



PAOLO RAMONI PERAZZI

**CLIMATE MODELS, NICHE MODELS AND THE
CONSERVATION OF OWLS IN BRAZIL**

**LAVRAS-MG
2022**

PAOLO RAMONI PERAZZI

**CLIMATE MODELS, NICHE MODELS AND THE CONSERVATION OF OWLS IN
BRAZIL**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ecologia Aplicada, área de concentração em Ecologia e Conservação de Recursos Naturais em Ecossistemas Fragmentados e Agrossistemas, para a obtenção do título de Doutor.

Prof. Dr. Marcelo Passamani
Orientador

Prof. Dr. Dirk Thielen
Coorientador

**LAVRAS–MG
2022**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Ramoni Perazzi, Paolo.

Climate models, niche models and the conservation of owls in Brazil /
Paolo Ramoni Perazzi. - 2022.
108 p. : il.

Orientador(a): Marcelo Passamani.

Coorientador(a): Dirk Thielen.

Tese (doutorado) - Universidade Federal de Lavras, 2022.

Bibliografia.

1. Strigidae. 2. BrazilClim. 3. Modelagem de Distribuição de
Espécies. I. Passamani, Marcelo. II. Thielen, Dirk. III. Título.

PAOLO RAMONI PERAZZI

**CLIMATE MODELS, NICHE MODELS AND THE CONSERVATION OF OWLS IN
BRAZIL**

**MODELOS CLIMATICOS, MODELOS DE NICHO E CONSERVAÇÃO DE
CORUJAS NO BRASIL**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ecologia Aplicada, área de concentração em Ecologia e Conservação de Recursos Naturais em Ecossistemas Fragmentados e Agrossistemas, para a obtenção do título de Doutor.

APROVADA em 11 de fevereiro de 2022.

Dr. Rafael Dudeque Zenni, UFLA

Dr. Ricardo Bomfim Machado, UNB

Dr. Lucas del Bianco Faria, UFLA

Dr. Thadeu Sobral de Souza, UFMT

Dr. Júlio Neil Cassa Louzada, UFLA

Prof. Dr. Marcelo Passamani
Orientador

Prof. Dr. Dirk Thielen
Coorientador

**LAVRAS–MG
2022**



*Für Heli
(und Alle)*

AGRADECIMENTOS

A Deus pela inacreditável sequência de fatos que permitiu a nossa saída da Venezuela e chegar para bom porto no Brasil: "Deem graças ao Senhor, clamem pelo seu nome, divulguem entre as nações o que ele tem feito". Crônicas 16:8-11

A mi esposa, Irma Alejandra Soto Werschitz y a mi hija, Heliangelus Ramoni Soto, por todo el apoyo, paciencia y amor. Esto se pudo hacer gracias a ustedes.

Ai me genidur (*in memoriam*), Edda Perazzi e Gian Franco Ramoni, g'a saress piasu' un sacc che m'avresan vist riva'a fin chi.

Alle mie sorelle Anna, Yudith e Pepita.

A mi suegra Irma Werschitz (*in memoriam*) y a mis cuñados Laritza, Jaime, César e Ivan por toda su enorme ayuda y apoyo, especialmente en los momentos más complicados.

Ao meu orientador Prof. Dr. Marcelo Passamani pela oportunidade de prosseguir meus estudos no Laboratório de Ecologia e Conservação de Mamíferos, e por todo o seu apoio inclusive fora do âmbito acadêmico.

A mi querido amigo y colega Dr. Dirk Thielen y a su hermosa familia, cotutor de esta tesis, por toda su ayuda y contribuciones, desde las que nos permitieron salir de Venezuela hasta todas las que ayudaron a mejorar este trabajo.

A mis amigos de toda la vida, Diego Bratta, Guillermo Bianchi Pérez, Lucia Pérez (*in memoriam*), Guillermo Bianchi Ballesteros, María Andreína Pacheco, Mariana Muñoz Romo, Suzana Cubillos, Mayerlin Uzcátegui, Kay Tucci (¡gracias por la maleta!), Carlos Echeverria, Magdiel Ablan, Giorgio Tonella, Irma Tonel, Andrea Cruz y Jorge Montejo por todo su cariño y apoyo desde la distancia. Con mucho aprecio y respeto recuerdo a los profesores Diego Dávila y Herbert Hoeger.

Aos meus novos amigos Luis Depablos, Gioconda Laurenti e Diego Nascimento, além do Hélio Parima, Olga Lucia Mondragón, a Regina Lewandowski, o Miguel Angel Díaz Herrera e suas respectivas famílias que fizeram este exílio muito mais grato e foram de grande apoio.

Aos meus coautores e amigos, o Dr. Carlos Padovani (Embrapa), o Prof. Dr. Marco Aurelio Arizapana Almonacid e Luis Fernando Landeo Atmi (Universidad Autónoma de Huanta, Peru) cujo apoio foi determinante para poder fazer este trabalho.

Ao Instituto Presbiteriano Gammon, otimamente dirigido pelo Dr. Allyson Massote Carvalho, pela bolsa de estudo para minha filha. Definitivamente: "a gente sai do Gammon, mas o Gammon não sai da gente".

Às Meninas Cantoras de Lavras, otimamente dirigidas pelo maestro José Maciel, pela acolhida para minha filha.

Aos colegas do LECOM, especialmente à Rayssa Pedroso, Luciano Querido e Hugo Mabilana. Também aos meus colegas da pós, especialmente à Marina Rufino, Raphaela Duarte, Alexandre Peressin, Tássia Chagas, Olívia Carolina Tonussi, Andressa Sene e Kassius Santos.

À Ellen Cristina de Carvalho pelo seu carinho, apoio e paciência!

As outras pessoas que foram de grande ajuda, entre as quais lembro agora aos Professores Antonio Chalfun-Junior, Kátia Poles e Norma Joseph, além dos Senhores Santiago Puentes, Carlos Torres, Máximo Ochoa, Maurício Mansur, Marco Talento.

À Universidade Federal de Lavras, ao Setor de Conservação e ao Programa de Pós-Graduação em Ecologia Aplicada pelo apoio e aprendizado. Ao Programa de Alianças para Educação e Capacitação (PAEC-PG), estabelecido entre a Organização dos Estados Americanos (OEA) e o Grupo Coimbra de Universidades Brasileiras (GCUB), pela bolsa de doutorado.

*"I wish it need not have happened in my time," said Frodo.
"So do I," said Gandalf, "and so do all who live to see such
times. But that is not for them to decide. All we have to
decide is what to do with the time that is given us."*

Tolkien, The Fellowship of the Ring

GENERAL ABSTRACT

The effective conservation and management of species or groups of them requires basic knowledge about their distribution, which is problematic in the case of highly mobile, ecologically diverse, little studied and nocturnal species such as owls (Strigidae). In addition, producing detailed distributions to megabiodiverse and continental-sized countries such as Brazil is difficult, especially when the base information is dispersed and non-standardized. Using Species Distribution Modeling (SDM) based on a maximum entropy approach, we evaluated the potential distribution of 21 species and 21 subspecies of owls recorded in Brazil. For this, we first evaluated the information on minimum and maximum temperatures, in addition to precipitation, provided by the Brazilian network of meteorological stations, comparing conventional and automatic stations. We found that both data sets have problems, but the evidence evaluated suggests that conventional stations provide slightly more reliable precipitation data, while automatic gauges are more consistent with regards to temperature information. Second, this information was used to evaluate the performance of the bioclimatic bases currently available. We found a better correspondence between the climate data measured in the field and the information provided by the Tropical Rainfall Measuring Mission (TRMM 3B43 v7), in the case of precipitation, and the surfaces provided by the National Oceanic and Atmospheric Administration (NOAA), in the case of temperatures, sources not traditionally used for SDMs. We gauge-calibrated these surfaces using the climatic information obtained in the field, through machine-learning algorithms (gradient boosting and random forest), and used these improved surfaces to create a base of bioclimatic variables we called BrazilClim. Third, we collected and filtered occurrence data for the Strigidae in Brazil, and generated particular SDMs for each species and subspecies, evaluating the similarity of niches between conspecific subspecies. Finally, we created maps of species richness, and contrasted this information with the areas of integral protection in Brazil. With 81% of the Brazilian species recorded, both the Atlantic Forest and the Cerrado have the highest richness, followed by the Amazon (67%), Pampa (62%), Caatinga (57%) and Pantanal (48%). However, the comparison of recorded and predicted richness suggests incomplete inventories, especially in the Caatinga and Pantanal. On the other hand, the subspecies presented marked divergences of niches, suggesting that the species richness of Strigidae is underestimated in Brazil. The Cerrado and the Atlantic Forest are the most threatened biomes, with relatively small and sparse preservation areas. Thus, our study is an urgent call to explore the diversification of owl strains in Brazil in order to improve efforts related to biodiversity conservation.

