by the Land-Use Harmonization[34] estimated by 2015, 2050 and 2080 (HURTT et al., 2011). The models MESSAGE-GLOBIOM and EMIND-MAGPIE were selected because they are consistent with the GHG emissions scenarios RCP4.5 and RCP8.5, respectively. As these data are available with a resolution of 0.25 degrees, we downscaled them by the bilinear method to 5 arc-minutes (c. 10 km) to have the same resolution of the other environmental variables. We quantified habitat loss by using the land use classes: C3 and C4 annual crops, C3 and C4 perennial crops, C3 nitrogen-fixing crops, urban, managed pastures, and rangelands. All crop categories were grouped under the category of croplands. Managed pastures and rangelands do not imply a total loss of natural cover but, given that they are anthropized, can harbor conditions that promote or demise the presence of a species, or even some species can adapt to the absence of natural habitat (KARP et al., 2012; MENDES; DE MARCO, 2017). Here, we considered that these land uses exert a negative impact on species in general.

---

[34] http://luh.umd.edu/code.shtml

## 2.5 MODELING PROCEDURES

Several correlative methods have been proposed for constructing ENMs, which may show variable performance depending on the condition of the modeling and its objective (ZHU; PETERSON, 2017). For this reason, ensemble models based on several algorithms are advisable (ARAÚJO; NEW, 2007). We used six ENM algorithms: Generalized Linear Models (GLM), Generalized Additive Models (GAM), Maximum Entropy (ME), Random Forest (RF), Support Vector Machine (SVM), and Gaussian Processes (GP). We fitted GLMs using the logistic link function and single predictor variables without interaction between them, no backward or forward covariates selection was used. For the GAMs, we used a binomial distribution and the same parametrization covariate as GLM, the Newton method was used to optimize the estimation of the smoothing parameter. MEs were fitted with the new approach based on inhomogeneous Poisson process procedure (PHILLIPS et al., 2017). We used linear, quadratic, product, and hinge features, default regularization values, 10,000 maximum background points and clog-log output format. The SVMs were performed with a radial basis kernel, with a constant cost value equal to one, and based on probabilities classes. The RFs were tuned automatically using the ‘tuneRF’ function of the *randomForest* R package, performed with 500 trees, default values of the step factor, and the improvement in out-of-bag error parameter. We considered those models with the minimum out-of-bag error as our final RF models. The GSs were constructed using Laplace approximation.

As the regional biological data lack true absences, we used pseudo-absences to fit the GLM, GAM, SVM, RF and GP models. The prevalence, call here as presence/pseudo-absence ratio, and the method of allocation of pseudo-absence can distinctively affect the algorithm performance (BARBET-MASSIN et al., 2012). Thus, we used presence/pseudo-absence equal to one for SVM, RF and GP approach, whereas for GLM and GAM we used 10,000 pseudo-absences. The pseudo-absences were allocated using environmental restriction based on a Bioclim model, allocating the pseudo-absences in the areas with the lowest suitability values. The area used to adjust ENMs must encompass the regions accessible to the dispersal capacities of species (i.e. M component of the BAM diagram; SOBERÓN; PETERSON, 2005). Such extent affects ENMs projections (ACEVEDO et al., 2017) and model accuracy (SAUPE et al., 2012). We define species-specific accessible areas bounded by the ecoregions where the occurrence records of each species are located. The ecoregion boundaries were sourced by WWF's Terrestrial Ecoregions of the World (OLSON et al., 2001).

## 2.6 MODEL EVALUATION, ENSEMBLE FORECAST, AND OVERPREDICTION CORRECTION

We used three approaches to evaluate ENM performance. For species with 5-15 occurrence records, we used the jackknife procedure (a.k.a. leave-one-out), where each partial model is constructed with $n$ – 1 records, hence the number of partitions were equal than the occurrences. For species with ≥ 20 occurrence records, we implemented a block fold-validation (Roberts et al., 2017) with two partitions (like a checkerboard) in order to control for spatial autocorrelation between training and testing data. To find a grid resolution for each species, we generated 20 grids with resolutions varying from 0.5 to 10 degrees, with a gradual increase of 0.5. The best grid was the one which simultaneously presented (i) the lower spatial autocorrelation, (ii) the maximum environmental similarity and (iii) the minimum difference of records between training and testing data. These three parameters were measured by Moran's I index, Multivariate environmental similarity surfaces, and the standard deviation, respectively. We used a k-fold cross-validation with five partitions for those species that did not find an efficient way to block fold-validation, i.e. the number of records of a single partition were ≤ 5. Finally, pseudo-absences were created within each partition group as explained above. We used the True Skill Statistic (TSS) as metric of model performance.

For the current environmental conditions and each GCM, we applied an ensemble forecast procedure to produce the final models for each species. Such procedure tends to reduce uncertainty presented by individual models (ARAÚJO et al., 2005). For this, we used the arithmetic average of the suitability predicted by the best algorithms of a species, i.e. those models with a performance greater than or equal to the algorithms’ average TSS. We built the final future projection by performing a new average of suitability values between the seven GCMs. We used the threshold that maximizes the sum of the sensitivity and specificity to transform continuous models (current or forecasted) to a binary one. This threshold was calculated based on the consensus model under current conditions. Thus, we constructed 9,336 models (Species x Algorithms), with 261,408 projection (GCMs x RCPs x Periods) that constituted the 6,224 final models (Species x RCPs X Periods).

Commonly, when ENMs are projected throughout the study region, they predict climatically suitable areas that can be far from the observed species distribution (PETERSON et al., 2011). To correct this overprediction, we only selected those suitable patches that met two criteria: (i) had at least one occurrence record or (ii) had no occurrences but were separated to those that did by less or equal than a certain distance *d*. The *d* was determined by the lower quartile of the pair-wise distance between patches with and without presences.

## 2.7 PROTECTED AREAS NETWORK

We collected different protected areas categories for the three countries to construct the protected area network of our study area. We used the indigenous lands as well as the municipal, departmental, and national protected areas from Bolivia sourced by the *Servicio Nacional de Areas Protegidas* updated to 2015[35], the public and private protected areas from Paraguay sourced by the *Secretaría del Ambiente* updated to 2007. The integral protection areas, sustainable use areas, and indigenous lands from Brazil, the first and second one sourced by the *Ministério do Meio Ambiente* updated to 2017, and the third one from *Fundação Nacional do Índio* updated to 2013, these data were sourced by *Laboratório de Processamento de Imagens e Geoprocessamento*[36]. We constructed a raster layer using this protected area dataset by rasterizing it to the resolution of the environmental data (i.e. 5 arc-min). We only considered as protected those grid cells that were overlapped by a protected area in a proportion ≥ 10%.

---

[35] http://geo.gob.bo
[36] http://maps.lapig.iesa.ufg.br/lapig.html

## 2.8 DATA ANALYSIS

We determined the effect of climate change on species distribution by considering a scenario of no-dispersion; i.e. future distribution ranges were delimited by the overlapping region between current and future geographical range. We considered this as a convenient procedure given the evidence that a no-dispersion situation can occur in plants (ZHU; WOODALL; CLARK, 2012). In addition, we did not have the large amount of data required to define the diversity of life cycles and forms of our modeled species, which are required to couple ENMs with dispersion models.

We assessed the relative species distributional loss for the whole study area and for each country caused by climate and climate plus land-use. These were calculated by the ratio between the range lost by different factors (i.e. climate and climate plus land-use) and the original distribution range (i.e. assuming a baseline landscape without anthropic land-use). We also calculated the contribution that each land-use category made to the distribution losses of the species.

The protection degree of species was calculated by the ratio between the species distribution area within the protected area network, either for current or future conditions, and the original distribution range for the current condition. The relative species distributional loss within the protected area network was based on the ratio between lost range of the species within protected areas for future condition and the current range within protected areas networks.

We assessed the species vulnerability combining two factors (i) the degree of conservation of a species, considered here as the proportion of a species' distribution within conservation units and (ii) the loss of species range outside protected areas. Under this scenario, a species can have a lower vulnerability because its range can be totally unprotected and be distributed in zones not affected by anthropic land-use. For future conditions, the vulnerability may increase as a species loses territory within protected areas caused by climate change or loses its range outside protected areas due to land-use expansion and/or contraction of species distribution by climate (FIGURE 1). The vulnerability of a species, for present and future condition, was calculated by the expression $V = 1 - (P + L)$, where $V$ is the vulnerability level, $P$ the protection degree of a species (see above), and $L$ the relative remnant distribution outside protected area network.

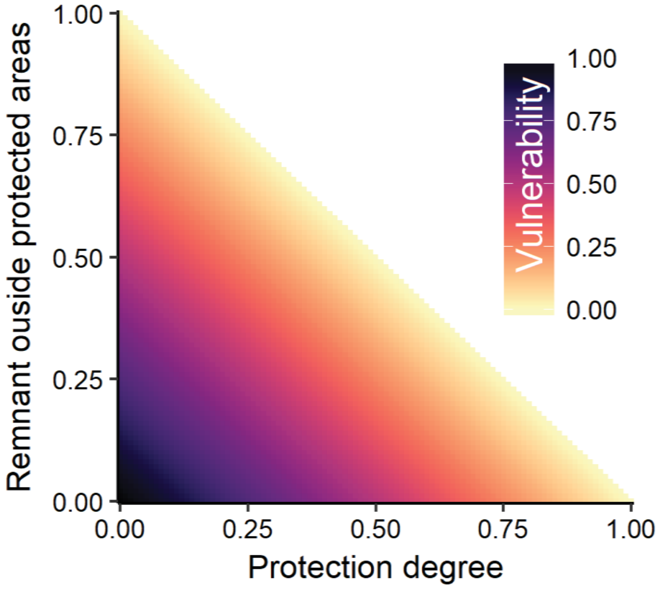


FIGURE 1 - Vulnerability degree based on the relationship between the species protection degree and the relative area outside protected areas. It is assumed a scenario in which protected areas are affected by the drift of species distribution because of climate change, while outside protected areas the range of species is affected by climate change and land-use.
SOURCE: the author (2018)

Data processing, construction of ENMs and analyses were conducted in the R environment version 3.4.1 (R CORE TEAM, 2017). We used the packages *maptools*, *raster*, *rgdal*, and *sp*, to handle spatial data; *rgbif* to download records from GBIF; *flora* for getting data from Brazilian Flora 2020. The *dismo* package was used to create pseudo-absences, model's prediction, and validation. The different ENMs approaches were fitted using *GRaF*, *kernlab*, *maxnet, mgcv, randomForest,* and *stats* packages.

## 3. RESULTS

We modeled 1,555 plant species comprising trees (151), shrubs (455), subshrubs (275), vines (88) and herbs (590). Models showed a satisfactory performance with mean TSS for all species of 0.76±0.16 (see APPENDIX 7). Under current conditions of land-use, the potential distribution of species was less than half of their complete modeled distribution (i.e. assuming a scenario without current land-use changes). These losses are different among nations. Bolivia is the country where species were least impacted by current land-use, while Paraguay suffered the greatest losses of its national flora (FIGURE 2a).

Our results showed that all species will have their distributions affected at some level for future climate conditions, even under the most optimistic scenarios. The RCP4.5 scenario predicts that species will lose an average of 34-40% of their distribution between 2050 and 2080, respectively; and 15-21 species are likely to become extinct (i.e. under future condition no suitable cell will be remain). Distribution loss tend to increase under RCP8.5 with average range loss of 43-60% for 2050 and 2080 respectively. Under this scenario, we estimated that 25 and 51 species may become extinct within each time period, respectively. These patterns of distributional loss showed an increase over time for all three countries. Paradoxically, species from Bolivia, which is the country with the lowest land-use effects, will have the greatest losses due to climate change in both scenarios. Despite an increase in distributional losses from 2050 to 2080 for species from Brazil and Paraguay, these species will experience lower distributional losses than those presented by species from Bolivia (FIGURE 2b).

We predicted a considerable increase in distributional losses when the effects of climate and land-use were combined for whole study region. Under the most optimistic scenario, distributional losses tend to stabilize over time. Species will lose on average 76 and 77% of their distribution by 2050 and 2080 respectively, with 15 and 21 species could become extinct for each year. The most pessimistic scenario estimated average distributional losses of 84 and 88%, with 26 and 55 species potentially going extinct for each time period (FIGURE 2c).

The combined effects of climate and land-use were different for the flora of each country. Although high, Bolivia will have the lowest effects of these changes, with average distributional losses by 2080 of 58 and 77% for RCP4.5 and RCP8.5, respectively. In Brazil, species distributional losses were estimated to be proportionally similar to those of the whole region, whereas Paraguay will have the highest distributional losses of its species, with average values for 2080 of 82 and 86% for each scenario, respectively (FIGURE 2c).

A total of 55 species can potentially go extinct due to climate change and land-use expansion under some of the scenarios and forecasting date for whole study area, with 13 of which are expected to be extinct for all projections. Of those 55 species, some are currently under some IUCN threat category, 11 are endangered, five vulnerable and two near threatened (see APPENDIX 8).

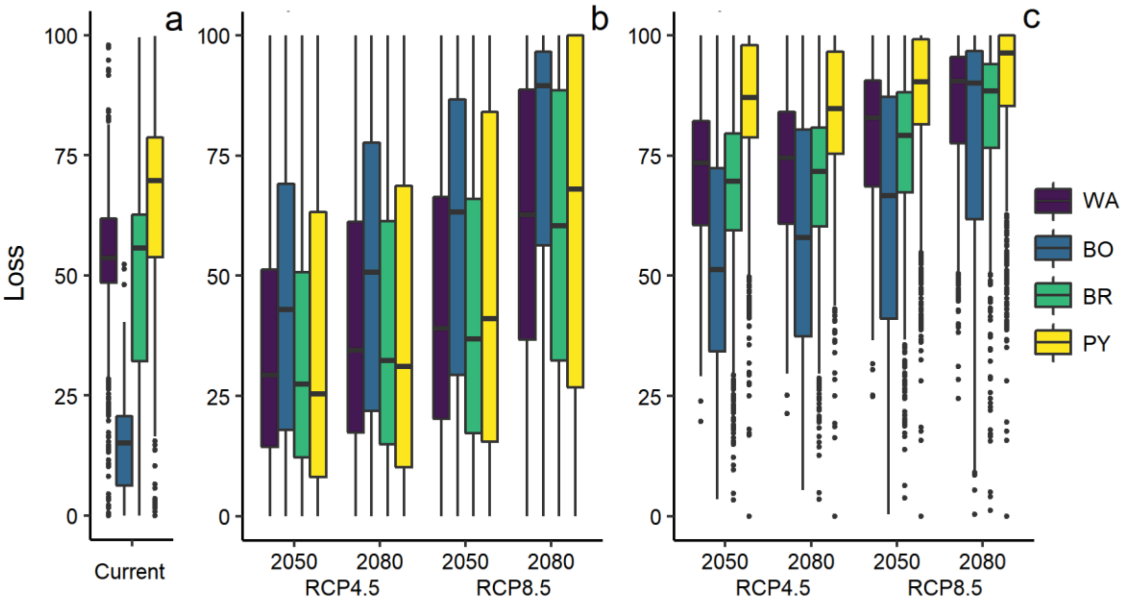


FIGURE 2 - Relative distributional loss of Cerrado’s plant species by land-use under current conditions (a), future climate (b) and future climate and land-use (c) under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080. Each color depicts different extents of assessment; for the whole study area (WA), and the nations of Bolivia (BO), Brazil (BR) and Paraguay (PY). Species losses assumed no dispersal to new suitable environmental conditions. The proportion of losses were calculated based on the potential distribution of species on the baseline landscape assuming unused primary vegetation. Distributional losses for each country were computed based on the original distribution of species within each nation.

SOURCE: the author (2018)

For the current land-use condition, the greatest species distributional losses for whole study area were caused by rangelands, followed by croplands. At national level, Bolivia and Brazil show that rangelands are the main causes; in Paraguay this land-use represented 50% of its losses since managed pastures are the second largest factor, and greater than previous countries. Under the optimistic scenario and considering the entire region, the effects of rangelands are reduced by the increase in croplands, such trend is similar for Bolivia and Brazil, while Paraguay shows a substantial increase of croplands by 2018. A considerable increase in croplands is estimated for the entire region for the pessimistic scenario. However, this will remain constant. At the national level, Bolivia and Paraguay are expected to present relative effects similar to the current state, while Brazil is predicted to significantly increase the effects of croplands (FIGURE 3).

The predicted richness pattern under current environmental conditions showed that the main concentration of species is located in the central and central-east area of the Cerrado ecoregion in Brazil, with a gradual decrease towards northern Brazil and to the west towards the Pantanal. In Bolivia, highest plant richness is concentrated in the eastern area in the Chiquitano Dry Forest and Dry Chaco ecoregions and eastern extreme of Humid Chaco ecoregions in Paraguay (FIGURE 4a). More importantly, regions with the highest predicted plant richness are concentrated over extensively disturbed areas in central Brazil. This situation is less pronounced in Bolivia, as the most disturbed regions are in places with lower richness, such as the Chiquitano Dry Forest and Dry Chaco (FIGURE 4a and b).

Most of the cells of our study area will reduce their species richness for both GHG emissions scenarios. Such species reduction will be higher under the pessimistic scenario (i.e. RCP8.5) than those predicted for optimistic scenario (FIGURE 4c). For 2080 and both GHG emissions scenarios, the region with the highest species richness will be in central-eastern and southern-eastern region of the Cerrado ecoregion in Brazil. In Bolivia, it is remarkable that the richest area will be Cerrado ecoregion near the Brazilian border, with additional important areas located in the Chiquitano Dry Forest, Dry Chaco, Beni Savanna. In Paraguay, plant richness will concentrate in the eastern of Humid Chaco ecoregion (FIGURE 4c). The areas with the most intense land-use will be coincident with that area highlighted as the richest for both forecasted year and GHG emissions scenario, with land-use expansion, around central Cerrado ecoregion in Brazil, the central region of Bolivia, and eastern and western Paraguay (FIGURE 4).