Keywords: Strigidae. BrazilClim. Species Distribution Modeling. Conservation Units.

RESUMO GERAL

A conservação e o manejo eficazes das espécies ou grupos delas requer conhecimento básico sobre a sua distribuição, o que é problemático no caso de espécies altamente móveis, ecologicamente diversas, pouco estudadas e noturnas como as corujas (Strigidae). Além disso, produzir distribuições detalhadas para países megabiodiversos e de tamanho continental como o Brasil é difícil, especialmente quando a informação base está dispersa e não padronizada. Usando Modelagem de Distribuição de Espécies (SDM no inglês) com base em uma abordagem de entropia máxima, avaliou-se a distribuição potencial de 21 espécies e 21 subespécies de corujas registradas no Brasil. Para isso, primeiro avaliou-se a informação sobre temperaturas mínimas e máximas, além da precipitação, fornecida pela rede de estações meteorológicas brasileiras comparando estações convencionais e automáticas. Encontrou-se que ambos os conjuntos de dados têm problemas, mas a evidência avaliada sugere que as estações convencionais fornecem dados de precipitação um pouco mais confiáveis, enquanto as automáticas são mais consistentes em relação à informação de temperaturas. Segundo, usou-se essa informação para avaliar o desempenho das bases bioclimáticas atualmente disponíveis. Encontrou-se uma melhor correspondência entre os dados climáticos medidos no campo e a informação fornecida pela Tropical Rainfall Measuring Mission (TRMM 3B43 v7) no caso da precipitação, e as superfícies fornecidas pela National Oceanic and Atmospheric Administration (NOAA) no caso das temperaturas, fontes não tradicionalmente usadas para SDMs. Calibraram-se estas superfícies usando a informação climática obtida no campo, usando algoritmos de machine-learning (gradient boosting e random forest), e estas superfícies melhoradas usaram-se para criar uma base de variáveis bioclimáticas que chamamos de BrazilClim. Terceiro, coletaram-se e filtraram-se dados de ocorrências para os Strigidae no Brasil, e geraram-se SDMs particulares para cada espécie e subespécie, avaliando a semelhança dos nichos entre as subespécies coespecíficas. Finalmente, criaram-se mapas de riqueza das espécies, e contrastaram-se essas informações com as áreas de proteção integral no Brasil. Com 81% das espécies brasileiras registradas, tanto a Mata Atlântica quanto o Cerrado têm a maior riqueza, seguidas pela Amazônia (67%), Pampa (62%), Caatinga (57%) e Pantanal (48%). No entanto, a comparação da riqueza registrada e prevista sugere inventários incompletos, especialmente na Pampa. Por outro lado, as subespécies apresentaram marcadas divergências de nichos, sugerindo que a riqueza de espécies de Strigidae está subestimada no Brasil. O Cerrado e a Mata Atlântica são os biomas mais ameaçados, com áreas protegidas relativamente pequenas e esparsas. Assim, nosso estudo constitui um chamado urgente para explorar a diversificação das linhagens de corujas no Brasil afim de melhorar os esforços relacionados à conservação da biodiversidade.

Palavras-chave: Strigidae. BrazilClim. Modelagem de Distribuição de Espécies. Unidades de Conservação.

LISTA DE ILUSTRAÇÕES

- Figura 1.1 – Comparação da quantidade de anos totais operativos, contabilizados dentro do período de 19 anos entre o primeiro de junho de 2000 e 30 de setembro de 2019, para cada mês e para cada tipo de estação no caso de: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima..... 36
- Figura 1.2 – Comparação da quantidade de anos totais efetivos, ou seja, com informação acorde aos critérios da WMO (2017), contabilizados dentro do período de 19 anos entre o primeiro de junho de 2000 e 30 de setembro de 2019, para cada mês e para cada tipo de estação no caso de: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima. 37
- Figura 1.3 – Proporção de gaps na informação climática mensal para cada tipo de estação no caso de: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima. 38
- Figura 1.4 – Comparação da informação fornecida pelas estações convencionais e automáticas homônimas em relação a: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima. Na base de cada gráfico encontra-se o resumo do teste bicaudal pareado *t* de Student: (N.S.) não significativo; (*) $p < 0,05$; (**) $p < 0,01$; (***) $p < 0,001$ 41
- Figura 1.5 – Diferenças resultantes de subtrair a precipitação média mensal registrada por cada estação convencional menos a registrada pela automática homônima. Em azul aqueles casos aonde a estação convencional $>$ estação automática; em vermelho quando estação convencional $<$ estação automática. Os histogramas mostram a frequências dessas diferenças..... 42
- Figura 1.6 – Diferenças resultantes de subtrair a temperatura máxima média mensal registrada por cada estação convencional menos a registrada pela automática homônima. Em azul aqueles casos aonde a estação convencional $>$ estação automática; em vermelho quando estação convencional $<$ estação automática. Os histogramas mostram a frequências dessas diferenças..... 43
- Figura 1.7 – Diferencias resultantes de subtrair a temperatura mínima média mensal registrada por cada estação convencional menos a registrada pela automática homônima. Em cinza escuro aqueles casos aonde a estação convencional $>$ estação automática; em cinza médio quando estação convencional $<$ estação automática. Os histogramas mostram a frequências dessas diferencias. 45
- Figura 1.8 – Comparação entre as estações convencionais e as automáticas homônimas dos valores mensais da raiz média quadrática dos erros relativos (RRMSE pelas suas siglas no inglês), calculados com os valores observados e interpolados para cada estação. (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima. Na base de cada gráfica encontra-se o resumo do teste bicaudal pareado *t* de Student: (N.S.) não significativo; (*) $p < 0,05$; (**) $p < 0,01$; (***) $p < 0,001$. Os eixos das ordenadas estão em escada logarítmica. ... 49

Figure 2.1 – Distribution of (A) climate and (B) biomes in mainland Brazil, according to the Brazilian Institute of Geography and Statistics (IBGE, 2021).	63
Figure 2.2 – Distribution of the meteorological stations providing climate information used in the present study. (A) Mean monthly accumulated precipitation; and (B) Mean maximum and mean minimum temperatures.....	66
Figure 2.3 – Workflow for BrazilClim creation. Evaluation strategies used are listed within dashed boxes. Interpolation (calibration) acronyms: (CK) Cokriging; (GBM) Generalized boosted regression modeling; (LM) Simple regression; (RF) Quantile random forest; and (TPS) Thin plate splines.....	67
Figure 2.4 – Overall (all months) Root Mean Squared Errors contrasting all the databases evaluated in this study (listed in Table 2.1), against BrazilClim (BrCl): our gauge-calibrated dataset. (A) Mean monthly accumulated precipitation; (B) Mean maximum temperature; (C) Mean minimum temperature.....	70
Figure 2.5– Overall (all months) Root Mean Squared Errors calculated comparing different interpolation methods. (A) Mean monthly accumulated precipitation; (B) Mean maximum temperature; (C) Mean minimum temperature; (CK) Cokriging; (GBM) Generalized boosted regression modeling; (LM) Simple regression; (RF) Quantile random forest; and (TPS) Thin plate splines. White boxes represent the results obtained by interpolating the measurements performed on-field (“direct-based”); gray boxes represent the interpolations of the differences resulting by subtracting the amounts measured-on-field minus the value provided in the corresponding conterminous pixels by the selected databases (TRMM in the case of the precipitation and NOAA in the case of the temperatures), and then adding such interpolated surfaces to the corresponding surfaces of the selected databases (“error-based”).....	71
Figure 2.6 – Mean monthly accumulated precipitation surfaces of BrazilClim, resulting from gauge-calibrating information from TRMM 3B43 v7 (Kummerow et al., 2000; TRMM, 2018) through error-based interpolations with generalized boosted regression modeling (GBM).	72
Figure 2.7 – Mean maximum temperatures surfaces of BrazilClim, resulting from gauge-calibrating information from the National Oceanic and Atmospheric Administration (NOAA) through error-based interpolations with generalized boosted regression modeling (GBM).....	74
Figure 2.8 – Mean minimum temperatures surfaces of BrazilClim, resulting from gauge-calibrating information from the National Oceanic and Atmospheric Administration (NOAA) through interpolations with random forest (RF).....	75
Figure 3.1 – Records of Brazilian Strigidæ used in the modeling process. Biome colors correlates with the abundance of records (the darkest the most abundant). State and countries acronyms follow ISO 3166-2 and ISO 3166-1 alpha-2, respectively.	96

Figure 3.2 – Predicted Strigidae species richness patterns in Brazil based on (A) species *sensu stricto*; and (B) *sensu lato* (including subspecies). Dashed polygons represent the strict protected areas. Biome limits in dark gray lines..... 100

LISTA DE TABELAS

- Tabela 1.1 – Algumas imprecisões detectadas na localidade das estações climáticas convencionais segundo as coordenadas geográficas disponibilizadas pelo INMET junto com a informação climática.....35
- Tabela 1.2 – Correlação entre as precipitações médias mensais de cada tipo de estação climatológica e os respectivos valores de precipitação médias mensais fornecidos pelo Tropical Rainfall Measuring Mission 3B43: (r_c) correlação no caso das estações convencionais, com $p < 0,001$ para todos os casos; (r_a) correlação no caso das estações automáticas com $p < 0,001$ para todos os casos; (p_{c-a}) probabilidade resultante da comparação das correlações acima mencionadas através das transformações de Fisher (1915, 1921). Em negrito o valor estatisticamente significativo da comparação das correlações com $\alpha = 0,05$46
- Tabela 1.3 – Correlações entre as temperaturas máximas mensais de cada tipo de estação climatológica, e as elevação e a latitude respectivas. (r_c) correlação no caso das estações convencionais; (r_a) correlação no caso das estações automáticas; (p_{c-a}) probabilidade resultante da comparação das correlações acima mencionadas através das transformações de Fisher (1915, 1921). Todas as correlações com $p < 0,001$47
- Tabela 1.4 – Correlações parciais entre as temperaturas mínimas mensais de cada tipo de estação climatológica, e a elevação e a latitude correspondentes. (r_c) correlação no caso das estações convencionais; (p_c) probabilidade da correlação das estações convencionais; (r_a) correlação no caso das estações automáticas; (p_a) probabilidade da correlação das estações automáticas; (p_{c-a}) probabilidade resultante da comparação das correlações acima mencionadas através das transformações de Fisher (1915, 1921).....48
- Table 2.1 – Basic information of the bioclimatic datasets used in this study. *Ppt*: mean monthly accumulated precipitation; T_{\max} : mean monthly maximum temperature; T_{\min} : mean monthly minimum temperature. “Time span” refers to the temporal series used for determining the climatology..... 64
- Table 3.1 – Account of the species and subspecies occurring in Brazil, according to the International Ornithological Committee v. 11.1 (Gill et al. 2021). “Distribution” refers to the range described in the aforementioned source, used jointly to biomes, major hydrographic basins and reinforced by other sources such www.xeno-canto.org, to proximately assign the occurrences to the different subspecies (our “interpretation”), to create the occurrences databases used to model the subspecies distributions. 89
- Table 3.2 – Numbers of records, omissions, number of species (*sensu stricto*) recorded and predicted and their differences for the different Brazilian biomes. Numbers in parentheses represent percentages based on the total of 2608 occurrences (2600 from collections/literature + eight generated for *Glaucidium mooreorum*) or the 21 species.97

Table 3.4 – The mean \pm standard deviation of the percent of contribution and permutation importance of the explanatory variables to the SDMs of each species and subspecies. The percent of contribution is the gain of the model by including a given environmental covariate at each step of the Maxent algorithm. The permutation importance is the influence of each environmental covariate on the final model, measured through the random permutation of the values of such covariate among the training points and measuring the resulting decrease in the training area under the curve. 98

SUMÁRIO

PRIMEIRA PARTE	16
1. INTRODUÇÃO GERAL	17
REFERÊNCIAS	20
SEGUNDA PARTE	24
ARTIGO 1 — O TRADICIONAL OU O MODERNO? UMA VISÃO DA INFORMAÇÃO DA REDE DE ESTAÇÕES METEOROLÓGICAS BRASILEIRAS	25
Resumo	25
Abstract	26
1. INTRODUÇÃO	26
2. MATERIAL E MÉTODOS	28
2.1. Avaliação da localização geográfica das estações	29
2.2. Grau de adequação da informação climática segundo critérios internacionais	29
2.3. Comparação da informação fornecida por estações homônimas	31
2.4. Correlação com outros parâmetros	31
2.5. Avaliação da consistência dos dados	32
3. RESULTADOS	33
3.1. Avaliação da localização geográfica das estações	33
3.2. Grau de adequação da informação climática segundo critérios internacionais	34
3.3. Comparação da informação fornecida por EMCs e EMAs homônimas	39
3.4. Correlação com outras variáveis	40
3.5. Avaliação da consistência dos dados	44
4. DISCUSSÃO E CONCLUSÕES	44
Agradecimentos	52
Bibliografia	52
7. Material Suplementar	57
8. Endereços de internet	57
ARTIGO 2 – BRAZILCLIM: THE OVERCOMING OF LIMITATIONS OF PREEXISTING BIOCLIMATE DATA	58
Abstract	58
1. INTRODUCTION	59
2. MATERIALS AND METHODS	61
2.1. The study area	61
2.2. Comparing observed and estimated measurements	62
2.3. Creating the definitive climatic layers	66
3. RESULTS	68
3.1. Comparing observed and estimated measurements	68

3.2.	Creating the definitive climatic layers	69
4.	DISCUSSION	73
5.	CONCLUDING REMARKS	75
6.	DATA ACCESSIBILITY	76
7.	ACKNOWLEDGEMENTS	76
8.	BIBLIOGRAPHY	77
	ARTIGO 3 – EVALUATING THE BRAZILIAN SYSTEM OF STRICTLY PROTECTED AREAS USING OWLS AS AN INDICATOR GROUP	85
	ABSTRACT	85
	INTRODUCTION	86
	METHODS	87
	Study área	87
	Species and subspecies account	88
	Environmental covariates	88
	Spatial distribution models	92
	Subspecies niche analysis	93
	RESULTS	94
	The occurrences	94
	Spatial distribution models	95
	Subspecies' niche analysis	101
	DISCUSSION	101
	Sampling biases, distribution patterns and taxonomic knowledge	101
	Conservation-related issues	103
	Acknowledgements	104
	REFERENCES	104
	DATA ACCESSIBILITY STATEMENT	109

PRIMEIRA PARTE

1 INTRODUÇÃO GERAL

Prever quando e onde as espécies estão localizadas tem sido de suma importância ao longo de nossa história como espécie, tanto durante nosso prolongado período como caçadores-coletores (LEE & DALY, 1999; WHITLEY, 2014) e hoje em dia, quando as atividades humanas estão ameaçando a biodiversidade em todo o mundo em escalas sem precedentes (ALLAN et al. 2019; TILMAN et al. 2017; ZABEL et al. 2019).

Esta situação tem promovido o desenvolvimento de esforços de conservação da biodiversidade, incluindo abordagens inovadoras para a determinação de áreas prioritárias (BAX & FRANCESCONI, 2019; BRUM et al. 2017; GUIBAN et al. 2013; MININ et al. 2017). A maioria dessas abordagens envolve, em certa medida, mapear distribuições individuais de espécies, com o intuito de determinar áreas 'hotspots' de riqueza determinadas através de contagens diretas, ou ponderadas considerando níveis de endemismo ou ameaça (JENKINS et al. 2015; PIMM et al. 2014; VEACH et al. 2017).

Portanto, a modelagem preditiva de distribuição de espécies (doravante SMDs) constitui uma alternativa confiável para remediar ou mesmo contornar as limitações de métodos mais tradicionais (ELITH & LEATHWICK, 2009; MAINALI et al. 2020; PETERSON, 2001; SYFERT et al. 2014). A SDM, também conhecida como modelagem de nicho, é uma técnica moderna que utiliza algoritmos de computador, dados de ocorrência e camadas de informações ambientais para obter modelos que expliquem a distribuição probabilística de uma espécie no ambiente, buscando reduzir tanto os falsos negativos quanto os erros de falsos positivos (JIMÉNEZ-VALVERDE, 2012; MENDES et al. 2020). MaxEnt (PHILLIPS et al. 2004, 2006; PHILLIPS & DUDÍK, 2008) está entre os algoritmos para SDMs mais usados; utiliza dados ambientais de localidades de presença conhecidas e caracteriza uma amostra do “background” da área de estudo para fazer inferências de regiões de adequação dentro de uma área a partir dessas informações incompletas.

Um aspecto fundamental das SDMs é o conjunto de camadas ambientais utilizadas, entre as quais recentemente foram desenvolvidos diferentes conjuntos de dados baseados no clima (bioclimáticos). A noção de que as condições ambientais moldam a distribuição espacial dos seres vivos é conhecida desde a antiguidade, mas avaliada quantitativamente somente após o desenvolvimento de técnicas adequadas de estimativa espacial na década de 1980, que permitiram a criação destas camadas bioclimáticas (BOOTH et al. 2014). Não obstante, a relação real entre essas variáveis bioclimáticas e as respectivas tolerâncias ambientais é desconhecida para a maioria das espécies (ALVARADO-SERRANO &

KNOWLES, 2014), concomitantemente, a seleção das variáveis ambientais deve ser considerada como fundamental dentro da modelagem de nicho.