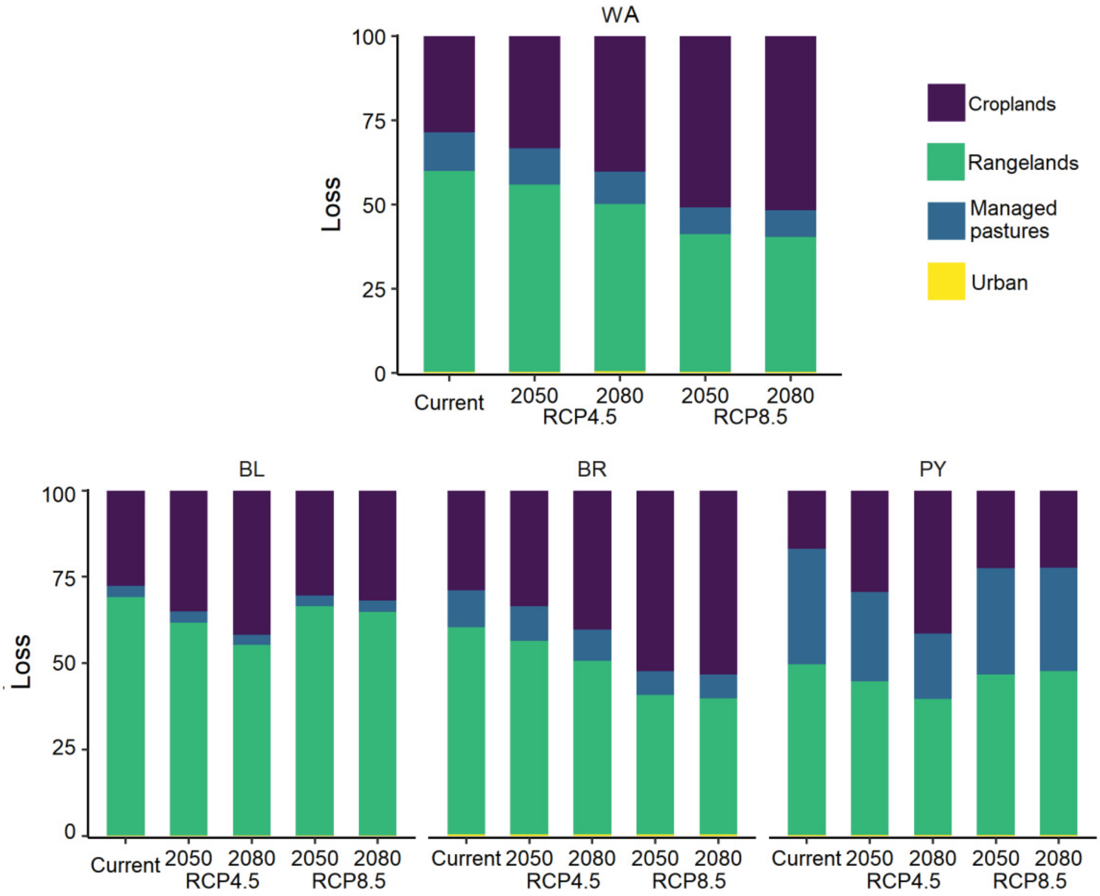


FIGURE 3 - Proportional distributional loss of Cerrado's plant species by different types of anthropogenic land-use for the whole study area (WA), and the nation of Bolivia (BL), Brazil (BR) and Paraguay (PY), under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080. These proportions were calculated based on distribution loss caused by land-use changes, losses by climate change were omitted.

SOURCE: the author (2018)

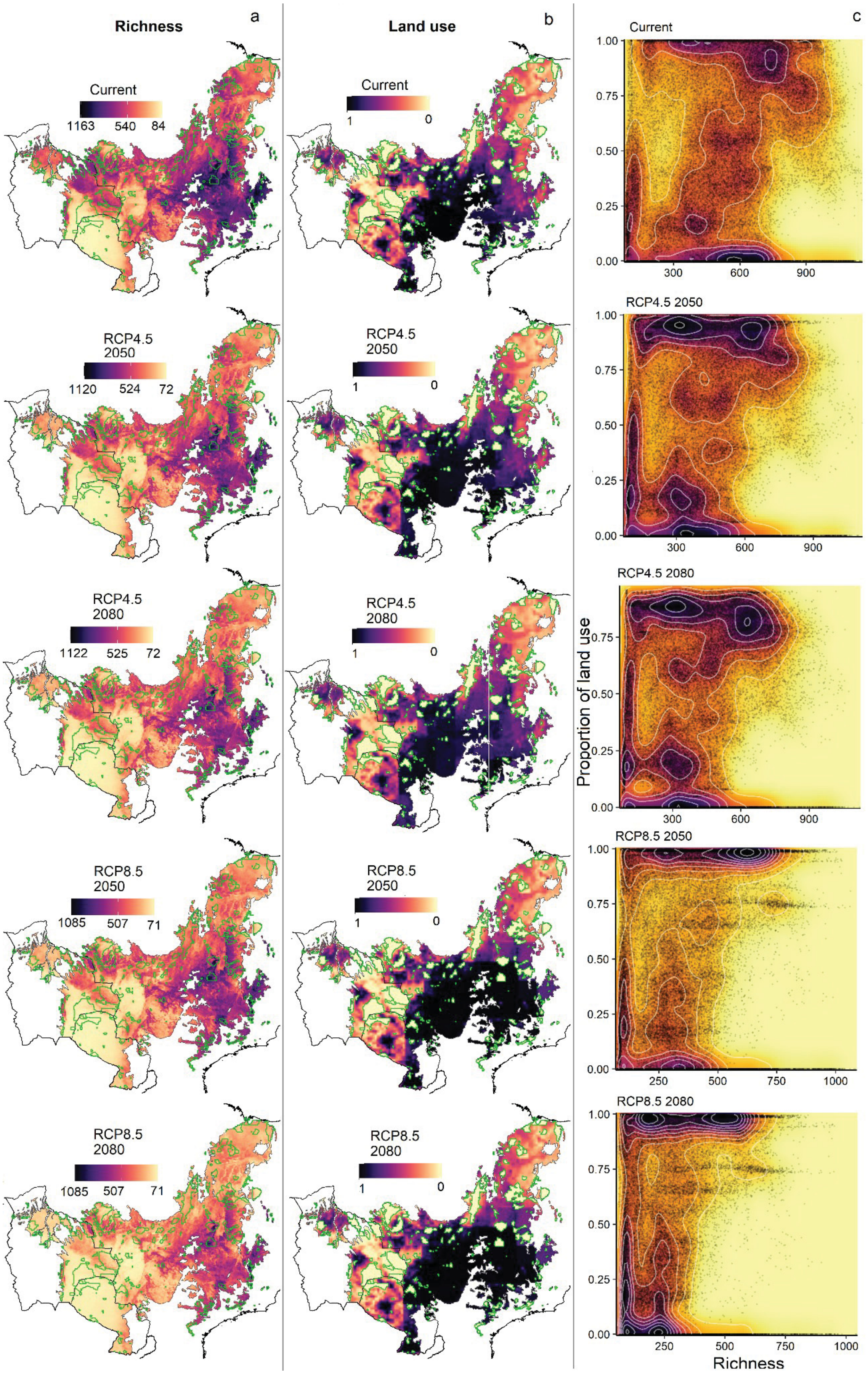

Richness
a
Land use
b
c
Current
1163
540
84
RCP4.5
2050
1120
524
72
RCP4.5
2080
1122
525
72
RCP8.5
2050
1085
507
71
RCP8.5
2080
1085
507
71
Current
1
0
RCP4.5
2050
1
0
RCP4.5
2080
1
0
RCP8.5
2050
1
0
RCP8.5
2080
1
0
Current
RCP4.5 2050
RCP4.5 2080
RCP8.5 2050
RCP8.5 2080
1.00
0.75
0.50
0.25
0.00
300
600
900
250
500
750
1000
Proportion of land use
Richness

FIGURE 4 - Current and future richness distribution of Cerrado plant species (a) intensity of land-use (b) and the relationship between richness and land-use (c) under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080. Future richness projections are based solely on species stable areas assuming a scenario without dispersion. Green polygons represent the current protected area network. Each point depicted in the panel (c) represents a grid cell of the study area.
SOURCE: the author (2018)

Regarding the patterns of species loss due to climate change, the greatest net losses for the optimistic scenario will occur mainly in the northern and northwestern regions of Bolivia and in the central region of the Cerrado in Brazil. In the pessimistic scenario, it is estimated that the highest values will be in the central and central-western region of Bolivia. In the ecoregion of Beni by 2050, these losses will increase in these regions by 2080. However, high losses will also occur in extensive areas of the central and central-western region of Brazil (see APPENDIX 9). In the case of relative loss values (i.e. the ratio between the richness that will be lost in the future and the current richness) the highest values for the optimistic scenario are concentrated in the ecoregion of Beni and southeast Bolivia, as well as in the Pantanal ecoregion. For the pessimistic scenario in addition to these areas, the southwest and north region of the Cerrado in Brazil stands out (APPENDIX 10).

Current conditions of the protection degree show that most species are poorly represented within the protected area network with an average value of degree of 16%. In fact, c. 1400 species have less than 25% of their distribution under protection. With respect to protection within each nation (that is, the relationship between the range of species within national protected areas and the area occupied by the species within that nation), Bolivia has the highest protection degree, followed by Brazil and Paraguay (FIGURE 5a).

For the optimistic scenario, species are expected to loss on average 30 and 36% of their current protected distribution by 2050 and 2080, respectively. Such losses will be higher for the pessimistic scenario, which will reach average losses of 39 and 57% for each time period, respectively (FIGURE 5b). At the national level, it is estimated that all countries will have an increase in the distributional losses within conservation units, with a maximum for the year 2080 and RCP8.5. In this way, Brazil will show the lowest values, but still high, as they increase to 50% by 2080. The situation in Bolivia and Paraguay will be more critical as these average values by 2080 will be greater than 75%. (FIGURE 5b).

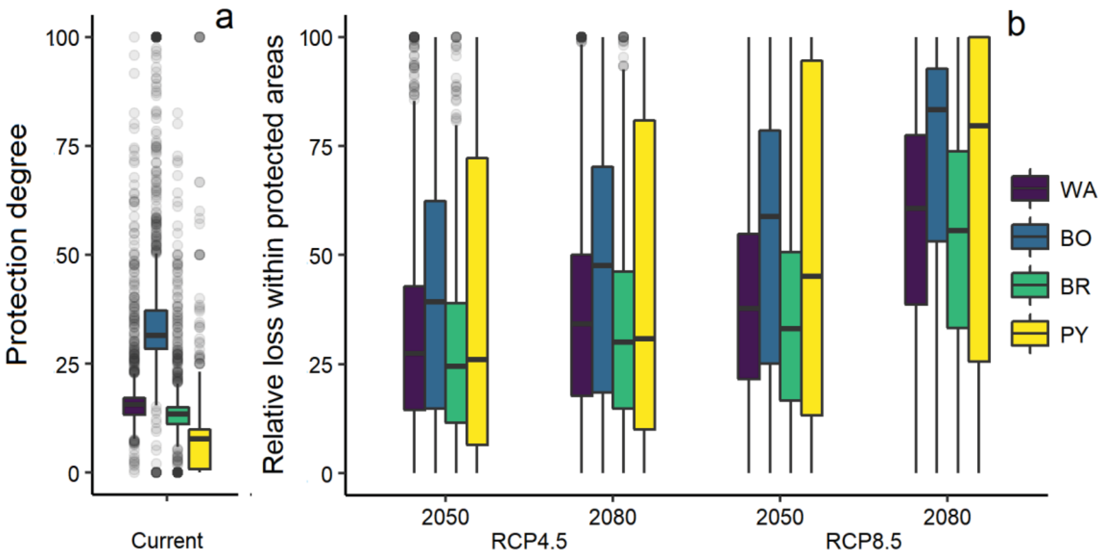


FIGURE 5 - Current relative protection (a) and future relative loss within protected areas (b) of Cerrado's plant species for whole study region (WR), Bolivia (BO), Brazil (BR) and Paraguay (PY). Future losses were forecasted for 2050 and 2080 under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios. The relative protected distribution was calculated based on the ratio between the protected area of a species and its complete range. The relative loss in protected area was based on the ratio between the stable climatic area into protected areas, i.e. assuming no dispersal to new suitable environment condition, and the current distribution under protection. Protection and losses for each country were based on the ratio of the stable and lost species range within national protected areas and the area occupied by the species within a nation.

SOURCE: the author (2018)

Regarding the vulnerability based on the relationship between the protection degree, and the remaining area outside the protected areas network; currently, the majority of species presents a vulnerability of 0.6 mainly due to the low species protection degree and secondly for the loss of their range outside protected area network (FIGURE 6a and b). The degree of vulnerability will be increased for all species but stable over time for the optimistic scenario, on average species will present a medium vulnerability (FIGURE 6a and c-d). Under the pessimistic scenario, an increase of loss area outside protected area network by climate and land-use will lead to increase significantly the vulnerability of species by 2050, however, the species-area loss within the protected area network will substantially contribute to the vulnerability of species by 2080. Thus, for this scenario, the average value of vulnerability will increase to 0.79 and 0.85 for 2050 and 2080 respectively (FIGURE 6a and e-f).

The relationship between species vulnerability and range size showed that, for the current environmental conditions, species with smaller range sizes show high variability of vulnerability. As the range increases, vulnerability tend to concentrate on the average values. However, this relationship tends to be negative regardless of the different scenarios or projected year, i.e. the vulnerability decreases as the range of species increases (see APPENDIX 11). Of the total modeled species, 93 had a vulnerability value equal or upper to the upper quartile for current condition and both RCP scenario and forecasted year. Of these species, 19 are currently under threat status, one as critically endangered, four vulnerable, four near threatened, and ten as least concerned (see APPENDIX 12).

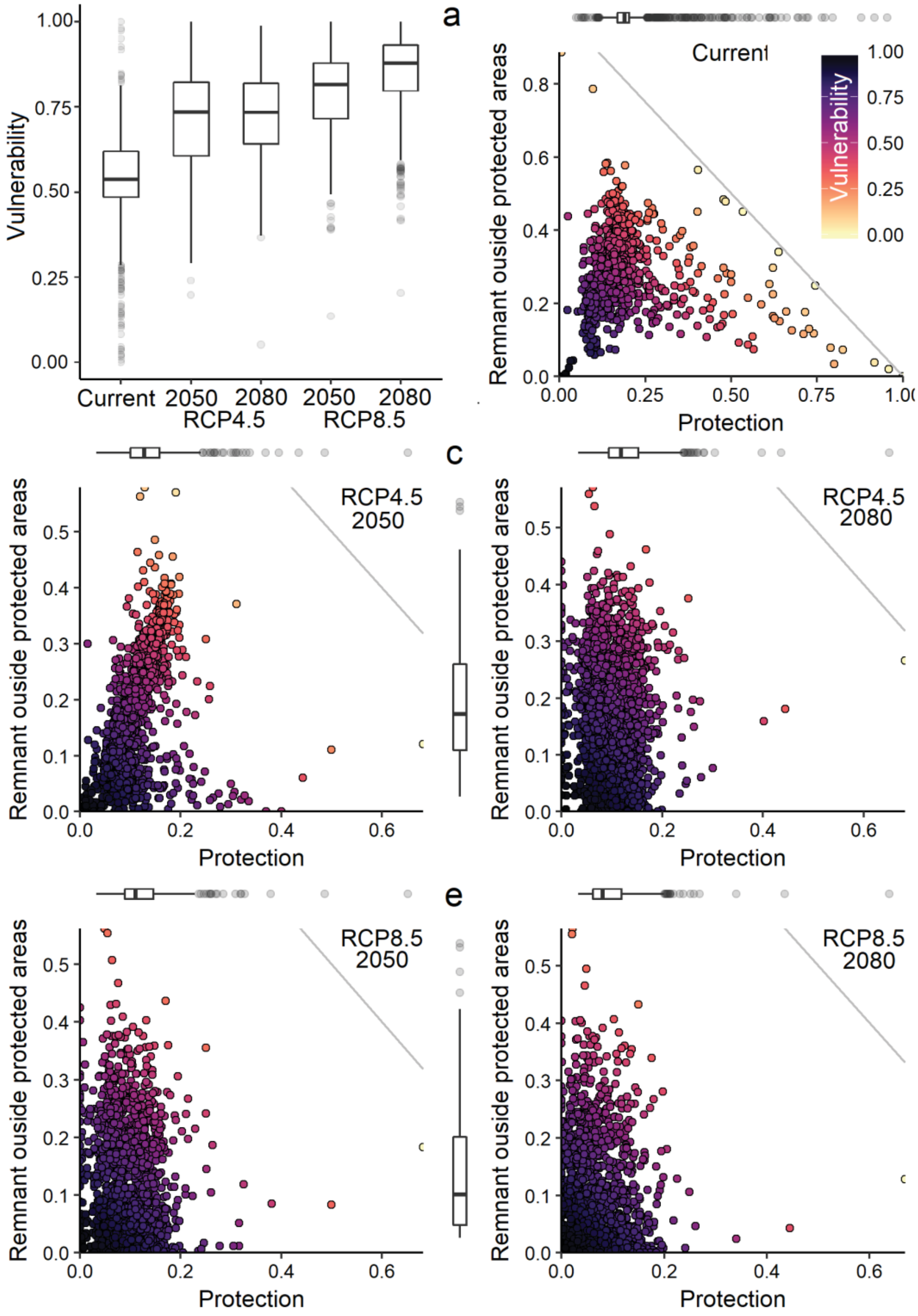


FIGURE 6 - Mean vulnerability level of 1555 Cerrado's plant species for current condition and future land-use and climate change under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080 (a). The relationship between vulnerability values and the relative protected distribution, and the relative distribution remnant outside protected areas for each greenhouse-gas emissions scenarios and period (b-f). There were used was used only climatic stable areas assuming a no dispersal to new suitable environment condition. Each depicted point represents a

modeled species. Gray diagonal line represents the situation of no vulnerability because of distribution range.
SOURCE: the author (2018)

## 4. DISCUSSION

We evaluated the effects of climate and land-use changes on the distribution of Cerrado plant species for two GHG emission scenarios, assessed the effectiveness of the protected area network to maintain these taxa, and measure their vulnerability under current and future conditions. We found that climate change and land-use will greatly reduce the geographic distribution of species by 2050 and 2080. The regions where current and future species richness were predicted to be greatest overlap with the most disturbed areas. The interaction between these two factors (climate and land-use) could cause substantial species distributional losses in each country, seriously compromising onservation efforts. Currently, the protected area network is not efficient in safeguarding Cerrado plant species nor it will be for future conditions due to the loss of suitable areas for species within conservation units. At the national level, the three studied countries showed different efficiencies of their protected area network. However, the different projections showed that countries that are promising for the conservation of the Cerrado flora under current conditions could be seriously affected in the future. This low protection degree and susceptibility to climate change, along with a more intensive and extensive land-use outside protected areas would lead the species to be severely threatened, even under the most optimistic scenario.

Green house gas emissions have increased since the pre-industrial era driven largely by economic and population growth (IPCC, 2014). Generally, country governance fails to account redistribution of species and its effects on policy and international decision (PECL et al., 2017). The consequences that climate change could have on the distribution of Cerrado’s plants show an unfavorable scenario. In fact, the latest international report on GHG emission trends show that our more pessimistic predictions could be very likely (UNEP, 2017). According to the last Emission Gap Report, keeping the global temperature increase in this century below 2°C (the Paris Agreement) is far from being achieved, as the gap between the necessary reductions and the national commitments made is still large (UNEP, 2017). At the regional level, Brazil was the main emitter of $CO_2$ equivalent in South America by 2015[37], nevertheless, it is one of the G20 members on track to meet the 2030 emission reduction targets (UNEP, 2017).

The mechanisms by which climate change affects a species are complex and, in addition to affecting its geographic distribution, it can alter another species characteristics such as their interactions with other species (e.g. pollinators, hosts, dispersers, competitors or pathogens), behavior, food abundance, among others (further examples see CAHILL et al., 2012 and references therein). Climate change effects on the phenology and interaction of some Cerrado's plants have already been suggested (VILELA et al., 2017). The negative effect of climate change, evaluated via ENMs, has been also reported for several species of the Brazilian Cerrado realm, such as trees (SIQUEIRA; PETERSON, 2003), edible and economical plants (DE OLIVEIRA et al., 2015; SIMON et al., 2013), and fauna (AGUIAR et al., 2016; OLIVEIRA et al., 2009). All these studies highlighted that the southern and southwestern regions of the Brazilian Cerrado will be the areas where species will tend to move or will be the most climatically stable. Our results showed similar patterns, however, the central region of the Cerrado is also predicted as an area of species richness concentration. Probably the differences between our results and the other studies can be caused by several factors such as the inclusion of soil data in our ENMs, our use of a greater species number than that of previous studies, and the no dispersion scenario. Further methodological differences between our study and others, we emphasize that a large part of the regions that would potentially concentrate the greatest remnant richness also are those that will suffer an increase in the expansion and intensity of land-use change. The loss of species distributional range both inside and outside of the protected areas would lead to most of the species be vulnerable to extinction.

At worldwide scale, land-use impact assessments predicted a species richness reduction of 3.4% by 2100 if assuming the trajectory of the pessimist land-use scenario (NEWBOLD et al., 2015). Also, anthropogenically land-uses such as plantation forest, cropland, pasture and urban areas significantly reduced species abundances and local species richness in comparison to primary vegetation

---

[37] http://www.fao.org/faostat/en/#data/GL

(NEWBOLD et al., 2015). We identified that the effects caused by land-use change to the diversity of Cerrado plants will tend to increase, mainly caused by rangelands and croplands. South American rangelands are a key factor in the economy of many countries (such as Brazil) as they support grazing and livestock, hence it is expected that anthropic activities in these areas will intensify in the future (YAHDJIAN; SALA, 2008). The increase in the effects of the agriculture expansion is in line with projections of future food demand, as this activity will need to produce almost 50 % more food, feed, and biofuel to meet the demand in 2050 (FAO, 2017). In addition, the population growth rate and its tendency to concentrate in urban areas (more than half of population is currently urban), has led to a rapid change of food consumption pattern (FAO, 2017), followed by an increase of livestock products (ALEXANDRATOS; BRUINSMA, 2012). In this sense, Brazil is considered one of the world's leading producers of agricultural commodities and, according to our forecasts, it is the country that will suffer the greatest impact of crop expansion on its plant diversity. Despite policies to halt deforestation in the Brazilian Cerrado, land-use data show that Cerrado continues to lose its natural cover. Forest areas and savannah lost 0.67 and 2.11 Mha respectively from 2010 to 2016, while farming expanded by 2.69 Mha[38]. Bolivia and Paraguay are also among the countries that have suffered heavy losses of natural areas in South America, mainly as a result of farming and livestock activities, whose production is partially destined to the foreign market (BAUMANN et al., 2017; FEHLENBERG et al., 2017; REDO; AIDE; CLARK, 2012). Therefore, they are also among the countries that supply global demands for agricultural goods, so the slowdown of their production would hardly occur.

Our findings show the enormous potential distributional losses that could be caused by climate change alone within protected areas. However, climate change can also affect the efficiency of protected areas in other ways, such as increasing disease risk or promoting population declines (MONZÓN; MOYER-HORNER; PALAMAR, 2011). In addition, there are other threats related to the political, social and legislative spheres of a nation. For example, most of the conservation units in Paraguay do not have a management plan (SEAM, 2016), the construction of a road through an important national park and indigenous area in Bolivia (FERNÁNDEZ-LLAMAZARES et al., 2018), or the potential mining development within Brazilian

---

[38] http://mapbiomas.org/stats

protected areas (VILLÉN-PÉREZ et al., 2017), which would affect the conservation efficiency in these countries.

At the national level and under current conditions, Bolivia showed the greatest protection degree of its national Cerrado flora. However, several of the species that occur there will be seriously affected by future climatic conditions, such effects are also observed in Paraguay (see APPENDIX 9 and 10). This may be partly due to the fact that these countries are rich in species distribution edges, which can be susceptible to be lost due to climate change (THOMAS, 2010). Sometimes, these distributional edges can serve as a starting point towards new environmentally suitable regions (CHANNELL; LOMOLINO, 2000). This latter effect could not be assessed here because we used a scenario in which species would not be able to disperse.

The expansion of protected areas to reach Aichi Biodiversity Target 11 may be threatened by the expansion of land-use (POUZOLS et al., 2014). As stated above, the countries where Cerrado flora distributes, showed a strong conversion of their lands to agriculture, and will probably expand to supply the domestic and foreign demand for farm-livestock products. For this reason, the strategic management of territories should simultaneously aim to expand protected areas, since all three countries have ratified the Convention on Biological Diversity; and maintain or increase the production of agricultural commodities.

The relationship between land-use and the distribution of the protected area network highlight the necessity of proper management and monitoring of protected areas, the creation of new ones in existing remaining areas, and the recovery of disturbed lands. Actions recently proposed in the Brazilian Cerrado, such as expansion from croplands to pasturelands, productivity improvement, increase protections and land-use planning among others (STRASSBURG et al., 2017), should be implemented in the neighboring countries of Paraguay and Bolivia. It would be appropriate to face the loss of plant species in the Cerrado through global and regional actions. Actions covering the entire study area (Bolivia, Brazil and Paraguay) could improve conservation efficiency. For instance, the allocation of protected areas throughout the entire domain can be more effective than those implemented within each nation (see MOILANEN et al., 2013). Also, seeds collected from widely-spaced populations can capture more genetic variability (HOBAN; SCHLARBAUM, 2014).

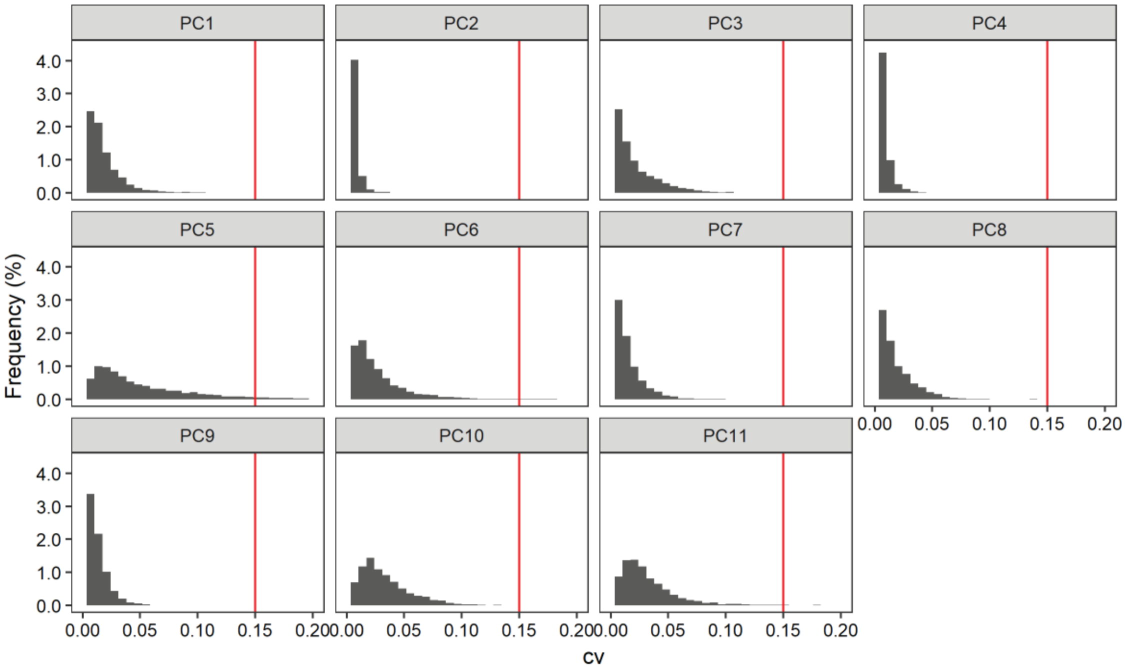
We showed that for most species with lower than 50% of their range within protected areas, there are remnants of their distributions without protection, probably in private areas. This result suggests that for a general conservation plan to be successful, it must also consider the protection of species in private areas. However, such actions may not be sufficient or reachable under current conservation plans. For instance, in Brazil, there is a Forest Code to regulate deforestation on private lands but it has been recently shown that the areas considered as legal for deforestation are much larger than those that would have to be restored to overcome such action (VIEIRA et al., 2017). The creation of private reserves may be considered an interesting way to maintain landscapes where species could persist. Nevertheless, actions within each nation would also be necessary in order to maintain its biological patrimony, this parochialism may also have positive points (see HUNTER; HUTCHINSON, 1994). New studies would be needed to assess what would be the priority areas needed for conserving biodiversity under the combined effect of climate and land-use changes projections thorough entire territory.

## 5. CONCLUSION

We demonstrated that climate change and land-use could cause great damage to the Cerrado flora by the years 2050 and 2080, even under optimistic conditions. Unfortunately, the greatest intensity and extent of land-use will be on the regions where the greatest species richness will be harbored. Conservation of Cerrado’s plant species will also be seriously affected. The current protected area network is not efficient (and will not be) in safeguarding these species under current and future conditions, owing to the considerable habitat loss predicted to be caused by climate change within the conservation units. The low protection degree coupled with the losses caused by climate change and the advance of anthropic land-use will lead to most species being highly vulnerable to extinction. Given that climate and land-use effects showed different interactions for each country, conservation strategies for protecting the Cerrado’s flora will have to be implemented at both transboundary/international and national levels.

# 6. SUPPORTING INFORMATION

APPENDIX 1 - Histogram depicting the frequency of Brazilian municipalities for classes of coefficient variation based on the environmental variability within municipalities' boundaries for each principal component. The variation coefficient of each municipality was calculated based on the cells of each principal component within each municipality. Most of the municipality from Brazil had variation coefficient lower the 15% allowing the use of most municipal georeferenced records.



SOURCE: the author (2018)

APPENDIX 2 - Climate and edaphic variables used to construct the ecological niche models.

| Source | Variables | Units |
|---|---|---|
| WorldClim | Annual mean temperature (BIO1) | °C |
| | Mean diurnal range (BIO2) | °C |
| | Isothermality (BIO3) | °C |
| | Temperature Seasonality (BIO4) | % |
| | Max temperature of warmest week (BIO5) | °C |
| | Min temperature of coldest week (BIO6) | °C |
| | Temperature annual range (BIO7) | °C |
| | Mean temperature of wettest quarter (BIO8) | °C |
| | Mean temperature of driest quarter (BIO9) | °C |
| | Mean temperature of warmest quarter (BIO10) | °C |
| | Mean temperature of coldest quarter (BIO11) | °C |
| | Annual precipitation (BIO12) | mm |
| | Precipitation of wettest week (BIO13) | mm |
| | Precipitation of driest week (BIO14) | mm |
| | Precipitation seasonality (BIO15) | - |
| | Precipitation of wettest quarter (BIO16) | mm |
| | Precipitation of driest quarter (BIO17) | mm |
| | Precipitation of warmest quarter (BIO18) | mm |
| | Precipitation of coldest quarter (BIO19) | mm |
| SoilGrid | Absolute depth to bedrock (in cm) | cm |
| | Clay content (0-2 micro meter) mass fraction * | % |
| | Coarse fragments volumetric * | % |
| | Predicted probability of occurrence of R horizon | % |
| | Sand content (50-2000 micrometer) mass fraction * | % |
| | Silt content (2-50 micrometer) mass fraction * | % |

*Data for seven depths

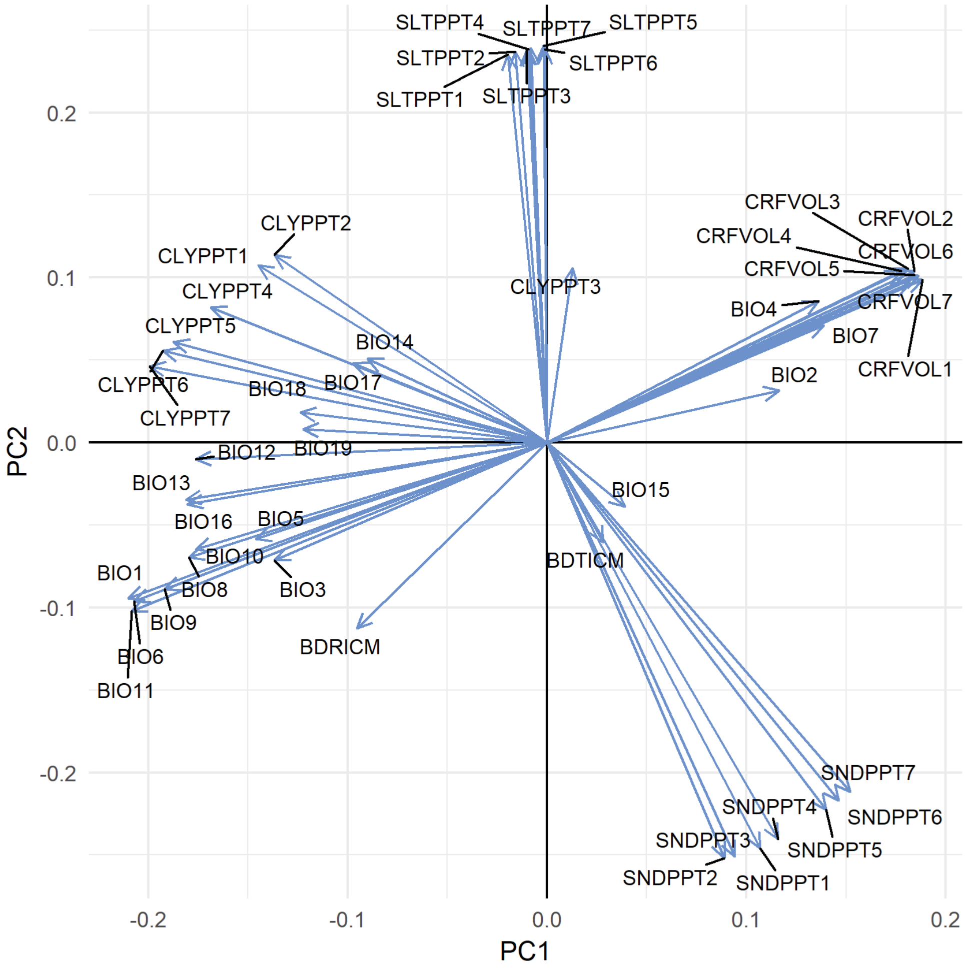
SOURCE: the author (2018)

APPENDIX 3 - Coefficients of the principal components selected from the principal component analysis performed with environmental data

| Principal Components | Eigenvalues | Variance explained for each PC | Cumulative variance explained |
|---|---|---|---|
| 1 | 4.0 | 32.9 | 32.9 |
| 2 | 3.5 | 25.0 | 57.9 |
| 3 | 2.6 | 13.3 | 71.2 |
| 4 | 2.3 | 10.4 | 81.7 |
| 5 | 1.6 | 5.1 | 86.7 |
| 6 | 1.3 | 3.2 | 90.0 |
| 7 | 1.0 | 2.2 | 92.2 |
| 8 | 1.0 | 1.9 | 94.0 |
| 9 | 0.8 | 1.3 | 95.3 |

SOURCE: the author (2018)

APPENDIX 4 - Ordination diagram for the first two axis of three PCAs conducted with environmental variables



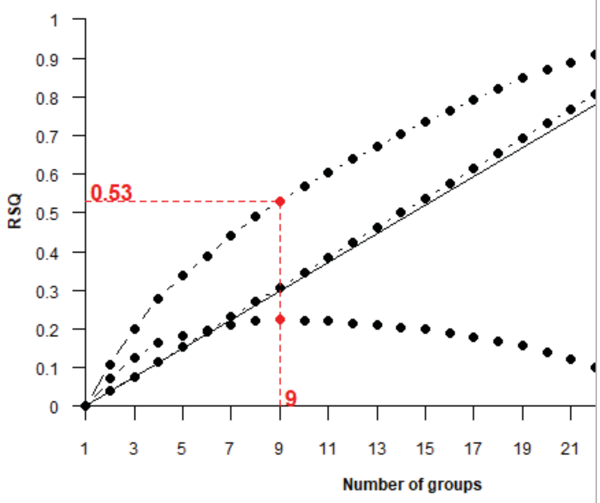
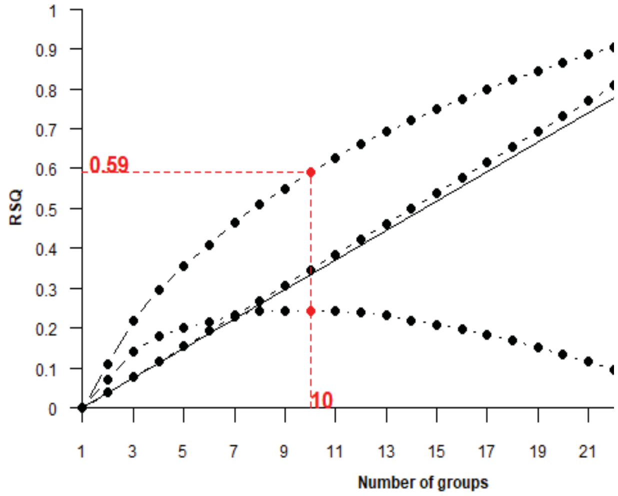
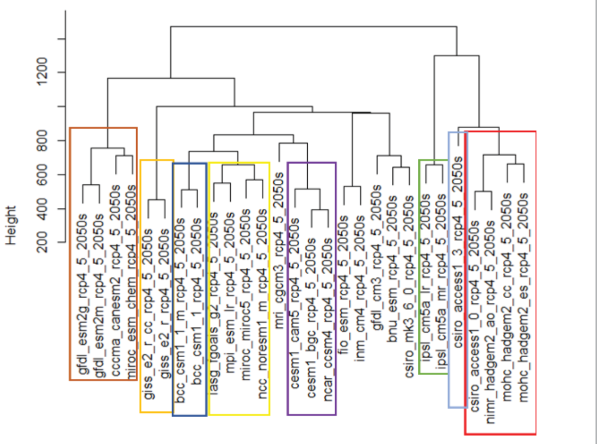
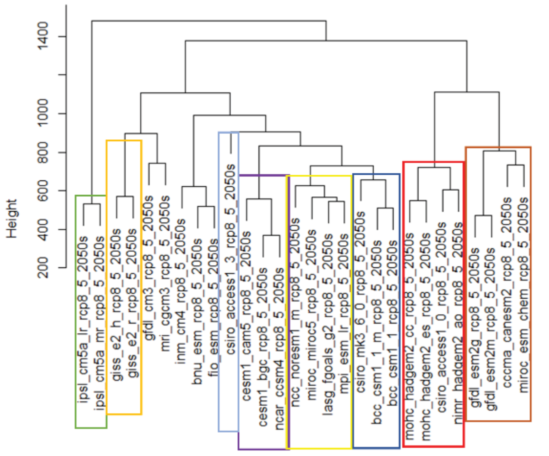
SOURCE: the author (2018)

APPENDIX 5 - List of Global Circulation Models and the explained variance for the first principal component (PCV) of Principal Component Analysis performed with the 19 bioclimatic variables for two emission scenarios greenhouse-gas emissions scenarios (RCP4.5 and RCP 8.5) forecasted for 2050

| GCMS | PCV | |
|---|---|---|
| | RCP4.5 | RCP8.5 |
| BCC-CSM1.1M | 0.518 | 0.520 |
| BCC-CSM1.1 | 0.520 | 0.518 |
| BNU-ESM | 0.513 | 0.515 |
| CCCMA-CANESM2 | 0.517 | 0.515 |
| CESM1-BGC | 0.516 | 0.510 |
| CESM1-CAM5 | 0.511 | 0.504 |
| CSIRO-ACCESS1.0 | 0.517 | 0.514 |
| CSIRO-ACCESS1.3 | 0.523 | 0.514 |
| CSIRO-MK3.6.0 | 0.522 | 0.529 |
| FIO-ESM | 0.528 | 0.524 |
| GFDL-CM3 | 0.508 | 0.503 |
| GFDL-ESM2G | 0.530 | 0.526 |
| GFDL-ESM2M | 0.530 | 0.527 |
| GISS-E2-R-CC | 0.515 | 0.517 |
| GISS-E2-R | 0.512 | 0.514 |
| INM-CM4 | 0.528 | 0.521 |
| IPSL-CM5A-LR | 0.529 | 0.525 |
| IPSL-CM5A-MR | 0.517 | 0.531 |
| LASG-FGOALS-G2 | 0.537 | 0.534 |
| MIROC-ESM-CHEM | 0.516 | 0.516 |
| MIROC-MIROC5 | 0.531 | 0.524 |
| MOHC-HADGEM2-CC | 0.519 | 0.507 |
| MOHC-HADGEM2-ES | 0.503 | 0.499 |
| MPI-ESM-LR | 0.524 | 0.525 |
| MRI-CGCM3 | 0.521 | 0.509 |
| NCAR-CCSM4 | 0.517 | 0.509 |
| NCC-NORESM1-M | 0.521 | 0.519 |
| NIMR-HADGEM2-AO | 0.511 | 0.504 |