A precipitação e a temperatura são os preditores mais utilizados e, muitas vezes, os mais importantes em estudos de nicho (BRADIE & LEUNG, 2017). Os conjuntos de dados bioclimáticos desses dois importantes parâmetros foram inicialmente construídos a partir de informações de redes de estações meteorológicas terrestres, capazes de fornecer observações confiáveis com alta precisão, mas cuja incerteza aumenta quando interpolada em direção a áreas remotas (BORGES et al. 2016), e piorando sobre regiões montanhosas ou aonde as redes são esparsas (CHUBB et al. 2016; TOZER et al. 2012), ou seja, a onde a biodiversidade é geralmente maior (KÖRNER, 2002). Essas limitações levaram ao desenvolvimento de conjuntos de dados bioclimáticos baseados em dados remotamente sentidos, para a busca de precisão (DEBLAUWE et al. 2016). No entanto, mesmo essas estimativas não são isentas de erros, sendo afetadas pela resolução, taxa de chuva e frequência amostral (NESBITT & ANDERS, 2009). Atualmente, existem diferentes conjuntos de dados bioclimáticos, gerados por meio de diversas abordagens, mas todas elas potencialmente problemáticas, pelo que é necessário explorar alternativas.

Do ponto de vista da conservação, as aves de rapina constituem um paradoxo uma vez que, em comparação com outras aves, eles são um grupo pouco estudado dadas suas baixas taxas reprodutivas e abundâncias, enquanto exercem grande apelo por apoio financeiro dado seus papéis ecológicos como topos nas suas redes tróficas, simbolismo, e níveis de ameaça (DONÁZAR et al. 2016). Particularmente, as corujas verdadeiras (Strigidae) são adequadas para fins de priorização de áreas de conservação, dada a sua diversidade mundial (> 220 espécies; GILL et al. 2021), cujas espécies variam muito em tamanho, rangos de distribuição (ampla ou muito restrita), especialização de habitat (desde generalistas até altamente especializados) e, assim, resposta diferencial à estrutura (BARROS & CINTRA, 2009; BURGAS et al. 2014; SERGIO et al. 2005) e alteração (ENRÍQUEZ, 2017; RULLMAN & MARZLUFF, 2014) de habitat. De fato, vários estudos têm sinalizado as corujas como bioindicadores confiáveis de qualidade ambiental (CLAUDINO et al. 2012; DAL PIZZOL et al. 2020; DAYANANDA et al. 2016; FRÖHLICH & CIACH, 2018, 2019; HARRISON & FAHRIG, 1995; SHEFFIELD, 1997).

Pelo fato de serem de hábitos noturnos, enquanto que os ornitólogos geralmente são "diurnos", as corujas podem estar sub-representadas nos inventários da fauna (SILVA & MEDELLÍN, 2001). Essa é provavelmente a razão por trás da insuficiência de informação biológica básica para a maioria das espécies e, em última instância, da notável escassez de

espécies de corujas nas listas oficiais da fauna ameaçada brasileira (MOTTA-JUNIOR & BRAGA, 2012; MOTTA-JUNIOR et al. 2017). Até as informações sobre sua distribuição são escassas, anedóticas e mal detalhadas (MOTTA-JUNIOR; BRAGA, 2012) e provavelmente muito incompletas.

O megadiverso Brasil alberga 21 espécies (GILL et al. 2021), ou ~10 % da lista Mundial, com o Caburé-de-Pernambuco (*Glaucidium mooreorum*) sendo endêmica e criticamente ameaçada (BIRDLIFE INTERNATIONAL, 2019) ou provavelmente extinta (PEREIRA, 2010), enquanto que a Corujinha-sapo (*Megascops atricapilla*), a Corujinha-do-sul (*M. sanctaecatarinae*), o Murucututu-de-barriga-amarela (*Pulsatrix koeniswaldiana*), a coruja-listrada (*Strix hylophila*), e o caburé-miudinho (*Glaucidium minutissimum*) podem ser considerados como "quase endêmicos" (MOTTA-JUNIOR et al. 2017). No entanto, essa biodiversidade pode ser em grande parte subestimada, como repetidamente sugerido por recentes levantamentos sistemáticos moleculares que têm relatado repetidamente que as espécies tropicais podem englobar subgrupos geneticamente divergentes, que potencialmente poderiam ser elevados ao status de espécies plenas após avaliações mais aprofundadas (DANTAS et al. 2021; ISLER et al. 2007; MILÁ et al. 2012; VIEITES et al. 2009).

A necessidade de completar nosso entendimento sobre informação da biologia, ecologia e distribuição das corujas verdadeiras no Brasil, especialmente sob cenários de transformações ambientais estendidas, profundas e rápidas (ESCOBAR, 2020; LIBONATI et al. 2020; SONTER et al. 2017), coloca a modelagem de sua distribuição como um esforço científico e de conservação desafiador e urgente. Assim, no presente estudo modelou-se a distribuição de todas as espécies de corujas verdadeiras registradas até o presente no Brasil segundo GILL et al. (2021).

Para isso, primeiro avaliou-se a informação sobre temperaturas mínimas e máximas, além da precipitação, fornecida pela rede de estações meteorológicas brasileiras comparando estações convencionais e automáticas. Segundo, usou-se essa informação para avaliar o desempenho das bases bioclimáticas atualmente disponíveis, e calibrou-se as melhores usando a informação climática obtida no campo, usando algoritmos de machine-learning (gradient boosting e random forest). Estas superfícies melhoradas usaram-se para criar uma base de variáveis bioclimáticas. Terceiro, coletaram-se e filtraram-se dados de ocorrências para os Strigidae no Brasil, e geraram-se SDMs particulares para cada espécie e subespécie, avaliando a semelhança dos nichos entre as subespécies coespecíficas.

REFERÊNCIAS

- ALLAN, J.R. et al. Hotspots of human impact on threatened terrestrial vertebrates. **PLOS Biology**, v. 17, n. 3, p. 1–18, 2019.
- ALVARADO-SERRANO, D.F.; KNOWLES, L.L. Ecological niche models in phylogeographic studies: applications, advances and precautions. **Molecular Ecology Resources**, v. 14, n. 2, p. 233–248, 2014.
- BARROS, O.G.; CINTRA, R. The effects of forest structure on occurrence and abundance of three owl species (Aves: Strigidae) in the Central Amazon forest. **Zoologia**, v. 26, p. 85–96, 2009.
- BAX, V.; FRANCESCONI, W. Conservation gaps and priorities in the Tropical Andes biodiversity hotspot: Implications for the expansion of protected areas. **Journal of Environmental Management**, v. 232, p. 387–396, 2019.
- BIRDLIFE INTERNATIONAL. **IUCN Red List for birds**. Disponível em: <<http://www.birdlife.org>>. Acesso em: 18 set. 2019.
- BOOTH, T.H. et al. bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. **Diversity and Distributions**, v. 20, n. 1, p. 1–9, 2014.
- BORGES, P. DE A. et al. Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. **Theoretical and Applied Climatology**, v. 123, n. 1, p. 335–348, 1 jan. 2016.
- BRADIE, J.; LEUNG, B. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. **Journal of Biogeography**, v. 44, n. 6, p. 1344–1361, 2017.
- BRUM, F.T. et al. Global priorities for conservation across multiple dimensions of mammalian diversity. **Proceedings of the National Academy of Sciences**, v. 114, n. 29, p. 7641–7646, 2017.
- BURGAS, D. et al. Raptors as surrogates of biodiversity along a landscape gradient. **Journal of Applied Ecology**, v. 51, n. 3, p. 786–794, 2014.
- CHUBB, T.H. et al. Evaluation of the AWAP daily precipitation spatial analysis with an independent gauge network in the Snowy Mountains. **Journal of Southern Hemisphere Earth Systems Science**, v. 66, p. 55–67, 2016.
- CLAUDINO, R. et al. New record of nesting site of Tropical screech owl (*Megascops choliba*) from Brazil. **Ornitología Neotropical**, v. 23, p. 137–141, 2012.
- DAL PIZZOL, G.E. et al. Biomonitoring of owls and their environment using pellets and feathers. **Bulletin of Environmental Contamination and Toxicology**, v. 105, n. 5, p. 685–691, 1 nov. 2020.
- DANTAS, S.M. et al. Multi-character taxonomic review, systematics, and biogeography of the Black-capped/Tawny-bellied Screech Owl (*Megascops atricapilla* - *M. watsonii*) complex (Aves: Strigidae). **Zootaxa**, v. 4949, n. 3, p. 401–444, 2021.
- DAYANANDA, S.K. et al. Effects of forest fragmentation on nocturnal Asian birds: A case study from Xishuangbanna, China. **Zoological research**, v. 37, n. 3, p. 151–158, 18 maio 2016.

- DEBLAUWE, V. et al. Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. **Global Ecology and Biogeography**, v. 25, n. 4, p. 443–454, 2016.
- DONÁZAR, J.A. et al. Roles of raptors in a changing world: From flagships to providers of key ecosystem services. **Ardeola**, v. 63, n. 1, p. 181–234, 2016.
- ELITH, J.; LEATHWICK, J.R. Species distribution models: Ecological explanation and prediction across space and time. **Annual Review of Ecology, Evolution, and Systematics**, v. 40, n. 1, p. 677–697, 2009.
- ENRÍQUEZ, P.L. **Neotropical owls: Diversity and conservation**. Cham: Springer, 2017.
- ESCOBAR, H. Deforestation in the Brazilian Amazon is still rising sharply. **Science**, v. 369, n. 6504, p. 613–613, 2020.
- FRÖHLICH, A.; CIACH, M. Noise pollution and decreased size of wooded areas reduces the probability of occurrence of Tawny Owl *Strix aluco*. **Ibis**, v. 160, n. 3, p. 634–646, 2018.
- FRÖHLICH, A.; CIACH, M. Nocturnal noise and habitat homogeneity limit species richness of owls in an urban environment. **Environmental Science and Pollution Research**, v. 26, n. 17, p. 17284–17291, 2019.
- GILL, F. et al. **IOC World bird list 11.1. World Bird Names, 2021**. Disponível em: <<http://www.worldbirdnames.org/ioc-lists/crossref>>. Acesso em: 10 abr. 2021
- GUISAN, A. et al. Predicting species distributions for conservation decisions. **Ecology Letters**, v. 16, n. 12, p. 1424–1435, 2013.
- HARRISON, S.; FAHRIG, L. **Landscape pattern and population conservation**. In: HANSSON, L.; FAHRIG, L.; MERRIAM, G. (Eds.). *Mosaic Landscapes and Ecological Processes*. Dordrecht: Springer Netherlands, 1995. p. 293–308.
- ISLER, M.L. et al. Species limits in antbirds (Thamnophilidae): The Warbling Antbird (*Hypocnemis cantator*) complex. **The Auk**, v. 124, n. 1, p. 11–28, 2007.
- JENKINS, C.N. et al. US protected lands mismatch biodiversity priorities. **Proceedings of the National Academy of Sciences**, v. 112, n. 16, p. 5081–5086, 2015.
- JIMÉNEZ-VALVERDE, A. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. **Global Ecology and Biogeography**, v. 21, n. 4, p. 498–507, 2012.
- KÖRNER, C. **Mountain biodiversity, its causes and function: an overview**. In: KÖRNER, C.; SPEHN, E.M. (Eds.). *Mountain Biodiversity: A Global Assessment*. London: CRC Press, 2002. p. 3–20.
- LEE, R.B.; DALY, R.H. **Introduction: Foragers & others**. In: LEE, R.B.; DALY, R.H. (Eds.). *The Cambridge Encyclopedia of hunters and gatherers*. Cambridge, UK: Cambridge University Press, 1999. p. 1–20.
- LIBONATI, R. et al. Rescue Brazil’s burning Pantanal wetlands. **Nature**, v. 588, p. 217–219, 2020.
- MAINALI, K. et al. Matching expert range maps with species distribution model predictions. **Conservation Biology**, v. 34, p. 1292–1304, 2020.

- MENDES, P. et al. Dealing with overprediction in species distribution models: How adding distance constraints can improve model accuracy. **Ecological Modelling**, v. 431, p. 109180, 2020.
- MILÁ, B. et al. A trans-Amazonian screening of mtDNA reveals deep intraspecific divergence in forest birds and suggests a vast underestimation of species diversity. **PLOS ONE**, v. 7, n. 7, p. 1–12, 2012.
- MININ, E.D. et al. Integrating biodiversity, ecosystem services and socio-economic data to identify priority areas and landowners for conservation actions at the national scale. **Biological Conservation**, v. 206, p. 56–64, 2017.
- MOTTA-JUNIOR, J.; BRAGA, A.C. Estado del conocimiento sobre la ecología y biología de búhos en Brasil. **Ornitología Neotropical**, v. 23, p. 227–234, 2012.
- MOTTA-JUNIOR, J.C. et al. **The owls of Brazil**. In: ENRIQUEZ, P.L. (Ed.). *Neotropical Owls: Diversity and Conservation*. Cham: Springer International Publishing, 2017. p. 97–158.
- NESBITT, S.W.; ANDERS, A.M. Very high resolution precipitation climatologies from the Tropical Rainfall Measuring Mission precipitation radar. **Geophysical Research Letters**, v. 36, n. 15, 16 ago. 2009.
- PEREIRA, G.A. Avifauna associada a três lagoas temporárias no estado do Rio Grande do Norte, Brasil. **Atualidades Ornitológicas**, v. 156, p. 53–60, 2010.
- PETERSON, A.T. Predicting species' geographic distributions based on ecological niche modeling. **The Condor**, v. 103, n. 3, p. 599–605, 2001.
- PHILLIPS, S.J. et al. Maximum entropy modeling of species geographic distributions. **Ecological Modelling**, v. 190, n. 3, p. 231–259, 2006.
- PHILLIPS, S.J.; DUDÍK, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. **Ecography**, v. 31, n. 2, p. 161–175, 2008.
- PHILLIPS, S.J. et al. A Maximum entropy approach to Species Distribution Modeling. **Proceedings of the Twenty-first International Conference on Machine Learning**, p. 655–662, 2004
- PIMM, S.L. et al. The biodiversity of species and their rates of extinction, distribution, and protection. **Science**, v. 344, n. 6187, 2014.
- RULLMAN, S.; MARZLUFF, J.M. Raptor presence along an urban–wildland gradient: Influences of prey abundance and land cover. **Journal of Raptor Research**, v. 48, n. 3, p. 257–272, 2014.
- SERGIO, F. et al. Top predators and biodiversity. **Nature**, v. 436, n. 7048, p. 192–192, 2005.
- SHEFFIELD, S.R. **Owls as biomonitors of environmental contamination**. In: DUNCAN, J.R.; JOHNSON, D.H.; NICHOLLS, T.H. (Eds.). *Biology and conservation of owls of the northern hemisphere*. 2nd International symposium. 1997, v. 190, p. 383–398.
- SILVA, H.G. DE; MEDELLÍN, R.A. Evaluating completeness of species lists for conservation and macroecology: A case study of Mexican land birds. **Conservation Biology**, v. 15, n. 5, p. 1384–1395, 2001.
- SONTER, L.J. et al. Mining drives extensive deforestation in the Brazilian Amazon. **Nature Communications**, v. 8, n. 1, p. 1013, 2017.

- SYFERT, M.M. et al. **Using species distribution models to inform IUCN Red List assessments.** *Biological Conservation*, v. 177, p. 174–184, 2014.
- TILMAN, D. et al. Future threats to biodiversity and pathways to their prevention. *Nature*, v. 546, n. 7656, p. 73–81, 2017.
- TOZER, C.R. et al. On the uncertainties associated with using gridded rainfall data as a proxy for observed. *Hydrology and Earth System Sciences*, v. 16, n. 5, p. 1481–1499, 2012.
- VEACH, V. et al. Species richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity. *Diversity and Distributions*, v. 23, n. 7, p. 715–726, 2017.
- VIEITES, D.R. et al. Vast underestimation of Madagascar’s biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences*, v. 106, n. 20, p. 8267–8272, 2009.
- WHITLEY, D.S. **Future directions in hunter-gatherer research: hunter-gatherer religion and ritual.** In: CUMMINGS, V.; JORDAN, P.; ZVELEBIL, M. (Eds.). *Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers*. Oxford, UK: Oxford University Press, 2014. p. 1221–1242.
- ZABEL, F. et al. Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. *Nature Communications*, v. 10, n. 1, p. 2844, 28 jun. 2019.

SEGUNDA PARTE

1 **ARTIGO 1 — O TRADICIONAL OU O MODERNO? UMA VISÃO DA**
2 **INFORMAÇÃO DA REDE DE ESTAÇÕES METEOROLÓGICAS BRASILEIRAS**

3
4 Paolo Ramoni Perazzi^{1,2}, Marcelo Passamani¹, Dirk Thielen³, Carlos Padovani⁴

5
6 ¹Laboratório de Ecologia e Conservação de Mamíferos, Universidade Federal de Lavras,
7 Lavras, Minas Gerais, Brasil.

8 ²Centro de Simulación y Modelos, Universidad de los Andes, Facultad de Ingeniería, Mérida,
9 Venezuela.

10 ³Laboratorio de Ecología del Paisaje y Clima, Centro de Ecología, Instituto Venezolano de
11 Investigaciones Científicas, Miranda, Venezuela.