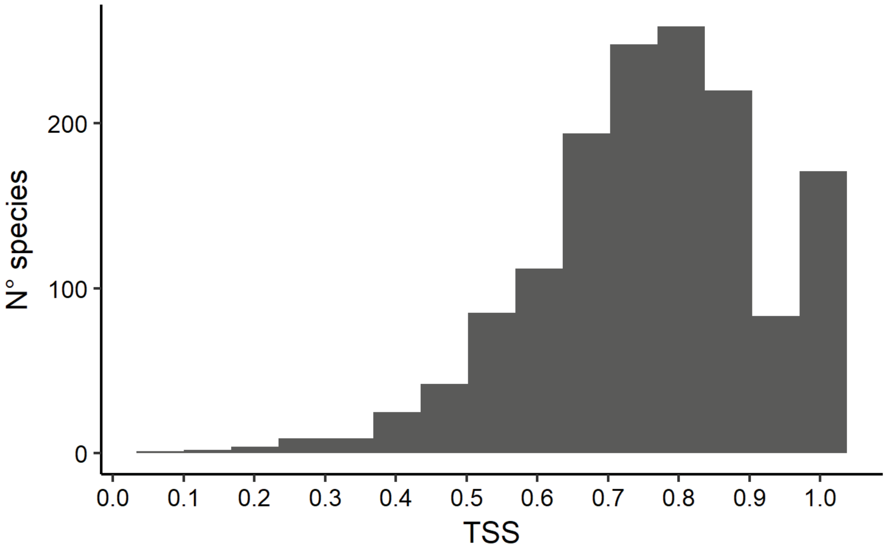
SOURCE: the author (2018)

APPENDIX 6 - Optimal number of cluster (upper panel) and cluster analysis (lower panel) based on the first principal component coming from Principal Components Analysis performed with the 19 bioclimatic variables of 28 Global Circulation Models (GCMs) for two greenhouse-gas emissions scenarios (RCP4.5 and RCP 8.5) forecasted for 2050. Similar GCM's group are depicted by equal square color in cluster dendrograms (lower panel).

| RCP4.5 | RCP8.5 |
|---|---|
|  |  |
|  |  |

SOURCE: the author (2018)

APPENDIX 7 - Model performance of 1555 plant species of Cerrado measured by TSS (True Skill Statistic). Each species value was based on the average of TSS of the best algorithm used to construct ensembled models



SOURCE: the author (2018)

APPENDIX 8 - List of species that may potentially be extinct due to climate change and land-use change for optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080.

| Species | RCP4.5 | | RCP8.5 | | Threat |
|---|---|---|---|---|---|
| | 2050 | 2080 | 2050 | 2080 | status |
| *Arachis guaranitica* | ● | ● | | ● | |
| *Arachis kempff-mercadoi* | | | | ● | |
| *Arachis magna* | ● | ● | ● | ● | |
| *Arachis matiensis* | | | | ● | |
| *Arrojadoa dinae* | | | | ● | NT |
| *Arthrocereus glaziovii* | ● | ● | | ● | EN |
| *Begonia lindmanii* | | ● | ● | ● | |
| *Brasilia sickii* | | | | ● | |
| *Chronopappus bifrons* | ● | ● | ● | ● | VU |
| *Cipocereus minensis* | | | | ● | VU |
| *Cuphea luteola* | ● | ● | ● | ● | |
| *Dalechampia herzogiana* | | | ● | ● | |
| *Dasyphyllum reticulatum* | | | | ● | NT |
| *Dimerostemma grazielae* | ● | ● | ● | ● | VU |
| *Dimerostemma humboldtianum* | ● | ● | ● | ● | |
| *Dyckia brachyphylla* | ● | ● | ● | ● | |
| *Dyckia ferruginea* | | | | ● | |
| *Echinocoryne echinocephala* | | | | ● | EN |
| *Eitenia polyseta* | | | | ● | |
| *Eremanthus argenteus* | | | | ● | EN |
| *Fosterella hatschbachii* | | | | ● | |
| *Galianthe kempffiana* | | | | ● | |

| Species | RCP4.5 | | RCP8.5 | | Threat |
|---|---|---|---|---|---|
| | 2050 | 2080 | 2050 | 2080 | status |
| *Hyptis meridionalis* | | | | ● | |
| *Hyptis tuberosa* | ● | ● | ● | ● | 🟩 EN |
| *Hyptis woodii* | ● | ● | ● | ● | |
| *Ichthyothere palustris* | | | | ● | |
| *Ipomoea psammophila* | | | ● | ● | |
| *Justicia sarothroides* | | | ● | ● | |
| *Lepidaploa tombadorensis* | | | | ● | |
| *Lessingianthus bishopii* | | ● | | ● | |
| *Lessingianthus eitenii* | | | ● | ● | 🟨 EN |
| *Lessingianthus erythrophilus* | | ● | ● | ● | |
| *Lessingianthus exiguus* | | | | ● | 🟨 VU |
| *Lessingianthus fonsecae* | | | | ● | |
| *Lessingianthus lanuginosus* | | | | ● | |
| *Lychnophora diamantinana* | | ● | ● | ● | 🟨 EN |
| *Lychnophora markgravii* | | | | ● | 🟨 EN |
| *Manihot arenaria* | | ● | ● | ● | |
| *Manihot fabianae* | | | ● | ● | |
| *Mimosa lupinoides* | | | ● | ● | |
| *Minaria grazielae* | | | | ● | 🟨 EN |
| *Moquiniastrum hatschbachii* | | ● | ● | ● | |
| *Paralychnophora glaziouana* | | | | ● | |
| *Paspalum crispatum* | | | | ● | |
| *Philodendron flumineum* | | | | ● | |
| *Pilosocereus fulvilanatus* | ● | ● | ● | ● | 🟨 EN |
| *Piptocarpha matogrossensis* | | | | ● | |
| *Piptolepis schultziana* | ● | ● | ● | ● | |
| *Pitcairnia chiquitana* | ● | | ● | ● | |
| *Pitcairnia platystemon* | ● | ● | ● | ● | |
| *Proteopsis argentea* | | | | ● | 🟨 VU |
| *Rauvolfia gracilis* | ● | ● | ● | ● | |
| *Richterago lanata* | ● | ● | ● | ● | 🟨 EN |
| *Schefflera glaziovii* | | | | ● | 🟨 EN |
| *Syagrus deflexa* | | ● | ● | ● | |

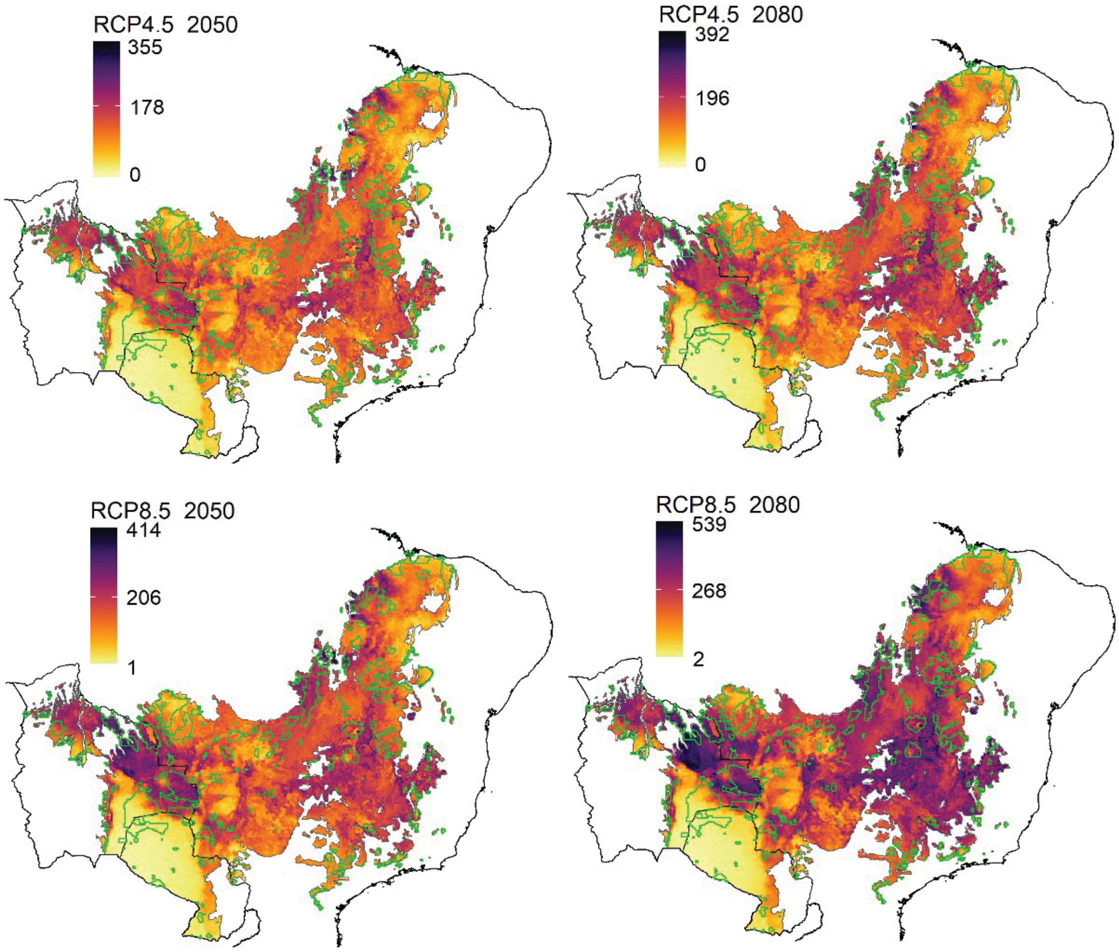
🟨: Brazilian Flora 2020 (http://floradobrasil.jbrj.gov.br/)

🟦: MARTINELLI, G.; MESSINA, T.; SANTOS FILHO, L. Livro vermelho da flora do Brasil – Plantas raras do Cerrado; 1. ed. - Rio de Janeiro: Andrea Jakobsson: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro: CNCFlora, 2014. 320p.

🟩 : MAMANI, F. et al. Libro Rojo de las Plantas de los Cerrados del Oriente Boliviano. Santa Cruz, Bolivia: John R I Wood, 2010. 153p.

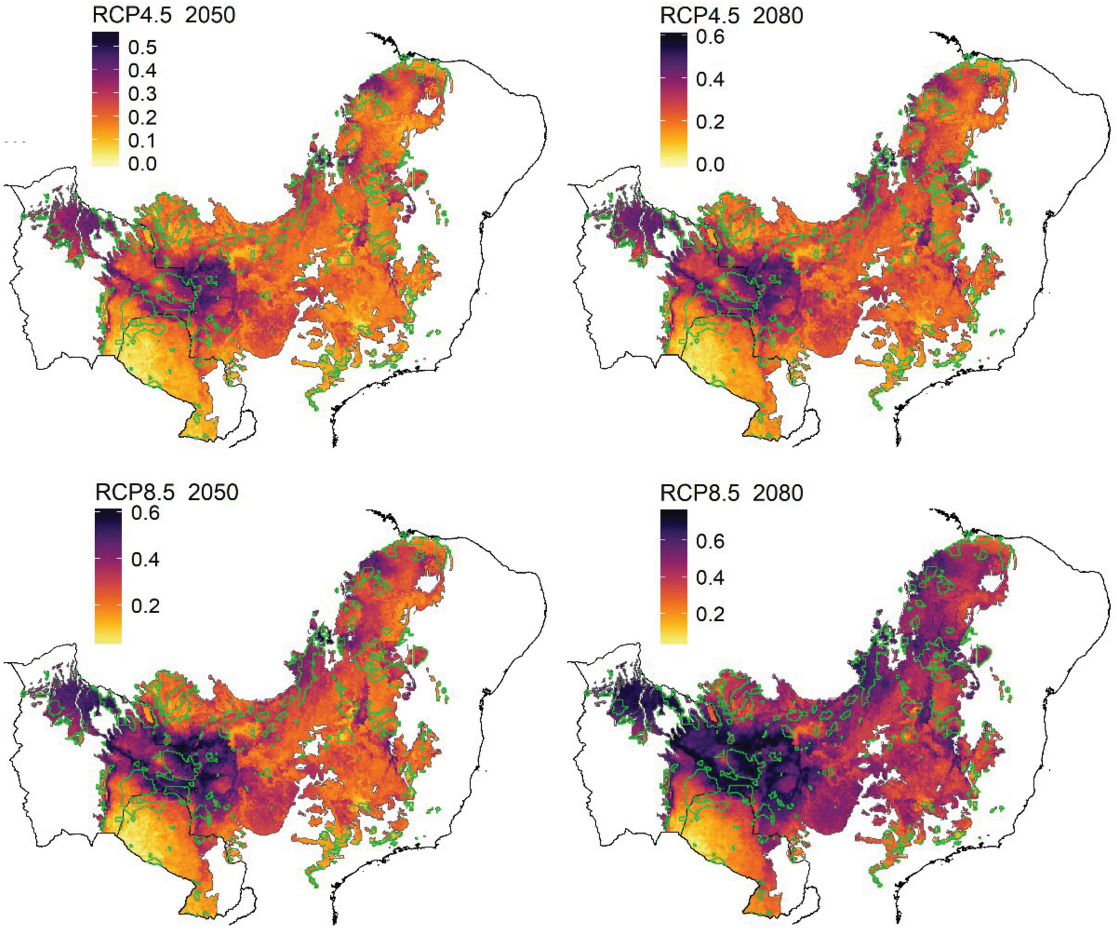
SOURCE: the author (2018)

APPENDIX 9 - Total species loss as consequence of climate change under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080. Green polygons depict the protected area network.



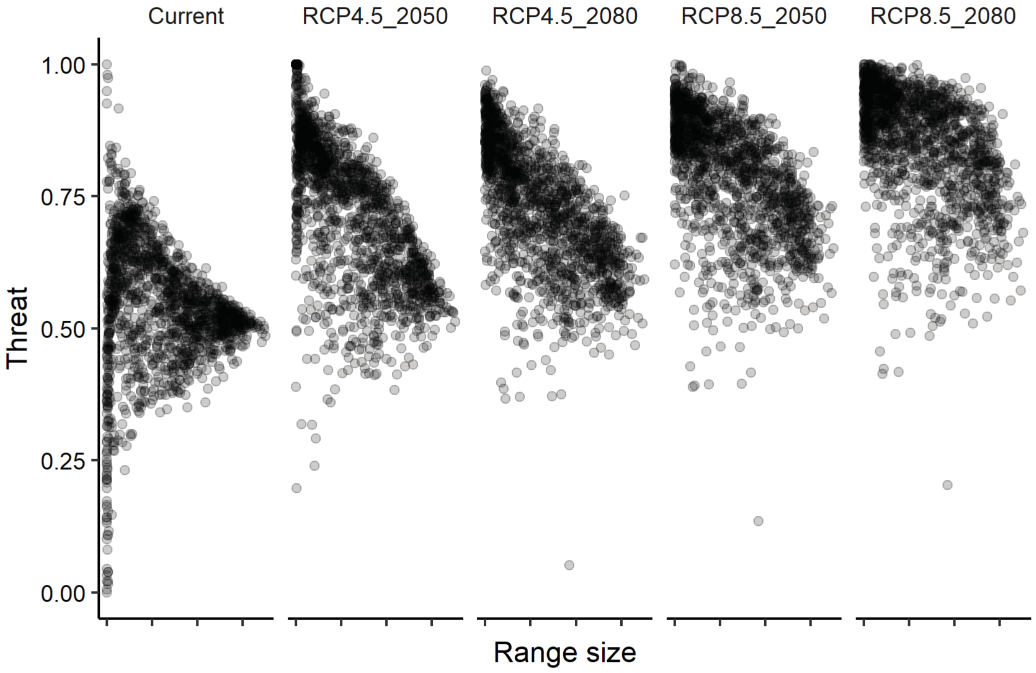
SOURCE: the author (2018)

APPENDIX 10 - Relative species loss as consequence of climate change under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080. Green polygons depict the protected area network. Relative values by each cell were based on the ration between the richness that will be loss in the future and the current richness



SOURCE: the author (2018)

APPENDIX 11 - Relationship species range size and vulnerability level for current and future condition under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080.