12 ⁴Empresa Brasileira de Pesquisa Agropecuária, Centro de Pesquisa Agropecuária do Pantanal,
13 Mato Grosso do Sul, Brasil.

14
15 Autor de correspondência: Paolo Ramoni-Perazzi; rpaolo1967@gmail.com.

16
17 (Artigo publicado na *Revista Brasileira de Meteorologia*, [http://dx.doi.org/10.1590/0102-](http://dx.doi.org/10.1590/0102-77863630151)
18 [77863630151](http://dx.doi.org/10.1590/0102-77863630151))

19
20 **Resumo**

21
22 Apresenta-se uma análise comparativa com base em dados médios mensais de precipitação
23 acumulada e temperaturas máximas e mínimas registradas durante o período 2000-2019 pelas
24 estações meteorológicas convencionais e automáticas existentes no Brasil. Foi avaliada a
25 exatidão da localização das estações (coordenadas geográficas) fornecida pelo Instituto
26 Nacional de Meteorologia (INMET), foi comparada gráfica e estatisticamente a informação
27 fornecida por estações homônimas, foi correlacionada esta informação com variáveis
28 independentes (elevação e latitude no caso das temperaturas e informação de satélite da
29 Tropical Rainfall Measuring Mission -TRMM 3B43- no caso da precipitação), e foi avaliada a
30 consistência da informação fornecida pelas estações comparada com os dados das cinco
31 estações geograficamente mais próximas, através de interpolações espaciais baseadas em
32 distâncias inversas ponderadas sob um esquema de Jaccknife. Ainda que a informação
33 fornecida ambos os tipos de estações podem ter boa concordância em muitos casos, existem

34 incongruências que em alguns casos podem ser muito marcadas. Tais diferenças podem ser
35 atribuídas tanto a erros sistemáticos como aleatórios, de origem indeterminada. Ambos os
36 conjuntos de dados têm problemas, mas a evidência avaliada sugere que as estações
37 convencionais fornecem dados de precipitação um pouco mais confiáveis, enquanto que as
38 automáticas são mais consistentes em relação a informação de temperaturas.

39 Palavras-chave: estação meteorológica, clima, parâmetros meteorológicos, Brasil.

40

41 **The Traditional or the Modern? A Vision of the Information from the Brazilian** 42 **Weather Stations Network**

43

44 **Abstract**

45

46 A comparative analysis is presented based, on one side, mean monthly accumulated
47 precipitation and, on the other, maximum and minimum monthly mean temperatures,
48 recorded during the period 2000-2019 by conventional and automatic weather stations
49 existing in Brazil. The accuracy of the location of the stations (geographic coordinates)
50 provided by the National Institute of Meteorology (INMET) was evaluated, comparing
51 graphically and statistically the information provided by homonymous stations, this
52 information was correlated with independent variables (elevation and latitude in the case of
53 temperatures, and satellite information from the Tropical Rainfall Measuring Mission -
54 TRMM 3B43-, in the case of precipitation), and evaluate the consistency of the information
55 provided by the stations compared with data from the five stations geographically closest
56 through spatial interpolations based on inverse-distance weighted under a Jaccknife scheme.
57 Although the information provided by both types of stations may have good agreement in
58 many cases, there are incongruities that in some cases can be very marked. Such differences
59 can be attributed to both systematic and random errors, of indeterminate origin. Both data sets
60 have problems, but the evidence evaluated suggests that conventional stations provide slightly
61 more reliable precipitation data, while automatic ones are more consistent with regards to
62 temperature information.

63 Keywords: weather stations, climate, meteorological parameters, Brazil.

64

65 **1. INTRODUÇÃO**

66

67 O Brasil é o quinto maior país em superfície do mundo, abrangendo condições
68 climáticas contrastantes como consequência dos grandes gradientes de altitude, latitude,
69 sazonalidade, entre outros. Num país com tais características o estudo da variabilidade
70 espacial dos parâmetros climáticos é essencial para o conhecimento, planejamento e ação de
71 uma grande variedade de aspetos. Entre estes, destacam a agricultura e pecuária,
72 biodiversidade, saúde pública, ou desenvolvimento sustentável. Por exemplo, esta informação
73 é fundamental na tomada de decisões por parte de órgãos dos governos ou empresas, tanto
74 nacionais como estrangeiras, podem ser usados pelas famílias ou pelas gerações futuras como
75 bens de consumo pessoal para o seu dia-a-dia, podem ser usados pelos cientistas como
76 entradas em modelos para explicar ou prever melhor eventos naturais ou induzidos pelo
77 homem, ou podem indicar o tipo de culturas adequadas para uma região determinando o
78 sucesso no desenvolvimento da economia agrária de um país (Das *et al.*, 2003).

79 O conceito de clima refere-se ao conjunto de condições atmosféricas médias,
80 normalmente medidas durante um período de 30 anos segundo a Organização Meteorológica
81 Mundial (WMO; pelas suas siglas no inglês; Planton, 2013), as quais caracterizam uma região
82 e influenciam marcadamente os ecossistemas (Köppen, 1900). Nas últimas décadas, a
83 informação climática tem sido obtida através de sensoriamento remoto, incluindo aquela
84 informação gerada espacialmente, mas tradicionalmente tem sido obtida pelo meio de
85 medidas feitas *in situ* usando estações meteorológicas (WMO, 2018). Uma estação
86 meteorológica é uma instalação dotada com instrumentos e sensores para medir as variáveis
87 atmosféricas, particularmente temperatura, pressão atmosférica, umidade, velocidade do
88 vento, direção do vento e precipitação. Nas estações meteorológicas convencionais (EMCs),
89 as medições são feitas por observadores humanos pelo menos uma vez por dia, enquanto nas
90 estações meteorológicas automatizadas (EMAs) o observador humano é substituído por
91 aparelhos autônomos e as medições são normalmente feitas e transmitidas pelo menos uma
92 vez por hora automaticamente (WMO, 2018).

93 Por ter sido as primeiras em ter sido utilizadas, as EMCs normalmente contam com
94 series de dados temporalmente mais prolongadas, mas a dependência diária do elemento
95 humano tem limitado geograficamente as possibilidades de sua instalação, além de
96 incrementar as probabilidades de erros na leitura/transcrição das medições. A partir do ano
97 2000, o Instituto Nacional de Meteorologia (INMET) adicionou a tecnologia de 450 estações
98 automáticas (EMAs) às 293 EMCs do sistema brasileiro de estações meteorológicas (Lucas *et al.*, 2010). A adoção de EMAs apresenta certas vantagens, que segundo Ahmad *et al.* (2017)
99 incluem: (1) a possibilidade de observação contínua; (2) os dados podem ser obtidos mesmo
100

101 quando não há pessoal presente; (3) em consequência podem ser instaladas em locais
102 inacessíveis; (4) reduz o número de observadores e os custos operacionais; (5) uma vez que os
103 dados meteorológicos são tomados como sinais elétricos, erros de observadores em leitura são
104 eliminados; (6) as técnicas de observação padronizadas permitem a homogeneização dos
105 dados observados em regiões aonde a observação meteorológica automática é adotada; (7)
106 novos elementos de observação podem ser adicionados com relativa facilidade instalando
107 novos instrumentos; (8) permite escolher instrumentos de medição ideais com o nível
108 adequado de precisão na medição; e (9) a necessidade de treinamento de observadores é
109 eliminada. Naturalmente, a sua adoção não está isenta de desvantagens ou fragilidades, por
110 exemplo: (a) na ausência de observadores, pode ocorrer vandalismo e roubo de equipamentos;
111 (b) requerem um programa de manutenção com técnicos especializados para defeitos
112 mecânicos, elétricos e eletrônicos intrínsecos; (c) problemas provocados por intempéries
113 naturais (insetos, aves, roedores, poeira, objetos que podem ser atirados contra as estações em
114 tempestades com ventos fortes; (d) problemas relacionados com a manutenção do local onde
115 estão as estações, com controle da vegetação; (e) proteção com cercamento para evitar
116 problemas com animais grandes como o gado ou animais grandes selvagens; (f) exigem um
117 período inicial de calibração comparativa com as estações convencionais o que nem sempre é
118 possível; (g) exigem calibrações regulares.

119 A fim de avaliar as limitações inerentes à geração de dados meteorológicos, subsidiar
120 possíveis melhorias e para alertar aos usuários qual fonte de informação deveria ser
121 selecionada preferencialmente para a aplicação desses dados nas mais diversas áreas de
122 conhecimento, comparamos os dados das redes de EMCs e de EMAs, particularmente dos
123 parâmetros precipitação, e temperaturas mínima e máxima mensais.

124

125 2. MATERIAL E MÉTODOS

126

127 Neste estudo, foram analisados 19 anos de informação climática, pois, segundo a
128 informação disponível, as primeiras EMAs iniciaram operar no dia primeiro de junho do
129 2000, e para este estudo foi incluída a informação obtida até 30 de setembro do 2019.

130 As informações correspondentes às EMCs foram obtidas diretamente via *on line* na sessão de
131 dados históricos do portal web do INMET, sob formato de valores diários. No caso das
132 EMAs, as informações foram fornecidas pela equipe da Seção de Armazenamento de Dados
133 Meteorológicos do INMET através de um disco digital versátil (DVD) com os dados horários

134 em formato compatível com o software usado neste estudo. Para realizar as comparações,
135 foram analisadas apenas as informações das EMCs disponíveis para o mesmo período das
136 EMAs. Todas as comparações e cálculos foram feitos usando a linguagem de programação R
137 versão 3.6.1 (R Core Team, 2019). Toda a informação geográfica foi projetada segundo o
138 Sistema de Referência Geocêntrico para as Américas (SIRGAS2000; IBGE, 2005).

139

140 **2.1. Avaliação da localização geográfica das estações**

141

142 O crescente desenvolvimento de ferramentas de modelagem espacial, além do
143 incremento na resolução espacial de dados usados em climatologia e meteorologia, aumentou
144 a necessidade do conhecimento da exatidão (concordância entre as coordenadas disponíveis e
145 a potencial localidade) e precisão (quantidade de posições decimais) das coordenadas
146 geográficas das estações meteorológicas (Carrega, 2013). Conseqüentemente, exploramos a
147 precisão e exatidão das coordenadas geográficas fornecidas pelo INMET conjuntamente com
148 os dados para cada estação. Primeiro, já que normalmente o nome das estações corresponde
149 ao nome do município, verificou-se se as coordenadas da estação concordam com aquelas do
150 município homônimo através da malha municipal fornecida *online* sob formato *shapefile* pelo
151 Instituto Brasileiro de Geografia e Estatística (IBGE, 2018; acessado no 10 de maio do 2020),
152 como uma primeira avaliação de quantas e quais estações poderiam ter inconsistências.
153 Segundo, as coordenadas de cada estação foram exploradas usando recursos *online*, tais como
154 Google Earth®, prestando atenção se elas correspondiam às locações homônimas e inclusive
155 a estações meteorológicas ou estruturas onde estas poderiam estar localizadas, tais como
156 aeroportos, etc.

157 Para as análises seguintes, as coordenadas das EMCs foram corrigidas com as
158 coordenadas atualizadas que o INMET fornece em outra parte separada do seu site web
159 (acessado no 10 de maio do 2020), quando estas estiverem disponíveis.

160

161 **2.2. Grau de adequação da informação climática segundo critérios internacionais**

162

163 A Organização Meteorológica Mundial é uma organização intergovernamental que
164 fornece liderança mundial e expertise em cooperação internacional na entrega e utilização de
165 serviços climáticos, hidrológicos e afins por seus Estados membros, incluindo o avance na
166 standardização de dados meteorológicos, publicando uma ampla variedade de manuais.

167 Consequentemente, o presente estudo está baseado em “normais provisionais”, tais
168 como foram definidas pela WMO (1989), ou seja, como os valores médios calculados com
169 base em observações que se estendem por um período curto de pelo menos dez anos. Mais
170 ainda, adotou-se o critério 4/10 recomendado pela WMO (2011) segundo o qual um valor
171 mensal, ou a média dos valores diários daquele mês, não deve ser calculado se num mês dado
172 faltam observações para (a) cinco ou mais dias consecutivos ou (b) onze ou mais dias
173 dispersos.

174 Os números de meses nominalmente disponíveis foram determinados para cada
175 estação, como o total de meses contabilizados entre o primeiro de junho do 2000 e o 30 de
176 setembro do 2019. Não obstante, estas datas variaram, dependendo se uma estação dada foi
177 colocada em serviço após a data inicial indicada, ou parou de funcionar antes da data final
178 indicada, em cujos casos os meses foram contados a partir do primeiro ou até o último registro
179 climatológico disponível, segundo o caso. Adicionalmente, foi contada a quantidade de meses
180 efetivamente disponíveis para cada estação, entendidos como aqueles que seguem os critérios
181 da WMO previamente indicados. Finalmente, para cada um dos parâmetros considerados
182 neste estudo foram criados diagramas de *boxplot* mensais, e testes de *t* de Student foram
183 aplicados para comparar a quantidade de anos totais operativos, a quantidade de anos totais
184 efetivos e a proporção de gaps na informação climática mensal. Os entalhes dos *boxplot*
185 mostram o intervalo de confiança ao redor da mediana e, ainda que não se constitua num teste
186 formal, se os entalhes das duas caixas não se sobrepõem, existe uma “forte evidencia” (95%
187 de confiança) que as suas medianas diferem (Chambers *et al.*, 1983).

188 Aqueles estações que cumpriram os critérios estabelecidos pelo WMO previamente
189 mencionados foram usadas nas análises remanescentes, após completar a informação diária
190 faltante, estimada de duas maneiras complementares. Primeiro, transversalmente (cada dia por
191 vez) para todas as estações com dados faltantes, diante interpolações espaciais, naqueles casos
192 quando dados de 20 ou mais estações estiveram disponíveis. As interpolações são predições
193 de valores em pontos de interesse dentro do âmbito do estudo, e no caso das interpolações
194 espaciais atualmente existem muitas técnicas disponíveis, com diferentes capacidades e
195 desempenhos (Li e Heap, 2011; Kresse e Danko, 2012; Raju, 2016; Zhou *et al.*, 2017), entre
196 as quais foi escolhida a distância inversa ponderada (IDW, pelas suas siglas no inglês),
197 inicialmente proposta por Shepard (1968), usando o pacote *gstat* (Pebesma, 2004; Gräler *et*
198 *al.*, 2016) na qual o peso das cinco estações com informação e geograficamente mais
199 próximas foi assumido como inversamente proporcional à distância geográfica até a estação
200 alvo, usando um valor da potência igual a 0,5. Esta escolha foi baseada no fato de que a IDW:

201 (1) constitui uma técnica determinística matematicamente simples, cuja demanda
202 computacional é pouca; e (2) tem sido amplamente usada com resultados satisfatórios (Chen e
203 Liu, 2012; Bagheri, 2016). Segundo, longitudinalmente (cada estação por separado) no caso
204 daqueles dados faltantes que não puderam ser estimados diante interpolações espaciais, diante
205 interpolações temporais baseadas em modelos de estado-espaco inicialmente proposto por
206 Kálmán (1960), implementado no pacote imputeTS (Moritz e Bartz-Beielstein, 2017). A esta
207 escolha foi baseada no fato que a representação do estado-espaco e as equações recursivas que
208 caracterizam o filtro de Kalman são ideais para analisar series com dados faltantes (Brockwell
209 e Davis, 1991), tendo sido particularmente recomendado para o seu uso em dados
210 meteorológicos (Joyce e Xie, 2011; Sivagami *et al.*, 2019).

211

212 **2.3. Comparação da informação fornecida por estações homônimas**

213

214 Foi calculada e comparada a informação média mensal fornecida por aquelas EMCs e
215 EMAs que tinham nomes semelhantes e coordenadas geográficas próximas (afastadas até por
216 de 25 km). Para cada parâmetro, foram criados diagramas de boxplot mensais, e foram
217 aplicados testes pareados de t de Student.

218 Além disso, para cada parâmetro foram comparadas as magnitudes das diferenças
219 entre os valores fornecidos pelas EMCs em relação às respectivas EMAs, através de subtração
220 simples. Representou-se espacialmente os resultados destas subtrações por meio de mapas de
221 bolhas, usando as coordenadas geográficas das EMAs, o que permitiu explorar visualmente a
222 existência de possíveis padrões espaciais por meio de agregação das diferenças, além de
223 histogramas para visualizar a distribuição das frequências de tais diferenças.

224

225 **2.4. Correlação com outros parâmetros**

226

227 É bem sabido que a distribuição da temperatura é determinada por diferentes fatores,
228 particularmente pela latitude e a altitude (Trapasso, 2008; Hartmann, 2016; Collier, 2016).
229 Assim, foi explorado o grau de correlação entre as temperaturas mínimas e máximas
230 registradas pelo conjunto de EMCs e EMAs homônimas e as latitudes fornecidas pelas
231 estações e com as altitudes correspondentes a tais coordenadas segundo o modelo digital de
232 elevação GMTED2010 (Danielson e Gesch, 2011), com resolução de 30 arcossegundos
233 (aproximadamente 90 m no Equador). Estas correlações foram parciais, ou seja, controlando o

234 efeito da elevação ou da latitude enquanto se avaliou a correlação das temperaturas com a
235 latitude ou a elevação, usando o pacote ppcor (Kim, 2015).