SOURCE: the author (2018)

APPENDIX 12 - Most vulnerable species of the 1555 species evaluated based on a procedure that considers the degree of protection and loss of habitat for current and future condition under optimistic (RCP4.5) and pessimistic (RCP8.5) greenhouse-gas emissions scenarios forecasted for 2050 and 2080. They are listed only those species with a vulnerability level greater or equal to the upper quartile for all period and scenarios

| Species | Current | RCP4.5_2050 | | RCP8.5_2050 | | Threat status |
|---|---|---|---|---|---|---|
| | | 2050 | 2080 | 2050 | 2080 | |
| *Acrocomia hassleri* | 0.691 | 0.922 | 0.912 | 0.946 | 0.980 | LC |
| *Adenocalymma peregrinum* | 0.641 | 0.875 | 0.872 | 0.905 | 0.971 | |
| *Adiantum ornithopodum* | 0.690 | 0.881 | 0.884 | 0.917 | 0.964 | VU |
| *Ageratum fastigiatum* | 0.700 | 0.832 | 0.831 | 0.906 | 0.960 | |
| *Aldama squalida* | 0.658 | 0.845 | 0.901 | 0.931 | 0.992 | |
| *Alstroemeria julieae* | 0.735 | 0.850 | 0.858 | 0.917 | 0.973 | |
| *Alstroemeria plantaginea* | 0.684 | 0.845 | 0.891 | 0.936 | 0.989 | |
| *Alstroemeria rupestris* | 0.710 | 0.864 | 0.880 | 0.921 | 0.951 | |
| *Anacardium corymbosum* | 0.682 | 0.903 | 0.867 | 0.930 | 0.959 | |
| *Aristolochia stomachoides* | 0.755 | 0.902 | 0.949 | 0.975 | 0.999 | |
| *Aspilia cylindrocephala* | 0.671 | 0.886 | 0.885 | 0.920 | 0.932 | VU |
| *Aspilia platyphylla* | 0.633 | 0.833 | 0.962 | 0.998 | 1.000 | |
| *Aspilia riedelii* | 0.651 | 0.878 | 0.935 | 0.951 | 0.995 | |
| *Astronium nelson-rosae* | 0.709 | 0.871 | 0.886 | 0.924 | 0.940 | |
| *Baccharis concinna* | 0.658 | 0.832 | 0.889 | 0.940 | 0.987 | VU |
| *Baccharis pseudoalpestris* | 0.621 | 0.889 | 0.923 | 0.938 | 0.966 | |

| Species | Current | RCP4.5_2050 | | RCP8.5_2050 | | Threat status |
|---|---|---|---|---|---|---|
| | | 2050 | 2080 | 2050 | 2080 | |
| *Bidens flagellaris* | 0.633 | 0.880 | 0.899 | 0.925 | 0.975 | |
| *Byttneria oblongata* | 0.635 | 0.829 | 0.839 | 0.911 | 0.957 | LC |
| *Calea verticillata* | 0.813 | 0.920 | 0.888 | 0.916 | 0.969 | |
| *Calliandra foliolosa* | 0.841 | 0.893 | 0.942 | 0.980 | 0.974 | |
| *Campuloclinium hirsutum* | 0.643 | 0.861 | 0.884 | 0.931 | 0.958 | |
| *Campuloclinium megacephalum* | 0.705 | 0.858 | 0.866 | 0.913 | 0.962 | |
| *Campuloclinium riedelii* | 0.751 | 0.888 | 0.835 | 0.900 | 0.950 | LC |
| *Chionanthus trichotomus* | 0.725 | 0.834 | 0.849 | 0.914 | 0.969 | |
| *Chresta scapigera* | 0.650 | 0.851 | 0.873 | 0.916 | 0.977 | |
| *Chresta sphaerocephala* | 0.710 | 0.841 | 0.853 | 0.923 | 0.956 | LC |
| *Chromolaena caaguazuensis* | 1.000 | 1.000 | 0.949 | 0.955 | 0.985 | |
| *Chromolaena cylindrocephala* | 0.703 | 0.857 | 0.866 | 0.933 | 0.971 | |
| *Croton cinerellus* | 0.660 | 0.910 | 0.876 | 0.919 | 0.964 | |
| *Cyrtopodium poecilum* | 0.638 | 0.892 | 0.839 | 0.882 | 0.968 | |
| *Ditassa obcordata* | 0.663 | 0.886 | 0.871 | 0.909 | 0.963 | |
| *Duguetia glabriuscula* | 0.705 | 0.951 | 0.904 | 0.917 | 0.960 | LC |
| *Echinocoryne holosericea* | 0.658 | 0.867 | 0.900 | 0.928 | 0.958 | |
| *Echinocoryne stricta* | 0.677 | 0.837 | 0.880 | 0.932 | 0.970 | |
| *Elephantopus micropappus* | 0.704 | 0.902 | 0.907 | 0.934 | 0.962 | |
| *Eremanthus elaeagnus* | 0.669 | 0.893 | 0.915 | 0.946 | 0.978 | |
| *Eremanthus incanus* | 0.619 | 0.824 | 0.937 | 0.974 | 0.987 | |
| *Eryngium goyazense* | 0.669 | 0.843 | 0.848 | 0.900 | 0.936 | |
| *Eryngium hemisphaericum* | 0.651 | 0.856 | 0.889 | 0.921 | 0.955 | |
| *Evolvulus scoparioides* | 0.626 | 0.829 | 0.850 | 0.929 | 0.980 | |
| *Gochnatia pulchra* | 0.840 | 0.877 | 0.832 | 0.907 | 0.960 | |
| *Gomesa hydrophila* | 0.641 | 0.859 | 0.852 | 0.902 | 0.965 | |
| *Gomphrena pohlii* | 0.705 | 0.832 | 0.909 | 0.952 | 0.974 | |
| *Guatteria sellowiana* | 0.690 | 0.827 | 0.890 | 0.960 | 0.986 | LC |
| *Habenaria goyazensis* | 0.688 | 0.835 | 0.882 | 0.940 | 0.962 | |
| *Habenaria guilleminii* | 0.681 | 0.857 | 0.894 | 0.938 | 0.978 | |
| *Hemipogon irwinii* | 0.640 | 0.883 | 0.907 | 0.959 | 0.992 | LC |
| *Hoehnephytum trixoides* | 0.707 | 0.868 | 0.862 | 0.922 | 0.961 | LC |
| *Hyptis lavandulacea* | 0.676 | 0.843 | 0.854 | 0.911 | 0.951 | |
| *Ichthyothere mollis* | 0.697 | 0.834 | 0.883 | 0.942 | 0.966 | |
| *Ipomoea virgata* | 0.701 | 0.886 | 0.838 | 0.879 | 0.938 | |
| *Justicia clivalis* | 0.693 | 0.880 | 0.865 | 0.905 | 0.963 | NT |
| *Kielmeyera trichophora* | 0.747 | 0.841 | 0.839 | 0.891 | 0.956 | CR |
| *Koanophyllon myrtilloides* | 0.683 | 0.900 | 0.883 | 0.916 | 0.951 | |
| *Lepidaploa tombadorensis* | 0.667 | 0.951 | 0.936 | 0.936 | 0.979 | |
| *Lippia rotundifolia* | 0.679 | 0.827 | 0.884 | 0.960 | 0.993 | |
| *Lychnophora ericoides* | 0.646 | 0.845 | 0.882 | 0.934 | 0.972 | NT |
| *Marlierea lituatinervia* | 0.661 | 0.892 | 0.918 | 0.927 | 0.937 | |
| *Matelea purpurea* | 0.636 | 0.833 | 0.847 | 0.904 | 0.961 | |
| *Mikania acuminata* | 0.684 | 0.908 | 0.918 | 0.928 | 0.933 | |
| *Mikania linearifolia* | 0.673 | 0.888 | 0.919 | 0.935 | 0.960 | |
| *Mikania microcephala* | 0.754 | 0.844 | 0.832 | 0.915 | 0.948 | |
| *Mikania pohlii* | 0.689 | 0.869 | 0.895 | 0.945 | 0.981 | |
| *Mikania ramosissima* | 0.653 | 0.875 | 0.919 | 0.957 | 0.991 | |
| *Mikania reticulata* | 0.646 | 0.889 | 0.867 | 0.906 | 0.955 | |
| *Mikania sessilifolia* | 0.710 | 0.845 | 0.841 | 0.892 | 0.934 | |
| *Mimosa lupinoides* | 0.948 | 0.972 | 0.940 | 0.954 | 0.983 | |
| *Ocotea spixiana* | 0.681 | 0.852 | 0.866 | 0.931 | 0.988 | |
| *Otachyrium grandiflorum* | 0.658 | 0.847 | 0.885 | 0.927 | 0.978 | |
| *Oxypetalum erectum* | 0.688 | 0.835 | 0.869 | 0.928 | 0.950 | |
| *Paspalum flaccidum* | 0.749 | 0.832 | 0.825 | 0.914 | 0.944 | |
| *Pilosocereus aurisetus* | 0.692 | 0.854 | 0.866 | 0.922 | 0.942 | |
| *Piriqueta rosea* | 0.687 | 0.869 | 0.845 | 0.895 | 0.956 | |
| *Polygala nudicaulis* | 0.637 | 0.826 | 0.897 | 0.932 | 0.948 | LC |

| Species | Current | RCP4.5_2050 | | RCP8.5_2050 | | Threat status |
|---|---|---|---|---|---|---|
| | | 2050 | 2080 | 2050 | 2080 | |
| *Pteroglossa macrantha* | 0.716 | 0.939 | 0.883 | 0.909 | 0.949 | NT |
| *Rhynchosia clausseni* | 0.669 | 0.857 | 0.833 | 0.916 | 0.962 | |
| *Rhynchospora warmingii* | 0.627 | 0.877 | 0.873 | 0.912 | 0.961 | LC |
| *Ruellia eurycodon* | 0.642 | 0.903 | 0.927 | 0.956 | 0.981 | |
| *Scleria acanthocarpa* | 0.699 | 0.823 | 0.822 | 0.905 | 0.951 | |
| *Selaginella arroyoana* | 0.675 | 0.969 | 0.917 | 0.948 | 0.991 | VU |
| *Senecio apensis* | 0.814 | 0.896 | 0.884 | 0.942 | 0.979 | |
| *Senecio macrotis* | 0.653 | 0.878 | 0.888 | 0.914 | 0.932 | |
| *Siphanthera dawsonii* | 0.722 | 0.864 | 0.890 | 0.942 | 0.967 | |
| *Skeptrostachys gigantea* | 0.729 | 0.823 | 0.857 | 0.969 | 0.992 | |
| *Stevia heptachaeta* | 0.672 | 0.923 | 0.887 | 0.903 | 0.941 | |
| *Syngonanthus fischerianus* | 0.715 | 0.828 | 0.855 | 0.935 | 0.970 | |
| *Trichogonia hirtiflora* | 0.773 | 0.902 | 0.946 | 0.982 | 0.990 | NT |
| *Trimezia juncifolia* | 0.621 | 0.845 | 0.823 | 0.885 | 0.949 | |
| *Utricularia praelonga* | 0.716 | 0.844 | 0.896 | 0.957 | 0.984 | |
| *Vellozia caruncularis* | 0.677 | 0.828 | 0.821 | 0.894 | 0.934 | |
| *Vernonanthura cymosa* | 0.778 | 0.977 | 0.880 | 0.894 | 0.934 | |
| *Vernonanthura mariana* | 0.690 | 0.843 | 0.863 | 0.920 | 0.942 | |
| *Zornia confusa* | 0.716 | 0.868 | 0.886 | 0.944 | 0.972 | |

🟨: Brazilian Flora 2020 (http://floradobrasil.jbrj.gov.br/)

🟦: MARTINELLI, G.; MESSINA, T.; SANTOS FILHO, L. Livro vermelho da flora do Brasil – Plantas raras do Cerrado; 1. ed. - Rio de Janeiro: Andrea Jakobsson: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro: CNCFlora, 2014. 320p.

🟩 : MENESES, R.I.; BECK, S. Especies Amenazadas de la Flora de Bolivia. Herbario Nacional de Bolivia. La Paz, 2005.

SOURCE: the author (2018)

# REFERENCES

ACEVEDO, P.; JIMÉNEZ-VALVERDE, A.; LOBO, J. M.; REAL R. Predictor weighting and geographical background delimitation: two synergetic sources of uncertainty when assessing species sensitivity to climate change. **Climatic Change**, v. 145, p. 131–143, 2017.

AGUIAR, L. M. S.; BERNARD, E.; RIBEIRO, V.; MACHADO. R. B.; JONES, G. Should I stay or should I go? Climate change effects on the future of Neotropical savannah bats. **Global Ecology and Conservation**, v. 5, p. 22–33, 2016.

AGUIRRE-GUTIÉRREZ, J.; CARVALHEIRO, L. G.; POLCE, C.; VAN LOON, E. E.; RAES, N.; REEMER, M.; BIES, J. C. Fit-for-Purpose: Species Distribution Model Performance Depends on Evaluation Criteria – Dutch Hoverflies as a Case Study. **PLoS ONE**, v. 8, n. 5, p. e63708, 2013.

ALBUQUERQUE, F.; BEIER, P. Predicted rarity-weighted richness, a new tool to prioritize sites for species representation. **Ecology and Evolution**, v. 6, n. 22, p. 8107–8114, 2016.

ALEXANDRATOS, N.; BRUINSMA, J. **World agriculture towards 2030/2050: the 2012 revision**: ESA Working paper No. 12-03. FAO, Rome: 2012.

ALLOUCHE, O.; TSOAR, A.; KADMON, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. **Journal of Applied Ecology**, v. 43, n. 6, p. 1223–1232, 2006.

ALVES, R. J. V.; KOLBEK, J. Can campo rupestre vegetation be floristically delimited based on vascular plant genera? **Plant Ecology**, v. 207, n. 1, p. 67–79, 2010.

ALVES, R. SILVA, N. G.; OLIVEIRA, J.A.; MEDEIROS, D. Circumscribing campo rupestre; megadiverse Brazilian rocky montane savanas. **Brazilian Journal of Biology**, v. 74, n. 2, p. 355–362, 2014.

AMARAL, A.G.; MUNHOZ, C. B. R.; WALTER, B. M. T.; AGUIRRE-GUTIÉRREZ, J.; RAES, N. Richness pattern and phytogeography of the Cerrado herb-shrub flora and implications for conservation. **Journal of Vegetation Science**, v. 28, n. 4, p. 848–858, 2017.

ANAGIOSPERM PHYLOGENY GROUP. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society. v. 161, p. 105–121, 2009.

ANDERSON, D. W. The effect of parent material and soil development on nutrient cycling in temperate ecosystems. **Biogeochemistry**, v. 5, n. 1, p. 71–97, 1988.

ANDERSON, R. P.; GONZALEZ, I. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. **Ecological Modelling**, v. 222, n. 15, p. 2796–2811, 2011.

ANDERSON, R. P.; RAZA, A. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela: Effect of study region on models of distributions. **Journal of Biogeography**, v. 37, n. 7, p. 1378–1393, 2010.

ARAÚJO, M. B.; WHITTAKER, R. B.; LADLE, R. J.; ERHARD, M. Reducing uncertainty in projections of extinction risk from climate change: Uncertainty in species' range shift projections. **Global Ecology and Biogeography**, v. 14, n. 6, p. 529–538, 2005.

ARAÚJO, M.; NEW, M. Ensemble forecasting of species distributions. **Trends in Ecology & Evolution**, v. 22, n. 1, p. 42–47, 2007.

ARMENTERAS, D.; ESPELTA, J. M.; RODRÍGUEZ, N.; RETANA, J. Deforestation dynamics and drivers in different forest types in Latin America: Three decades of studies (1980–2010). **Global Environmental Change**, v. 46, p. 139–147, 2017.

ARNOLD, R. W. Pedology and pedogenesis. In: CHESWORTH, W. (Ed.). **Encyclopedia of Soil Science Arnold**. Dordrecht, Netherlands: Springer, 2008. p. 902.

AUSTIN, M. P. Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. In: SCOTT, J. M. et al. (Eds.). **Predicting Species Occurrences: Issues of Accuracy and Scale**. Covelo, California: Island Press, 2002. p. 73–82.

AUSTIN, M. P.; SMITH, T. M. A new model for the continuum concept. **Vegetatio**, v. 83, n. 1–2, p. 35–47, 1989.

AUSTIN, M. P.; VAN NIEL, K. P. Impact of landscape predictors on climate change modelling of species distributions: a case study with Eucalyptus fastigata in southern New South Wales, Australia: Impact of landscape predictors on climate change modelling. **Journal of Biogeography**, v. 38, n. 1, p. 9–19, 2011.

AUSTIN, M. Species distribution models and ecological theory: A critical assessment and some possible new approaches. **Ecological Modelling**, v. 200, n. 1–2, p. 1–19, 2007.

BALDI, G.; TEXEIRA, M.; MARTIN, O. A.; GRAU, H. R.; JOBBÁGY, E. G. Opportunities drive the global distribution of protected areas. **PeerJ**, v. 5, p. e2989, 2017.

BARBET-MASSIN, M.; JIGUET, F.; ALBERT, C H.; THUILLER, W. Selecting pseudo-absences for species distribution models: how, where and how many?: How to use pseudo-absences in niche modelling? **Methods in Ecology and Evolution**, v. 3, n. 2, p. 327–338, 2012.

BARBET-MASSIN, M.; THUILLER, W.; JIGUET, F. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? **Ecography**, v. 33, n. 5, p. 878–886, 2010.

BÁRCENAS-ARGÜELLO, M. L.; GUTIÉRREZ-CASTORENA, M. DEL C.; TERRAZAS, T. The Role of Soil Properties in Plant Endemism – A Revision of Conservation Strategies. In: SORIANO, M. C. M. C. H. (Ed.). **Soil Processes and Current Trends in Quality Assessment**. Rijeka, Croatia: InTech, 2013.

BARQUET, K.; LUJALA, P.; RØD, J. K. Transboundary conservation and militarized interstate disputes. **Political Geography**, v. 42, p. 1–11, 2014.

BARVE, N.; BARVE, V.; JIMÉNEZ-VALVERDE, A.; LIRA-NORIEGA, A.; MAHER S. P.; PETERSON, A. T.; SOBERÓN, J.; VILLALOBOS, F. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. **Ecological Modelling**, v. 222, n. 11, p. 1810–1819, 2011.

BATEMAN, B. L.; PIDGEON, A.M.; RADELOFF, V. C.; ALLSTADT, A. J.; AKÇAKAYA, H. R.; THOGMARTIN, W. E.; VAVRUS, S. J.; HEGLUND, P. J. The importance of range edges for an irruptive species during extreme weather events. **Landscape Ecology**, v. 30, n. 6, p. 1095–1110, 2015.

BAUMANN, M.; ISRAEL, C.; PIQUER-RODRÍGUEZ, M.; GAVIER-PIZARRO, G.; VOLANTE, J. N.; KUEMMERLE, T. Deforestation and cattle expansion in the Paraguayan Chaco 1987–2012. **Regional Environmental Change**, v. 17, n. 4, p. 1179–1191, abr. 2017.

BEALE, C. M.; LENNON, J. J. Incorporating uncertainty in predictive species distribution modelling. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 367, n. 1586, p. 247–258, 2012.

BEAUREGARD, F.; BLOIS, S. Beyond a climate-centric view of plant distribution: edaphic variables add value to distribution models. **PLoS ONE**, v. 9, n. 3, p. e92642, 2014.

BECK, J.; BÖLLER, M.; ERHARDT, A.; SCHWANGHART, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. **Ecological Informatics**, v. 19, p. 10–15, 2014.

BECK, S. G. Las regiones y zonas de vegetación. In: JØRGENSEN, P. M.; NEE, M. H.; BECK, S. G. (Eds.). **Catálogo de las plantas vasculares de Bolivia**. St. Louis, Missouri: Missouri Botanical Garden Press, 2015. p. 3–20.

BERTRAND, R.; PEREZ, V.; GÉGOUT, J.-C. Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: the case of Quercus pubescens in France. **Global Change Biology**, v. 18, n. 8, p. 2648–2660, 2012.

BOYLE, B.; HOPKINS, N.; LU, Z.; GARAY, J. A. R.; MOZZHERIN, D.; REES, T.; MATASCI, N.; NARRO, M. L.; PIEL, W. H.; MCKAY, S. J. The taxonomic name resolution service: an online tool for automated standardization of plant names. **BMC bioinformatics**, v. 14, n. 1, p. 16, 2013.

BREEMEN, N. VAN; BUURMAN, P. **Soil formation**. Dordrecht; Boston: Kluwer Academic, 2002.

BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5–32, 2001.

BRIDGEWATER, S.; RATTER, J. A.; RIBEIRO, J. F. Biogeographic patterns, β-diversity and dominance in the cerrado biome of Brazil. **Biodiversity & Conservation**, v. 13, n. 12, p. 2295–2317, 2004.

CAHILL, A. E.; AIELLO-LAMMENS, M. E.; FISHER-REID, M. C.; HUA, X.; KARANEWSKY, C. J.; YEONG RYU, H.; SBEGLIA, G. C.; SPAGNOLO, F.; WALDRON, J. B.; WARSI, O.; WIENS, J. J. How does climate change cause extinction? **Proceedings of the Royal Society B: Biological Sciences**, v. 280, n. 1750, p. 20121890–20121890, 2012.

CARRETERO, M. A.; SILLERO, N. Evaluating how species niche modelling is affected by partial distributions with an empirical case. **Acta Oecologica**, v. 77, p. 207–216, 2016.

CASAJUS, N.; PÉRIÉ, C.; LOGAN, T.; LAMBERT, M. C.; DE BLOIS, S.. BERTEAUX, D. An Objective Approach to Select Climate Scenarios when Projecting Species Distribution under Climate Change. **PLOS ONE**, v. 11, n. 3, p. e0152495, 2016.

CEDDIA, M. B.; VIEIRA, S. R.; VILLELA, A. L. O.; MOTA, L. S., ANJOS, L. H. C.; CARVALHO, D. F. Topography and spatial variability of soil physical properties. **Scientia Agricola**, v. 66, n. 3, p. 338–352, 2009.

CHANNELL, R.; LOMOLINO, M. V. Dynamic biogeography and conservation of endangered species. **Nature**, v. 403, n. 6765, p. 84, 2000.

CHAUDHARY, A.; PFISTER, S.; HELLWEG, S. Spatially Explicit Analysis of Biodiversity Loss Due to Global Agriculture, Pasture and Forest Land Use from a Producer and Consumer Perspective. **Environmental Science & Technology**, v. 50, n. 7, p. 3928–3936, 2016.

COLLEVATTI, R. G.; LIMA-RIBEIRA, M. S.; SOUZA-NETO, A. C.; FRANCO, A. A.; OLIVEIRA, G.; TERRIBILE, L. C. Recovering the demographical history of a Brazilian Cerrado tree species Caryocar brasiliense: coupling ecological niche modeling and coalescent analyses. **Natureza & Conservação**, v. 10, n. 2, p. 169–176, 2012.

COLWELL, R. K.; RANGEL, T. F. Hutchinson's duality: the once and future niche. **Proceedings of the National Academy of Sciences**, v. 106, n. Supplement 2, p. 19651–19658, 2009.

COUDUN, C.; GÉGOUT, J. C.; PIEDALLU, C.; RAMEAU, J. C. Soil nutritional factors improve models of plant species distribution: an illustration with Acer campestre (L.) in France. **Journal of Biogeography**, v. 33, n. 10, p. 1750–1763, 2006.

CRAIN, B. J.; WHITE, J. W.; STEINBERG, S. J. Geographic discrepancies between global and local rarity richness patterns and the implications for conservation. **Biodiversity and Conservation**, v. 20, n. 14, p. 3489–3500, 2011.

CRUZ-CÁRDENAS, G.; LÓPEZ-MATA, L.; VILLASEÑOR, J. L.; ORTIZ, E. Potential species distribution modeling and the use of principal component analysis as predictor variables. **Revista Mexicana de Biodiversidad**, v. 85, n. 1, p. 189–199, 2014.

DALLIMER, M.; STRANGE, N. Why socio-political borders and boundaries matter in conservation. **Trends in Ecology & Evolution**, v. 30, n. 3, p. 132–139, 2015.

DE OLIVEIRA, G.; LIMA-RIBEIRO, M. S.; TERRIBILE, L. C.; DOBROVOLSKI, R.; TELLES, M. P. D. C.; DINIZ-FILHO, J. A. F. Conservation biogeography of the Cerrado's wild edible plants under climate change: Linking biotic stability with agricultural expansion. **American Journal of Botany**, v. 102, n. 6, p. 870–877, 2015.

DIEKMANN, M.; MICHAELIS, J.; PANNEK, A. Know your limits – The need for better data on species responses to soil variables. **Basic and Applied Ecology**, v. 16, n. 7, p. 563–572. 2015.

DIFFENBAUGH, N. S.; FIELD, C. B. Changes in Ecologically Critical Terrestrial Climate Conditions. **Science**, v. 341, n. 6145, p. 486–492, 2013.

DINIZ-FILHO, J. A. F.; BINI, L. M.; OLIVEIRA, G.; BARRETO, B. DE S.; SILVA, M. M. F. P.; TERRIBLE, L. C.; RANGEL, T. F. L. V. B.; PINTO, M. P.; SOUSA, N. P. R.; VIEIRA, L. C. G.; MELO, A. S.; DE MARCO JÚNIOR, P.; VIEIRA, C. M.; BALMIRES, D.; BASTOS, R. P.; CARVALHO, P.; FERREIRA, L. G.; TELLES, M. P. DE C.; RODRIGUES, F. M.; SILVA, D. M.; SILVA JÚNIOR, N. J.; SOARES, T. N. Macroecologia, biogeografia e áreas prioritárias para conservação no cerrado. **Oecologia Brasiliensis**, v. 13, n. 03, p. 470–497, 2009.

DINIZ-FILHO, J. A. F.; BINI, L. M.; RANGEL, T. F. L. V. B.; LOYOLA, R. D.; HOF, C.; NOGUÉS-BRAVO, D.; ARAÚJO, M. B. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. **Ecography**, v. 32, n. 6, p. 897–906, 2009.

DINIZ-FILHO, J. A. F.; OLIVEIRA, G. DE, BINI, L. M.; LOYOLA, R. D.; NABOUT, J. C.; RANGEL, T. F. L. V. B. Biogeografia da conservação e mudanças climáticas no cerrado brasileiro. **Natureza & Conservação**, v. 7, n. 2, 2009.

DONOGHUE, M. J.; EDWARDS, E. J. Biome shifts and niche evolution in plants. **Annual Review of Ecology, Evolution, and Systematics**, v. 45, p. 547–572, 2014.

DORMANN, C. F.; ELITH, J.; BACHER, S.; BUCHMANN, C.; CARL, G.; CARRÉ, G.; MARQUÉZ, J. R. G.; GRUBER, B.; LAFOURCADE, B.; LEITÃO, P. J.; MÜNKEMÜLLER, T., MCCLEAN, C.; OSBORNE, P. E.; REINEKING, B.; SCHRÖDER, B.; SKIDMORE, A. K.; ZURELL, D.; LAUTENBACH, S. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, v. 36, n. 1, p. 27–46, 2013.

DUAN, R. Y.; KONG, X. Q.; HUANG, M. Y.; FAN, W. Y.; WANG, Z. G. The Predictive Performance and Stability of Six Species Distribution Models. **PLoS ONE**, v. 9, n. 11, p. e112764, 2014.

DUBUIS, A.; GIOVANETTINA, S.; PELLISSIER, L.; POTTIER, J.; VITTOZ, P.; GUISAN, A. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. **Journal of Vegetation Science**, v. 24, n. 4, p. 593–606, 2013.

DUBUIS, A.; POTTIER, J.; RION, V.; PELLISSIER, L.; THEURILLAT, J. P.; GUISAN, A. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches: Predicting plant species richness. **Diversity and Distributions**, v. 17, n. 6, p. 1122–1131, 2011.

DUPIN, M.; REYNAUD, P.; JAROŠÍK, V.; BAKER, R.; BRUNEL, S.; EYRE, D.; PERGL, J.; MAKOWSKI, D. Effects of the Training Dataset Characteristics on the Performance of Nine Species Distribution Models: Application to Diabrotica virgifera virgifera. **PLoS ONE**, v. 6, n. 6, p. e20957, 2011.

ELITH, J.; GRAHAM, C. H.; ANDERSON, R. P.; DUDÍK, M.; FERRIER, S.; GUISAN, A.; HIJMANS, R. J.; HUETTMANN, F.; LEATHWICK, J. R.; LEHMANN, A.; LI, J.; LOHMANN, L. G.; LOISELLE, B. A.; MANION, G.; MORITZ, C.; NAKAMURA, M.; NAKAZAWA, Y.; OVERTON, J. M. M.; PETERSON, A. T.; PHILLIPS, S. J.; RICHARDSON, K.; SCACHETTI-PEREIRA, R.; SCHAPIRE, R. E.; SOBERÓN, J.; WILLIAMS, S.; WISZ, M. S.; ZIMMERMANN, N. E. Novel methods improve prediction of species' distributions from occurrence data. **Ecography**, v. 29, n. 2, p. 129–151, 2006.

ELITH, J.; PHILLIPS, S. J.; HASTIE, T.; DUDÍK, M.; CHEE, Y. E.; YATES, C. J. A statistical explanation of MaxEnt for ecologists: Statistical explanation of MaxEnt. **Diversity and Distributions**, v. 17, n. 1, p. 43–57, 2011.

FAHRIG, L. How much habitat is enough? **Biological conservation**, v. 100, n. 1, p. 65–74, 2001.

FAO. **The future of food and agriculture: trends and challenges**. Rome: Food and Agriculture Organization of the United Nations, 2017.

FEHLENBERG, V.; BAUMANN, M.; GASPARRI, N. I.; PIQUER-RODRIGUEZ, M.; GAVIER-PIZARRO, G.; KUEMMERLE, T. The role of soybean production as an underlying driver of deforestation in the South American Chaco. **Global Environmental Change**, v. 45, p. 24–34, 2017.

FEI, S.; DESPREZ, J. M.; POTTER, K. M.; JO, I.; KNOTT, J. A.; OSWALT, C. M. Divergence of species responses to climate change. **Science Advances**, v. 3, p. e1603055, 2017.

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.

FERNÁNDEZ-DELGADO, M.; CERNADAS, E.; BARRO, S.; AMORIM, D. Do we need hundreds of classifiers to solve real world classification problems? **The Journal of Machine Learning Research**, v. 15, n. 1, p. 3133–3181, 2014.

FERNÁNDEZ-LLAMAZARES, Á.; HELLE, J.; EKLUND, J.; BALMFORD, A.; MORAES, R. M.; REYES-GARCÍA, V.; CABEZA, M. New law puts Bolivian biodiversity hotspot on road to deforestation. **Current Biology**, v. 28, n. 1, p. R15–R16, 2018.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas: NEW CLIMATE SURFACES FOR GLOBAL LAND AREAS. **International Journal of Climatology**, 2017.

FISICHELLI, N.; WRIGHT, A.; RICE, K.; MAU, A.; BUSCHENA, C.; REICH, P.B. First-year seedlings and climate change: species-specific responses of 15 North American tree species. **Oikos**, v. 123, n. 11, p. 1331–1340, 2014.

FOLEY, J. A.; DEFRIES, R.; ASNER, G. P.; BARFORD, C.; BONAN, G.; CARPENTER, S. R.; CHAPIN, F. S.; COE, M. T.; DAILY, G. C.; GIBBS, H. K.; HELKOWSKI, J. H.; HOLLOWAY, T.; HOWARD, E. A.; KUCHARIK, C. J.; MONFREDA, C.; PATZ, J. A.; PRENTICE, I. C.; RAMANKUTTY, N.; SNYDER, P. K. Global Consequences of Land Use. **Science**, v. 309, n. 5734, p. 570–574, 2005.

FORDHAM, D. A.; BERTELSMEIER, C.; BROOK, B. W.; EARLY, R.; NETO, D.; BROWN, S. C.; OLLIER, S.; ARAÚJO, M. B. How complex should models be? Comparing correlative and mechanistic range dynamics models. **Global Change Biology**, 2017.

FOURCADE, Y.; ENGLER, J. O.; RÖDDER, D.; SECONDI, J. Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. **PLoS ONE**, v. 9, n. 5, p. e97122, 2014.

FOX, J.; WEISBERG, S. **An R Companion to Applied Regression.** Second edition ed. Thousand Oaks, California, USA, p.472, 2011

FRANÇOSO, R. D.; HAIDAR, R. F.; MACHADO, R. B. Tree species of South America central savanna: endemism, marginal areas and the relationship with other biomes. **Acta Botanica Brasilica**, v. 30, n. 1, p. 78–86, 2016.

FRANKLIN, J. **Mapping Species Distributions: spatial Inference and prediction**. United States of America: Cambridge University Press, 2009.

GASTON, K. J. Conservation and rarity. In: **Rarity**. Dordrecht: Springer Netherlands, 1994. p. 136–152.

GASTON, K. J. **The structure and dynamics of geographic ranges**. Oxford: Oxford University Press, 2003.

GIBBS, J. P. Distribution of woodland amphibians along a forest fragmentation gradient. **Landscape Ecology**, v. 13, n. 4, p. 263–268, 1998.

GOODWIN, Z. A.; HARRIS, D. J.; FILER, D.; WOOD, J. R.; SCOTLAND, R. W. Widespread mistaken identity in tropical plant collections. **Current Biology**, v. 25, n. 22, p. R1066–R1067, 2015.

GRAHAM, C. H.; ELITH, J.; HIJMANS, R. J.; GUISAN, A.; PETERSON, A. T.; LOISELLE, B. A. The influence of spatial errors in species occurrence data used in distribution models: Spatial error in occurrence data for predictive modelling. **Journal of Applied Ecology**, v. 45, n. 1, p. 239–247, 2007.

GRAHAM, C. H.; MORITZ, C.; WILLIAMS, S. E. Habitat history improves prediction of biodiversity in rainforest fauna. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 3, p. 632–636, 2006.

GRAMMONT, P. C.; CUARÓN, A. D. An evaluation of threatened species categorization systems used on the American continent. **Conservation Biology**, v. 20, n. 1, p. 14–27, 2006.

GRAY, C. L.; HILL, S. L. L.; NEWBOLD, T.; HUDSON, L. N.; BÖRGER, L.; CONTU, S.; HOSKINS, A. J.; FERRIER, S.; PURVIS, A.; SCHARLEMANN, J. P. W. Local biodiversity is higher inside than outside terrestrial protected areas worldwide. **Nature Communications**, v. 7, p. 12306, 2016.

GRÖGER, A.; HUBER, O. Rock outcrop habitats in the Venezuelan Guayana lowlands: their main vegetation types and floristic components. **Brazilian Journal of Botany**, v. 30, n. 4, p. 599–609, 2007.

GUERIN, G. R.; LOWE, A. J. 'Sum of inverse range-sizes' (SIR), a biodiversity metric with many names and interpretations. **Biodiversity and Conservation**, v. 24, n. 11, p. 2877–2882, 2015.

GUERIN, G. R.; RUOKOLAINEN, L.; LOWE, A. J. A georeferenced implementation of weighted endemism. **Methods in Ecology and Evolution**, v. 6, n. 7, p. 845–852, 2015.

GUILLERA-ARROITA, G.; LAHOZ-MONFORT, J. J.; ELITH, J. Maxent is not a presence-absence method: a comment on Thibaud *et al* . **Methods in Ecology and Evolution**, v. 5, n. 11, p. 1192–1197, 2014.

GUISAN, A.; EDWARDS, T. C.; HASTIE, T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. **Ecological modelling**, v. 157, n. 2, p. 89–100, 2002.

GUISAN, A.; RAHBEK, C. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages: Predicting spatio-temporal patterns of species assemblages. **Journal of Biogeography**, v. 38, n. 8, p. 1433–1444, 2011.

HANBERRY, B. B.; HE, H. S.; PALIK, B. J. Pseudoabsence generation strategies for species distribution models. **PLoS ONE**, v. 7, n. 8, p. e44486, 2012.

HARRIS, R. M. B.; PORFIRIO, L. L.; HUGH, S.; LEE, G.; BINDOFF, N. L.; MACKEY, B.; BEETON, N. J. To be or not to be? Variable selection can change the projected fate of a threatened species under future climate. **Ecological Management & Restoration**, p. 1–5. 2013.

HASTIE, T.; TIBSHIRANI, R. Generalized Aditive Models. **Statistical Science**, v. 1, n. 3, p. 297–318, 1986.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning Data Mining, Inference, and Prediction**. 2. ed. Springer-Verlag New York, 2009.

HAUTIER, Y., TILMAN, D., ISBELL, F., SEABLOOM, E. W., BORER, E. T., REICH, P. B. Anthropogenic environmental changes affect ecosystem stability via biodiversity. **Science**, v. 348, n. 6232, p. 333–336, 2015.

HENGL, T.; DE JESUS, J. M.; MACMILLAN, R. A.; BATJES, N. H.; HEUVELINK, G. B. M.; RIBEIRO, E.; SAMUEL-ROSA, A.; KEMPEN, B.; LEENAARS, J. G. B.; WALSH, M. G.; GONZALEZ, M. R. SoilGrids1km — Global Soil Information Based on Automated Mapping. **PloS ONE**, v. 9, n. 8, p. e105992, 2014.

HENGL, T.; DE JESUS, J.M.; HEUVELINK, G.B.; GONZALEZ, M.R.; KILIBARDA, M.; BLAGOTIĆ, A.; SHANGGUAN, W.; WRIGHT, M.N.; GENG, X.; BAUER-MARSCHALLINGER, B.; GUEVARA, M. A.; VARGAS, R.; MACMILLAN, R. A.; BATJES, N. H.; LEENAARS, J. G. B.; RIBEIRO, E.; WHEELER, I.; MANTEL, S.; KEMPEN, B. SoilGrids250m: Global gridded soil information based on machine learning. **PloS one**, v. 12, n. 2, p. e0169748, 2017.

HENGL, T.; HEUVELINK, G. B. M.; KEMPEN, B.; LEENAARS, J. G. B.; WALSH, M. G.; SHEPHERD, K. D.; SILA, A.; MACMILLAN, R. A.; MENDES DE JESUS, J.; TAMENE, L.; TONDOH, J. E. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. **PloS ONE**, v. 10, n. 6, p. e0125814, 2015.

HEUVELINK, G. B; WEBSTER, R. Modelling soil variation: past, present, and future. **Geoderma**, v. 100, n. 3–4, p. 269–301, 2001.

HIJMANS, R. J. **raster: Geographic Data Analysis and Modeling. R package version 2.5-2**, 2015. Disponível em: <https://CRAN.R-project.org/package=raster>

HIJMANS, R. J.; PHILLIPS, S.; LEATHWICK, J.; ELITH, J. **dismo: Species Distribution Modeling. R package version 1.0-15**. 2017. Disponível em: <https://CRAN.R-project.org/package=dismo>

HIRZEL, A. H.; HAUSSER, J.; CHESSEL, D.; PERRIN, N. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? **Ecology**, v. 83, n. 7, p. 2027–2036, 2002.

HOBAN, S.; SCHLARBAUM, S. Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. **Biological Conservation**, v. 177, p. 90–99, 2014.

HORNIK, K.; MEYER, D.; KARATZOGLOU, A. Support vector machines in R. **Journal of statistical software**, v. 15, n. 9, p. 1–28, 2006.

HORTAL, J.; JIMÉNEZ-VALVERDE, A.; GÓMEZ, J. F.; LOBO, J. M.; BASELGA, A. Historical bias in biodiversity inventories affects the observed environmental niche of the species. **Oikos**, v. 117, n. 6, p. 847–858, 2008.

HUNTER, M. L.; HUTCHINSON, A. The Virtues and Shortcomings of Parochialism: Conserving Species That Are Locally Rare, but Globally Common. **Conservation Biology**, v. 8, n. 4, p. 1163–1165, 1994.

HURTT, G. C.; CHINI, L. P.; FROLKING, S.; BETTS, R. A.; FEDDEMA, J.; FISCHER, G.; FISK, J. P.; HIBBARD, K.; HOUGHTON, R. A.; JANETOS, A.; JONES, C. D.; KINDERMANN, G.; KINOSHITA, T.; KLEIN GOLDEWIJK, K.; RIAHI, K.; SHEVLIAKOVA, E.; SMITH, S.; STEHFEST, E.; THOMSON, A.; THORNTON, P.; VAN VUUREN, D.P.; WANG, Y. P. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. **Climatic Change**, v. 109, n. 1–2, p. 117–161, 2011.

HUTCHINSON, G. E. **An Introduction to Population Ecology**. New Haven, CT: Yale University Press, 1978.

IBGE. **Manual técnico da vegetação brasileira**. 2. ed. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2012.

IBISCH, P. L., BECK, S. G., GERKMANN, B., CARRETERO, A. La diversidad biológica: ecorregiones y ecosistemas. In: IBISCH, P. L.; MÉRIDA, G. (Eds.). . **La diversidad biológica: ecorregiones y ecosistemas**. Santa Cruz, Bolivia: Fundación Amigos de la Naturaleza, 2003. p. 47–88.

IPCC. **Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change**. Geneva, Switzerland: Intergovernmental Panel on Climate Change, 2014.

IŞIK, K. Rare and endemic species: why are they prone to extinction? **Turkish Journal of Botany**, v. 35, n. 4, p. 411–417, 2011.

IUCN STANDARDS AND PETITIONS SUBCOMMITTEE. **Guidelines for Using the IUCN Red List Categories and Criteria**. 13. Ed, p. 108, 2017

JABLONSKI, D.; ROY, K.; VALENTINE, J. W. Out of the Tropics: Evolutionary Dynamics of the Latitudinal Diversity Gradient. **Science**, v. 314, n. 5796, p. 102–106, 2006.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning**. New York, NY: Springer New York, 2013. v. 103

JANSSON, R. Global patterns in endemism explained by past climatic change. **Proceedings of the Royal Society of London B: Biological Sciences**, v. 270, n. 1515, p. 583–590, 2003.

JARVIS, A.; TOUVAL, J. L.; SCHMITZ, M. C.; SOTOMAYOR, L.; HYMAN, G. G. Assessment of threats to ecosystems in South America. **Journal for Nature Conservation**, v. 18, n. 3, p. 180–188, 2010.

JENNY, H. **Factors of soil formation: a system of quantitative pedology**. New York: Dover, 1994.

JEZKOVA, T.; WIENS, J. J. Rates of change in climatic niches in plant and animal populations are much slower than projected climate change. **Proceedings of the Royal Society B: Biological Sciences**, v. 283, n. 1843, p. 20162104, 2016.

JIMÉNEZ-VALVERDE, A.; LOBO, J. M. Threshold criteria for conversion of probability of species presence to either–or presence–absence. **Acta Oecologica**, v. 31, n. 3, p. 361–369, 2007.

JIMÉNEZ-VALVERDE, A.; LOBO, J. M.; HORTAL, J. Not as good as they seem: the importance of concepts in species distribution modelling. **Diversity and Distributions**, v. 14, n. 6, p. 885–890, 2008.

JIMÉNEZ-VALVERDE, A.; LOBO, J.; HORTAL, J. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. **Community Ecology**, v. 10, n. 2, p. 196–205, 2009.

JUNK, W. J.; CUNHA, C. N. DE. Pantanal: a large South American wetland at a crossroads. **Ecological Engineering**, v. 24, n. 4, p. 391–401, 2005.

KAMATH, C. **Scientific data mining: a practical perspective**. Philadelphia: Society for Industrial and Applied Mathematics, 2009.

KARATZOGLOU, A.; SMOLA, A.; HORNIK, K.; ZEILEIS, A. kernlab - An S4 Package for Kernel Methods in R. **Journal of Statistical Software**, v. 11, n. 9, p. 1–20, 2004.

KARGER, D. N.; CONRAD, O.; BÖHNER, J.; KAWOHL, T.; KREFT, H.; SORIA-AUZA, R. W.; ZIMMERMANN, N. E.; LINDER, H. P.; KESSLER, M. Climatologies at high resolution for the earth’s land surface areas. **Scientific Data**, v. 4, p. 170122, 2017.

KARK, S.; TULLOCH, A.; GORDON, A.; MAZOR, T.; BUNNEFELD, N.; LEVIN, N. Cross-boundary collaboration: key to the conservation puzzle. **Current Opinion in Environmental Sustainability**, v. 12, p. 12–24, 2015.

KARP, D. S.; ROMINGER, A. J.; ZOOK, J.; RANGANATHAN, J.; EHRLICH, P. R.; DAILY, G. C. Intensive agriculture erodes β-diversity at large scales. **Ecology Letters**, v. 15, n. 9, p. 963–970, set. 2012.

KATTAN, G. Rarity and vulnerability: the birds of the Cordillera Central of Colombia. **Conservation Biology**, v. 6, n. 1, p. 64–70, 1992.

KEITH, D. A.; MAHONY, M.; HINES, H.; ELITH, J.; REGAN, T. J.; BAUMGARTNER, J. B.; HUNTER, D.; HEARD, G. W.; MITCHELL, N. J.; PARRIS, K. M.; PENMAN, T.; SCHEELE, B.; SIMPSON, C. C.; TINGLEY, R.; TRACY, C. R.; WEST, M.; AKÇAKAYA, H. R. Detecting Extinction Risk from Climate Change by IUCN Red List

Criteria: Detecting Extinction Risk from Climate Change. **Conservation Biology**, v. 28, n. 3, p. 810–819, 2014.

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. New York, NY: Springer New York, 2013.

KUNIN, W. E.; GASTON, K. J. (EDS.). **The Biology of Rarity**. Dordrecht: Springer Netherlands, 1997.

LE SAOUT, S.; HOFFMANN, M.; SHI, Y.; HUGHES, A.; BERNARD, C.; BROOKS, T. M.; BERTZKY, B.; BUTCHART, S. H.; STUART, S. N.; BADMAN, T. Protected areas and effective biodiversity conservation. **Science**, v. 342, n. 6160, p. 803–805, 2013.

LENTH, R. V. Least-Squares Means: The R Package lsmeans. **Journal of Statistical Software**, v. 69, n. 1, 2016.

LEPPIG, G.; WHITE, J. W. Conservation of peripheral plant populations in California. **Madroño**, v. 53, n. 3, p. 264–274, 2006.

LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R News**, v. 2, n. 3, p. 18–22, 2002.

LIM, M. Governance criteria for effective transboundary biodiversity conservation. **International Environmental Agreements: Politics, Law and Economics**, v. 16, n. 6, p. 797–813, 2016.

LIU, C.; WHITE, M.; NEWELL, G. Measuring and comparing the accuracy of species distribution models with presence-absence data. **Ecography**, v. 34, n. 2, p. 232–243, 2011.

LLOYD, J.; DOMINGUES, T.F.; SCHRODT, F.; ISHIDA, F.Y.; FELDPAUSCH, T.R.; SAIZ, G.; QUESADA, C.A.; SCHWARZ, M.; TORELLO-RAVENTOS, M.; GILPIN, M.; MARIMON, B. S.; MARIMON-JUNIOR, B. H.; RATTER, J. A.; GRACE, J.; NARDOTO, G. B.; VEENENDAAL, E.; ARROYO, L.; VILLARROEL, D.; KILLEEN, T. J.; STEININGERA, M.; PHILLIPS, O. L. Edaphic, structural and physiological contrasts across Amazon Basin forest–savanna ecotones suggest a role for potassium as a key modulator of tropical woody vegetation structure and function. **Biogeosciences**, v. 12, n. 22, p. 6529–6571, 18. 2015.

LOBO, J. M.; JIMÉNEZ-VALVERDE, A.; HORTAL, J. The uncertain nature of absences and their importance in species distribution modelling. **Ecography**, v. 33, n. 1, p. 103–114, 2010.

LOBO, J. M.; JIMÉNEZ-VALVERDE, A.; REAL, R. AUC: a misleading measure of the performance of predictive distribution models. **Global Ecology and Biogeography**, v. 17, n. 2, p. 145–151, 2008.

LÓPEZ-HOFFMAN, L.; VARADY, R.G.; FLESSA, K.W.; BALVANERA, P. Ecosystem services across borders: a framework for transboundary conservation policy. **Frontiers in Ecology and the Environment**, v. 8, n. 2, p. 84–91, 2010.

LORENA, A. C.; JACINTHO, L. F. O.; SIQUEIRA, M. F.; GIOVANNI, R. D.; LOHMANN, L. G.; DE CARVALHO, A. C. P. L. F.; YAMAMOTO, M. Comparing machine learning classifiers in potential distribution modelling. **Expert Systems with Applications**, v. 38, n. 5, p. 5268–5275, 2011.

LUOTO, M.; PÖYRY, J.; HEIKKINEN, R. K.; SAARINEN, K. Uncertainty of bioclimate envelope models based on the geographical distribution of species. **Global Ecology and Biogeography**, v. 14, n. 6, p. 575–584, 2005.

MACE, G. M. Whose conservation? **Science**, v. 345, n. 6204, p. 1558–1560, 2014.

MAMANI, F.; POZO, P.; SOTO, D.; VILLARROEL, D.; WOOD, J. R. I. **Libro Rojo de las Plantas de los Cerrados del Oriente Boliviano**. Santa Cruz, Bolivia: John R I Wood, 2010.

MAMANI, F.; POZO, P.; SOTO, D.; VILLARROEL, D.; WOOD, J. R. I.; MAMANI, F. **Guía Darwin de las plantas de los cerrados de la Chiquitania**. Santa Cruz, Bolivia: Wood, John R I, 2011.

MARÉCHAUX, I.; RODRIGUES, A. S. L.; CHARPENTIER, A. The value of coarse species range maps to inform local biodiversity conservation in a global context. **Ecography**, 2016.

MARGULES, C. R.; PRESSEY, R. L. Systematic conservation planning. **Nature**, v. 405, n. 6783, p. 243–253, 2000.

MARIA, B.; UDO, S. Why input matters: Selection of climate data sets for modelling the potential distribution of a treeline species in the Himalayan region. **Ecological Modelling**, v. 359, p. 92–102, 2017.

MARTINELLI, G.; MESSINA, T.; SANTOS FILHO, L. **Plantas Raras do Cerrado**, p . 320, 2014.

MCCARTHY, K. P.; FLETCHER JR, R. J.; ROTA, C. T.; HUTTO, R. L. Predicting Species Distributions from Samples Collected along Roadsides. **Conservation Biology**, v. 26, n. 1, p. 68–77, 2012.

MCGARIGAL, K.; CUSHMAN, S. A.; ENE, E. **FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps**. Amherst: University of Massachusetts, 2012.

MCPHERSON, J.; JETZ, W.; ROGERS, D. J. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? **Journal of applied ecology**, v. 41, n. 5, p. 811–823, 2004.

MELEČKOVÁ, Z.; DÍTĚ, D.; JUN, P. E.; PÍŠ, V.; GALVÁNEK, D. Succession of Saline Vegetation in Slovakia after a Large-Scale Disturbance. **Annales Botanici Fennici**, v. 51, n. 5, p. 285–296, 2014.

MENDES, P.; DE MARCO, P. Bat species vulnerability in Cerrado: integrating climatic suitability with sensitivity to land-use changes. **Environmental Conservation**, p. 1–8, 2017.

MENDONÇA, R. C. DE, FELFILI, J. M.; WALTER, B. M. T.; JÚNIOR, M. C. S.; REZENDE, A. V.; FILGUEIRAS, T. S.; NOGUEIRA, P. E.; FAGG, C. W. Flora vascular do bioma Cerrado: checklist com 12.356 espécies. In: SANO, S. M.; ALMEIDA, S. P. DE; RIBEIRO, J. F. (Eds.). . **Cerrado: ecologia e flora**. [s.l.] Embrapa Cerrados/Embrapa Informação Tecnológica, 2008. v. 1, p. 422–442.

MERELES, M. F. Acerca de las extensiones de Cerrados y Cerradones en el Paraguay. **Paraquaria Natural**, v. 1, n. 1, p. 35–38, 2013.

MERELES, M. F. Aportes al conocimiento de la flora y las comunidades vegetales en la cuenca del lago Ypoá, región Oriental, Paraguay. **Miscelánea**, v. 14, p. 160, 2005.

MERELES, M. F.; CARTES, J. L.; CLAY, R. P.; CACCIALI, P.; PARADEDA, C.; RODAS, O.; YANOSKY, A. Análisis cualitativo para la definición de las ecorregiones de Paraguay occidental. **Paraquaria Natural**, v. 1, n. 2, p. 12–20, 2013.

MEYER, C.; KREFT, H.; GURALNICK, R.; JETZ, W. Global priorities for an effective information basis of biodiversity distributions. **Nature Communications**, v. 6, p. 8221, 2015.

MOD, H. K.; SCHERRER, D.; LUOTO, M.; GUISAN, A. 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.

MOILANEN, A.; ANDERSON, B. J.; ARPONEN, A.; POUZOLS, F. M.; THOMAS, C. D., A. et al. Edge artefacts and lost performance in national versus continental conservation priority areas. **Diversity and Distributions**, v. 19, n. 2, p. 171–183, 2013.

MOILANEN, A.; KUJALA, H.; LEATHWICK, J. R. The Zonation framework and What we use is not what we know: environmental predictors in plant distribution models software for conservation prioritization. **Spatial conservation prioritization**, p. 196–210, 2009.

MONZÓN, J.; MOYER-HORNER, L.; PALAMAR, M. B. Climate Change and Species Range Dynamics in Protected Areas. **BioScience**, v. 61, n. 10, p. 752–761, 2011.

MUBOKO, N. The role of transfrontier conservation areas and their institutional framework in natural resource-based conflict management: A review. **Journal of Sustainable Forestry**, v. 36, n. 6, p. 583–603, 2017.

MUSCARELLA, R.; GALANTE, P. J.; SOLEY-GUARDIA, M.; BORIA, R. A.; KASS, J. M.; URIARTE, M.; ANDERSON, R. P. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. **Methods in Ecology and Evolution**, v. 5, n. 11, p. 1198–1205, 2014.

MUYLAERT, R. L.; STEVENS, R. D.; RIBEIRO, M. C. Threshold effect of habitat loss on bat richness in cerrado-forest landscapes. **Ecological Applications**, v. 26, n. 6, p. 1854–1867, 2016.

MWANJALOLO JACKSON-GILBERT, M. et al. Soil Fertility in relation to Landscape Position and Land Use/Cover Types: A Case Study of the Lake Kivu Pilot Learning Site. **Advances in Agriculture**, v. 2015, p. 1–8, 2015.

NAVARRO, G.; MOLINA, J. A.; DE MOLAS, L. P. Classification of the forests of the northern Paraguayan Chaco. **Phytocoenologia**, v. 36, n. 4, p. 473–508, 2006.

NAVARRO, G.; MOLINA, J. A.; VEGA, S. Soil factors determining the change in forests between dry and wet Chacos. **Flora - Morphology, Distribution, Functional Ecology of Plants**, v. 206, n. 2, p. 136–143, 2011.

NEWBOLD, T. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. **Progress in Physical Geography**, v. 34, n. 1, p. 3–22, 2010.

NEWBOLD, T.; HUDSON, L. N.; ARNELL, A. P.; CONTU, S.; PALMA, A. D.; FERRIER, S.; HILL, S. L. L.; HOSKINS, A. J.; LYSENKO, I.; PHILLIPS, H. R. P.; BURTON, V. J.; CHNG, C. W. T.; EMERSON, S.; GAO, D.; PASK-HALE, G.; HUTTON, J.; JUNG, M.; SANCHEZ-ORTIZ, K.; SIMMONS, B. I.; WHITMEE, S.; ZHANG, H.; SCHARLEMANN, J. P. W.; PURVIS, A. Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. **Science**, v. 353, n. 6296, p. 286–291, 2016.

NEWBOLD, T.; HUDSON, L. N.; HILL, S. L. L.; CONTU, S.; LYSENKO, I.; SENIOR, R. A.; BÖRGER, L.; BENNETT, D. J.; CHOIMES, A.; COLLEN, B.; DAY, J.; DE PALMA, A.; DÍAZ, S.; ECHEVERRIA-LONDOÑO, S.; EDGAR, M. J.; FELDMAN, A.; GARON, M.; HARRISON, M. L. K.; ALHUSSEINI, T.; INGRAM, D. J.; ITESCU, Y.; KATTGE, J.; KEMP, V.; KIRKPATRICK, L.; KLEYER, M.; CORREIA, D. L. P.; MARTIN, C. D.; MEIRI, S.; NOVOSOLOV, M.; PAN, Y.; PHILLIPS, H. R. P.; PURVES, D. W.; ROBINSON, A.; SIMPSON, J.; TUCK, S. L.; WEIHER, E.; WHITE, H. J.; EWERS, R. M.; MACE, G. M.; SCHARLEMANN, J. P. W.; PURVIS, A. Global effects of land use on local terrestrial biodiversity. **Nature**, v. 520, n. 7545, p. 45–50, 2015.

OLIVEIRA, U.; SOARES-FILHO, B. S.; PAGLIA, A. P.; BRESCOVIT, A. D.; DE CARVALHO, C. J. B.; SILVA, D.P.; REZENDE, D.T.; LEITE, F.S.F.; BATISTA, J. A. N.; BARBOSA, J. P. P. P.; STEHMANN, J. R.; ASCHER, J. S.; DE VASCONCELOS, M. F.; DE MARCO, P.; LÖWENBERG-NETO, P.; FERRO, V. G.; SANTOS, A. J. Biodiversity conservation gaps in the Brazilian protected areas. **Scientific Reports**, v. 7, n. 1, 2017.

OLIVER, T. H.; MORECROFT, M. D. Interactions between climate change and land use change on biodiversity: attribution problems, risks, and opportunities: Interactions between climate change and land use change. **Wiley Interdisciplinary Reviews: Climate Change**, v. 5, n. 3, p. 317–335, 2014.

OLSON, D. M.; DINERSTEIN, E.; WIKRAMANAYAKE, E. D.; BURGESS, N.D.; POWELL, G. V.; UNDERWOOD, E. C.; D'AMICO, J. A.; ITOUA, I.; STRAND, H.E.; MORRISON, J. C. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. **BioScience**, v. 51, n. 11, p. 933–938, 2001.

OVERBECK, G. E.; VÉLEZ-MARTIN, E.; SCARANO, F. R.; LEWINSOHN, T. M.; FONSECA, C. R.; MEYER, S. T.; MÜLLER, S. C.; CEOTTO, P.; DADALT, L.; DURIGAN, G.; GANADE, G.; GOSSNER, M. M.; GUADAGNIN, D. L.; LORENZEN, K.; JACOBI, C. M.; WEISSER, W. W.; PILLAR, V. D. Conservation in Brazil needs to include non-forest ecosystems. **Diversity and Distributions**, v. 21, n. 12, p. 1455–1460, 2015.

PAYNE, B. L.; BRO-JØRGENSEN, J. A framework for prioritizing conservation translocations to mimic natural ecological processes under climate change: A case study with African antelopes. **Biological Conservation**, v. 201, p. 230–236, 2016.

PEARSON, R. G. Reasons to Conserve Nature. **Trends in Ecology & Evolution**, v. 31, n. 5, p. 366–371, 2016.

PEARSON, R. G.; STANTON, J. C.; SHOEMAKER, K. T.; AIELLO-LAMMENS, M. E.; ERSTS, P. J.; HORNING, N.; FORDHAM, D. A.; RAXWORTHY, C. J.; RYU, H. Y.; MCNEES, J.; AKÇAKAYA, H. R. Life history and spatial traits predict extinction risk due to climate change. **Nature Climate Change**, v. 4, n. 3, p. 217–221, 2014.

PECL, G. T.; ARAÚJO, M. B.; BELL, J. D.; BLANCHARD, J.; BONEBRAKE, T. C.; CHEN, I.-C.; CLARK, T. D.; COLWELL, R. K.; DANIELSEN, F.; EVENGÅRD, B.; FALCONI, L.; FERRIER, S.; FRUSHER, S.; GARCIA, R. A.; GRIFFIS, R. B.; HOBDAY, A. J.; JANION-SCHEEPERS, C.; JARZYNA, M. A.; JENNINGS, S.; LENOIR, J.; LINNETVED, H. I.; MARTIN, V. Y.; MCCORMACK, P. C.; MCDONALD, J.; MITCHELL, N. J.; MUSTONEN, T.; PANDOLFI, J. M.; PETTORELLI, N.; POPOVA, E.; ROBINSON, S. A.; SCHEFFERS, B. R.; SHAW, J. D.; SORTE, C. J. B.; STRUGNELL, J. M.; SUNDAY, J. M.; TUANMU, M.-N.; VERGÉS, A.; VILLANUEVA, C.; WERNBERG, T.; WAPSTRA, E.; WILLIAMS, S. E. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. **Science**, v. 355, n. 6332, p. eaai9214, 2017.

PETER FILZMOSER; FRITZ, H.; KALCHER, K. **pcaPP: Robust PCA by Projection Pursuit. R package version 1.9-60**, 2014. Disponível em: <https://CRAN.R-project.org/package=pcaPP>

PETERSON, A. T.; NAVARRO-SIGÜENZA, A. G.; GORDILLO, A. Assumption-versus data-based approaches to summarizing species’ ranges: Summarizing Species’ Ranges. **Conservation Biology**, nov. 2016.

PETERSON, A. T.; SOBERÓN, J. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. **Natureza & Conservação**, v. 10, n. 2, p. 102–107, 2012.

PETERSON, A. T.; SOBERÓN, J.; PEARSON, R. G.; ANDERSON, R. P.; MARTÍNEZ-MEYER, E.; NAKAMURA, M.; ARAÚJO, M. B. **Ecological niches and geographic distributions**. Princeton, N.J: Princeton University Press, 2011.

PHILLIPS, S. J.; ANDERSON, R. P.; DUDÍK, M.; SCHAPIRE, R. E.; BLAIR, M. E. Opening the black box: an open-source release of Maxent. **Ecography**, v. 40, p. 1–7, 2017.

PHILLIPS, S. J.; ANDERSON, R. P.; SCHAPIRE, R. E. Maximum entropy modeling of species geographic distributions. **Ecological Modelling**, v. 190, n. 3–4, p. 231–259, 2006.

PHILLIPS, S. J.; DUDÍK, M.; SCHAPIRE, R. E. **A maximum entropy approach to species distribution modeling**. Proceedings of the twenty-first international conference on Machine learning. **Anais**.ACM, 2004

PINEDA, E.; LOBO, J. M. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. **Journal of Animal Ecology**, v. 78, n. 1, p. 182–190, 2009.

PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D.; R CORE TEAM. **nlme: Linear and Non linear Mixed Effects Models. R package version 3.1-125**.

PLISCOFF, P. et al. Effects of alternative sets of climatic predictors on species distribution models and associated estimates of extinction risk: A test with plants in an arid environment. **Ecological Modelling**, v. 288, p. 166–177. 2014.

POTT, A.; OLIVEIRA, A. K. M.; DAMASCENO-JUNIOR, G. A.; SILVA, J. S. V. Plant diversity of the Pantanal wetland. **Brazilian Journal of Biology**, v. 71, n. 1, p. 265–273, 2011.

POUZOLS, F. M.; TOIVONEN, T.; DI MININ, E.; KUKKALA, A. S.; KULLBERG, P.; KUUSTERÄ, J.; LEHTOMÄKI, J.; TENKANEN, H.; VERBURG, P. H.; MOILANEN, A. Global protected area expansion is compromised by projected land-use and parochialism. **Nature**, v. 516, n. 7531, p. 383–386, 2014.

PRANCE, G. T.; SCHALLER, G. B. Preliminary study of some vegetation types of the Pantanal, Mato Grosso, Brazil. **Brittonia**, v. 34, n. 2, p. 228–251, 1982.

PRITI, H.; ARAVIND, N.A.; UMA SHAANKER, R.; RAVIKANTH, G. Modeling impacts of future climate on the distribution of Myristicaceae species in the Western Ghats, India. **Ecological Engineering**, v. 89, p. 14–23, 2016.

PUCKO, C.; BECKAGE, B.; PERKINS, T.; KEETON, W. S. Species shifts in response to climate change: Individual or shared responses? **The Journal of the Torrey Botanical Society**, v. 138, n. 2, p. 156–176, 2011.

QIAO, H.; SOBERÓN, J.; PETERSON, A. T. No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. **Methods in Ecology and Evolution**, v. 6, n. 10, p. 1126–1136, 2015.

R CORE TEAM. **R: A language and environment for statistical computing. R Foundation for Statistical Computing**.

RABINOWITZ, D. Seven from of rarity. In: SYNGE, J. (Ed.). . **The biological aspects of rare plant conservation**. Chichester, UK: Wiley, 1981.

RACHWAL, M. F. G.; CURCIO, G. R. Atributos pedológicos e ocorrência de caixeta no litoral paranaense, Brasil. **Scientia Forestalis**, v. 59, p. 156–163, 2001.

RANGEL, T. F.; LOYOLA, R. D. Labeling Ecological Niche Models. **Natureza & Conservação**, v. 10, n. 2, p. 119–126, 2012.

RATTER, J. A.; BRIDGEWATER, S.; RIBEIRO, J. F. Analysis of the floristic composition of the Brazilian cerrado vegetation III: comparison of the woody vegetation of 376 areas. **Edinburgh journal of botany**, v. 60, n. 01, p. 57–109, 2003.

RATTER, J. A.; RIBEIRO, J. F.; BRIDGEWATER, S. The Brazilian cerrado vegetation and threats to its biodiversity. **Annals of botany**, v. 80, n. 3, p. 223–230, 1997.

REDDY, S.; DÁVALOS, L. M. Geographical sampling bias and its implications for conservation priorities in Africa. **Journal of Biogeography**, v. 30, n. 11, p. 1719–1727, 2003.

REDO, D. J.; AIDE, T. M.; CLARK, M. L. The Relative Importance of Socioeconomic and Environmental Variables in Explaining Land Change in Bolivia, 2001–2010. **Annals of the Association of American Geographers**, v. 102, n. 4, p. 778–807, 2012.

REEF, R.; FELLER, I. C.; LOVELOCK, C. E. Nutrition of mangroves. **Tree Physiology**, v. 30, n. 9, p. 1148–1160, 2010.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma Cerrado. In: SANO, S. M.; ALMEIDA, S. P. DE; RIBEIRO, J. F. (Eds.). . **Cerrado: ecologia e flora**. Brasília, DF: Embrapa Cerrados, 2008. v. 1p. 151–212.

RIDGEWAY, G. **gbm: Generalized Boosted Regression Models. R package version 2.1.1**.

ROBERTS, D. R.; BAHN, V.; CIUTI, S.; BOYCE, M. S.; ELITH, J.; GUILLERA-ARROITA, G.; HAUENSTEIN, S.; LAHOZ-MONFORT, J. J.; SCHRÖDER, B.; THUILLER, W.; WARTON, D. I.; WINTLE, B. A.; HARTIG, F.; DORMANN, C. F. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. **Ecography**, 2017.

RODRIGUES, A. S.; GASTON, K. J. Rarity and conservation planning across geopolitical units. **Conservation Biology**, v. 16, n. 3, p. 674–682, 2002.

RUGGIERO, P. G. C.; BATALHA, M. A.; PIVELLO, V. R.; MEIRELLES, S. T. Soil-vegetation relationships in cerrado (Brazilian savanna) and semideciduous forest, Southeastern Brazil. **Plant Ecology**, v. 160, n. 1, p. 1–16, 2002.

SALAZAR, A.; BALDI, G.; HIROTA, M.; SYKTUS, J.; MCALPINE, C. Land use and land cover change impacts on the regional climate of non-Amazonian South America: A review. **Global and Planetary Change**, v. 128, p. 103–119, 2015.

SALCEDO-SANZ, S.; ROJO-ÁLVAREZ, J. L.; MARTÍNEZ-RAMÓN, M.; CAMPS-VALLS, G. Support vector machines in engineering: an overview. **Wiley**

**Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 4, n. 3, p. 234–267, 2014.

SÁNCHEZ-FERNÁNDEZ, D.; LOBO, J. M.; HERNÁNDEZ-MANRIQUE, O. L. Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. **Diversity and Distributions**, v. 17, n. 1, p. 163–171, 2010.

SAUPE, E. E.; BARVE, V.; MYERS, C. E.; SOBERÓN, J.; BARVE, N.; HENSZ, C. M.; PETERSON, A. T.; OWENS, H. L.; LIRA-NORIEGA, A. Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. **Ecological Modelling**, v. 237–238, p. 11–22, 2012.

SCHAETZL, R. J.; ANDERSON, S. **Soils genesis and geomorphology**. Cambridge; New York: Cambridge University Press, 2005.

SCHMIDT, B. C.; ROLAND, J. Moth Diversity in a Fragmented Habitat: Importance of Functional Groups and Landscape Scale in the Boreal Forest. **Annals of the Entomological Society of America**, v. 99, n. 6, p. 1110–1120, 2006.

SEAM. **Quinto informe nacional al convenio de diversidad biológica**. Asunción , Paraguay: Dirección general de protecciòny conservación de la biodiversidad secretaria del ambiente, 2016.

SHACKELFORD, G. E.; STEWARD, P. R.; GERMAN, R. N.; SAIT, S. M.; BENTON, T. G. Conservation planning in agricultural landscapes: hotspots of conflict between agriculture and nature. **Diversity and Distributions**, v. 21, n. 3, p. 357–367, 2015.

SILVA, J. M. C.; BATES, J. M. Biogeographic Patterns and Conservation in the South American Cerrado: A Tropical Savanna Hotspot. **BioScience**, v. 52, n. 3, p. 225, 2002.

SILVA, M. P.; MAURO, R.; MOURAO, G., COUTINHO, M. Distribuição e quantificação de classes de vegetação do Pantanal através de levantamento aéreo. **Revista brasileira de Botânica**, v. 23, n. 2, p. 143–152, 2000.

SIMON, L. M.; OLIVEIRA, G. DE, BARRETO, B. DE S.; NABOUT, J. C.; RANGEL, T. F. L. V. B.; DINIZ-FILHO, J. A. F. Effects of global climate changes on geographical distribution patterns of economically important plant species in cerrado. **Revista Árvore**, v. 37, n. 2, p. 267–274, 2013.

SINTHUMULE, N. I. Multiple-land use practices in transfrontier conservation areas: the case of Greater Mapungubwe straddling parts of Botswana, South Africa and Zimbabwe. **Bulletin of Geography. Socio-economic Series**, v. 34, n. 34, 2016.

SIQUEIRA, M. F. DE; PETERSON, A. T. Consequences of global climate change for geographic distributions of cerrado tree species. **Biota Neotropica**, v. 3, n. 2, p. 1–14, 2003.

SMITH, A. B.; LONG, Q. G.; ALBRECHT, M. A. Shifting targets: spatial priorities for ex situ plant conservation depend on interactions between current threats, climate change, and uncertainty. **Biodiversity and Conservation**, v. 25, n. 5, p. 905–922, 2016.

SOBERÓN, J. Grinnellian and Eltonian niches and geographic distributions of species. **Ecology Letters**, v. 10, n. 12, p. 1115–1123, 2007.

SOBERÓN, J.; PETERSON, A. T. Interpretation of models of fundamental ecological niches and species' distributional areas. **Biodiversity Informatics**, v. 2, p. 1–10, 2005.

SOUSA-SILVA, R.; ALVES, P.; HONRADO, J.; LOMBA, A. Improving the assessment and reporting on rare and endangered species through species distribution models. **Global Ecology and Conservation**, v. 2, p. 226–237, 2014.

SOUZA, H. A. V. E.; COLLEVATTI, R. G.; LIMA-RIBEIRO, M. S.; LEMOS-FILHO, J. P.; LOVATO, M. B. A large historical refugium explains spatial patterns of genetic diversity in a Neotropical savanna tree species. **Annals of Botany**, v. 119, n. 2, p. 239–252, 2017.

SOUZA, R. A.; DE MARCO, P. The use of species distribution models to predict the spatial distribution of deforestation in the western Brazilian Amazon. **Ecological Modelling**, v. 291, p. 250–259. 2014.

SPICHIGER, R.; CALENGE, C.; BISE, B. Geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin. **Journal of Biogeography**, v. 31, n. 9, p. 1489–1501, 2004.

STILL, S. M.; FRANCES, A. L.; TREHER, A. C.; OLIVER, L. Using Two Climate Change Vulnerability Assessment Methods to Prioritize and Manage Rare Plants: A Case Study. **Natural Areas Journal**, v. 35, n. 1, p. 106–121, 2015.

STRASSBURG, B. B. N.; BROOKS, T.; FELTRAN-BARBIERI, R.; IRIBARREM, A.; CROUZEILLES, R.; LOYOLA, R.; LATAWIEC, A. E.; OLIVEIRA FILHO, F. J. B.; SCARAMUZZA, C. A. M.; SCARANO, F. R.; SOARES-FILHO, B.; BALMFORD, A. Moment of truth for the Cerrado hotspot. **Nature Ecology & Evolution**, v. 1, n. 4, p. 0099, 2017.

SVENNING, J. C.; FLØJGAARD, C.; MARSKE, K. A.; NÓGUES-BRAVO, D.; NORMAND, S. Applications of species distribution modeling to paleobiology. **Quaternary Science Reviews**, v. 30, n. 21–22, p. 2930–2947, 2011.

SWIFT, T. L.; HANNON, S. J. Critical thresholds associated with habitat loss: a review of the concepts, evidence, and applications. **Biological Reviews**, v. 85, n. 1, p. 35–53, 2010.

SYPHARD, A. D.; FRANKLIN, J. Species traits affect the performance of species distribution models for plants in southern California. **Journal of Vegetation Science**, v. 21, n. 1, p. 177–189, 2010.

TEILLIER, S.; BECERRA, P. Flora y vegetacion del salar de Ascotan, andes del norte de chile. **Gayana. Botánica**, v. 60, n. 2, p. 114–122, 2003.

TEJADA, G.; DALLA-NORA, E.; CORDOBA, D.; LAFORTEZZA, R.; OVANDO, A.; ASSIS, T.; AGUIAR, A.P. Deforestation scenarios for the Bolivian lowlands. **Environmental Research**, v. 144, p. 49–63, 2016.

TER STEEGE, H. et al. Estimating the global conservation status of more than 15,000 Amazonian tree species. **Science Advances**, v. 1, n. 10, p. e1500936–e1500936, 2015.

THOMAS, C. D. Climate, climate change and range boundaries: Climate and range boundaries. **Diversity and Distributions**, v. 16, n. 3, p. 488–495, 2010.

THUILLER, W. On the importance of edaphic variables to predict plant species distributions - limits and prospects. **Journal of Vegetation Science**, v. 24, n. 4, p. 591–592, 2013.

TSOAR, A.; ALLOUCHE, O.; STEINITZ, O.; ROTEM, D.; KADMON, R. A comparative evaluation of presence-only methods for modelling species distribution. **Diversity and Distributions**, v. 13, n. 4, p. 397–405, 2007.

UNEP. **The Emissions Gap Report 2017: a UN Environment Synthesis report**. Nairobi: United Nations Environment Programme UNEP, 2017.

UNITED NATIONS. **Transforming our world: the 2030 Agenda for Sustain able Development, General Assembly A/RES/70/1**. [s.l: s.n.].

VÁCLAVÍK, T.; MEENTEMEYER, R. K. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? **Ecological Modelling**, v. 220, n. 23, p. 3248–3258, 2009.

VALE, C. G.; TARROSO, P.; BRITO, J. C. Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara-Sahel transition zone. **Diversity and Distributions**, v. 20, n. 1, p. 20–33, 2014.

VALLEJOS, M.; VOLANTE, J. N.; MOSCIARO, M. J.; VALE, L. M.; BUSTAMANTE, M. L.; PARUELO, J. M. Transformation dynamics of the natural cover in the Dry Chaco ecoregion: A plot level geo-database from 1976 to 2012. **Journal of Arid Environments**, v. 123, p. 3–11, 2015.

VAN VUUREN, D. P.; EDMONDS, J.; KAINUMA, M.; RIAHI, K.; THOMSON, A.; HIBBARD, K.; HURTT, G. C.; KRAM, T.; KREY, V.; LAMARQUE, J. F.; MASUI, T.; MEINSHAUSEN, M.; NAKICENOVIC, N.; SMITH, S. J.; ROSE, S. K. The representative concentration pathways: an overview. **Climatic Change**, v. 109, n. 1–2, p. 5–31, nov. 2011.

VANDERWAL, J. **SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. R package version 1.1-221.**, 2014. Disponível em: <https://CRAN.Rproject.org/package=SDMTools>

VELAZCO, S.J.E.; GALVÃO, F.; VILLALOBOS, F.; DE MARCO JR, P. Using worldwide edaphic data to model plant species niches: An assessment at a continental extent. **PloS one**, v. 12, n. 10, p. e0186025, 2017.

VIEIRA, R. R. S.; RIBEIRO, B. R.; RESENDE, F. M.; BRUM, F. T.; MACHADO, N.; SALES, L. P.; MACEDO, L.; SOARES-FILHO, B.; LOYOLA, R. Compliance to Brazil's Forest Code will not protect biodiversity and ecosystem services. **Diversity and Distributions**, 2017.

VILELA, A. A.; DEL CLARO, V. T. S.; TOREZAN-SILINGARDI, H. M.; DEL-CLARO, K. Climate changes affecting biotic interactions, phenology, and reproductive success in a savanna community over a 10-year period. **Arthropod-Plant Interactions**, 2017.

VILLARROEL, D. **Campos e savanas do subandino boliviano: flora, diversidade e fitogeografia**. Brasília, DF: Universidade de Brasilia, 2017.

VILLARROEL, D.; MUNHOZ, C. B. R.; PROENÇA, C. E. B. Campos y Sabanas del Cerrado en Bolivia: Delimitación, Síntesis Terminológica y sus Características Fisionómicas. **Kempffiana**, v. 12, n. 1, p. 47–80, 2016.

VILLÉN-PÉREZ, S.; MENDES, P.; NÓBREGA, C.; CÓRTES, L. G.; DE MARCO, P. Mining code changes undermine biodiversity conservation in Brazil. **Environmental Conservation**, p. 1–4, 2017.

WAN, J.; WANG, C.; YU, J.; NIE, S.; HAN, S.; LIU, J.; ZU, Y.; WANG, Q. Developing conservation strategies for Pinus koraiensis and Eleutherococcus senticosus by using model-based geographic distributions. **Journal of Forestry Research**, v. 27, n. 2, p. 389–400, 2016.

WAROUX, Y. LE P. DE, GARRETT, R. D.; HEILMAYR, R.; LAMBIN, E. F. Land-use policies and corporate investments in agriculture in the Gran Chaco and Chiquitano. **Proceedings of the National Academy of Sciences**, v. 113, n. 15, p. 4021–4026, 2016.

Wells, J. V.; Robertson, B.; Rosenberg, K. V.; Mehlman, D. W. Global versus Local Conservation Focus of U.S. State Agency Endangered Bird Species Lists. **PLoS ONE**, v. 5, n. 1, p. e8608, 2010.

WERNECK, F. P.; NOGUEIRA, C.; COLLI, G. R.; SITES, J. W.; COSTA, G. C. Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot: Climatic stability and biodiversity in the Cerrado. **Journal of Biogeography**, v. 39, n. 9, p. 1695–1706, 2012.

WILLIAMS, J. N.; SEO, C.; THORNE, J.; NELSON, J.K.; ERWIN, S.; O'BRIEN, J. M.; SCHWARTZ, M. W. Using species distribution models to predict new occurrences for rare plants. **Diversity and Distributions**, v. 15, n. 4, p. 565–576. 2009.

YAHDJIAN, L.; SALA, O. E. Climate change impacts on South American rangelands. **Rangelands**, v. 30, n. 3, p. 34–39, 2008.

ZHU, G.-P.; PETERSON, A. T. Do consensus models outperform individual models? Transferability evaluations of diverse modeling approaches for an invasive moth. **Biological Invasions**, v. 19, n. 9, p. 2519–2532, 2017.

ZHU, K.; WOODALL, C. W.; CLARK, J. S. Failure to migrate: lack of tree range expansion in response to climate change. **Global Change Biology**, v. 18, n. 3, p. 1042–1052, 2012.