236 O padrão de distribuição espacial da precipitação é consideravelmente mais complexo
237 que a temperatura, porque vem influenciada por fatores envolvidos tanto no movimento
238 vertical da atmosfera, quanto na natureza da mesma, a estabilidade/instabilidade atmosférica,
239 ou as suas características de umidade e termais (Granger, 2008; Hartmann, 2016; Collier,
240 2016). Assim, a precipitação normalmente está pouco correlacionada com características
241 topográficas (Daly *et al.*, 1994, 2008; Ramoni-Perazzi *et al.*, 2016), um dos motivos pelos
242 quais o sensoriamento remoto, especialmente aquele baseado no uso do satélite, tornou-se a
243 principal fonte de dados precisos e espacialmente contínuos para fazer estimativas alternativas
244 ou suplementares da precipitação (Javanmard *et al.*, 2010; Shrivastava *et al.*, 2014; Manta *et al.*,
245 2015). Por estes motivos, foi explorada a correlação entre as precipitações médias mensais
246 registradas pelas EMCs e EMAs homônimas em relação à informação fornecida pela Tropical
247 Rainfall Measuring Mission TRMM 3B43. Trata-se de um satélite cujo radar mediu a
248 precipitação espacialmente durante mais de 17 anos a partir do 1997, baseado nos sensores de
249 radiação infravermelha e micro-ondas (em mm/h) e estimativas do erro na precipitação
250 respeito as estações meteorológicas (Kummerow *et al.*, 2000). Esta combinação utiliza as
251 medições do sensor na região espectral das micro-ondas, as quais tem melhor relação entre as
252 taxas de chuva e a microfísica das nuvens, para calibrar imagens do canal infravermelho, as
253 quais tem resolução espacial e temporal suficientes para monitorar sistemas convectivos
254 (Levizzani, 2000). Os dados da TRMM, tem encontrado aplicabilidade nas ciências agrárias
255 (Cashion *et al.*, 2005; Arvor *et al.*, 2014), modelagem hidrológica (Meng *et al.*, 2014) e seca
256 meteorológica (Sahoo *et al.*, 2015), mostrando um bom desempenho (Dinku *et al.*, 2007; Cao
257 *et al.*, 2018).

258 Em cada caso, as informações climáticas fornecidas pelas EMCs e as EMAs foram
259 comparadas pelo seu grau de correlação com os parâmetros independentes respectivos antes
260 indicados, por meio de transformações de Fisher (1915, 1921). Estas transformações, usadas
261 para encontrar intervalos de confiança tanto para o valor da correlação quanto para diferenças
262 entre correlações, têm sido mais comumente usadas para testar a significância da diferença
263 entre dois coeficientes de correlação.

264

265 **2.5. Avaliação da consistência dos dados**

266

267 Foi examinada a coerência dos dados de temperaturas medias mensais e dos totais
 268 mensais de precipitação entre cada uma das estações EMCs e EMAs homônimas a partir de
 269 interpolações espaciais. Usamos as distâncias inversas ponderadas já mencionadas, a fim de
 270 estimar o valor esperado para cada estação, usando a informação das estações remanescentes
 271 do mesmo tipo, sob um esquema de Jackknife (Efron, 1982). O pressuposto é de que a
 272 coerência entre valores observados e interpolados é um bom indicativo da qualidade da
 273 informação. Estes valores interpolados foram comparados com os valores medidos usando a
 274 raiz quadrática média dos erros relativos (RRMSE pelas suas siglas no inglês), uma das
 275 medidas de desempenho das interpolações proposta por Li e Heap (2011), que subtrai o efeito
 276 da unidade/escala, que por ser quadrática mostra que as diferenças extremas têm maior peso, e
 277 cuja fórmula é:

$$278 \quad RRMSE = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{[p_i - o_i]}{o_i} \right)^2 \right]^{1/2} * 100$$

279 onde n é o número de estações, o é o valor medido do parâmetro e p corresponde aos valores
 280 interpolados.

281

282 3. RESULTADOS

283

284 3.1. Avaliação da localização geográfica das estações

285

286 Foi avaliado um total de 265 EMCs e 423 EMAs, no Brasil. Uma comparação das
 287 coordenadas geográficas atribuídas às estações mostrou claramente uma maior exatidão no
 288 caso das EMAs. Efetivamente, o INMET fornece as coordenadas em formato de grau
 289 decimal, que no caso das EMCs consiste em apenas duas posições decimais (pouca exatidão),
 290 indicativo de uma localidade que pode estar próxima à localização real, mas imprecisa, e não
 291 se ajusta aos requerimentos estabelecidos pela WMO (2018), uma vez que as coordenadas
 292 têm que estar em graus, minutos e segundos inteiros, ou seja, pelo menos quatro posições
 293 decimais. Em contrapartida, as coordenadas das EMAs contêm um maior número de posições
 294 decimais (seis), que em muitas ocasiões correspondem com estruturas nas quais é comum o
 295 estabelecimento de estações climatológicas (aeroportos, institutos de ensino, etc.), ou nas
 296 quais inclusive foi possível observar instalações similares àquelas das estações climatológicas
 297 típicas através dos aplicativos usados (Google Earth®). Isso sugere que no caso das EMCs, a

298 informação fornecida em muitos casos é antiga e pode não ter sido atualizada com GPS de
299 precisão.

300 No caso das EMCs, 17 delas têm coordenadas geográficas que claramente ficam fora dos
301 limites dos municípios homônimos (afastadas até 79,7 km; Tabela 1.1) ou, inclusive, em
302 estados vizinhos. Ainda que, como já foi indicado, em alguns casos o INMET forneça as
303 coordenadas atualizadas em outra parte da sua plataforma web, a disponibilização das
304 informações da maneira atual pode gerar erros.

305 Em contrapartida, a única inconsistência que encontramos na informação associada às
306 EMAs foi um erro no nome da estação "Novo Mundo" (código WMO: 86643), que
307 corresponde à comunidade "Mundo Novo", município Brasnorte, Mato Grosso.

308

309 **3.2. Grau de adequação da informação climática segundo critérios internacionais**

310

311 As EMCs já estavam ativas para a data inicial considerada neste estudo (o primeiro de
312 junho do 2000), enquanto as EMAs foram-se agregando depois paulatinamente, o que ficou
313 evidenciado no fato que as primeiras contam com um período de operação significativamente
314 maior (média geral 18,4 anos; Fig. 1.1), do que as segundas (média geral 12,5 anos), durante o
315 período considerado neste estudo. Similarmente, as quantidades de anos disponíveis cuja
316 informação concorda com os critérios da WMO (pelo menos 10 anos de dados sob o critério
317 de quatro dias consecutivos ou 11 dias não consecutivos por mês) é maior nas EMCs (média
318 geral 17,7 anos; Fig. 1.2) que nas EMAs (média geral 11,3 anos). As EMAs mostraram uma
319 proporção significativamente maior de lacunas na informação (médias gerais: 0,15 no caso da
320 precipitação, e 0,12 no caso das temperaturas; Fig. 1.3) do que nas EMCs (médias gerais: 0,03
321 na precipitação, e 0,06 e 0,05 no caso das temperaturas máximas e mínimas respectivamente).

Tabela 1.1 – Algumas imprecisões detectadas na localidade das estações climáticas convencionais segundo as coordenadas geográficas disponibilizadas pelo INMET junto com a informação climática.

Nome da estação	Município das coordenadas	Erro (km)
(83582) Bambuí - MG	Divinópolis - MG	79,7
(83687) Lavras - MG	São Thomé das Letras - MG	39,3
(83386) Januária - MG	Itacarambira - MG	20,5
(82693) Cruzeta - RN	Acari - RN	18,4
(83446) Guaratinga - BA	Itabela - BA	13,4
(82596) Ceará-Mirim - RN	Poço Branco - RN	12,0
(82975) Bom Jesus do Piauí - PI	Santa Luz - PI	6,7
(82487) Guaramiranga - CE	Mulungu - CE	5,3
(82296) Luzilândia Lag. do Piauí - PI	São Bernardo - MA	3,8
(83195) Itabaianinha - SE	Riachão do Dantas - SE	3,1
(82598) Natal - RN	Parnamirim - RN	2,5
(83985) Pelotas - RS	Capão do Leão - RS	1,7
(82789) Triunfo - PE	São José de Princesa - PB	1,4
(83395) Janaúba - MG	Nova Porteirinha - MG	1,2
(83037) Coronel Pacheco - MG	Goianá - MG	0,1
(82863) Pedro Afonso - TO	Tupirama - TO	0,1
(82797) Surubim - PE	Casinhas - PE	0,1

Figura 1.1 – Comparação da quantidade de anos totais operativos, contabilizados dentro do período de 19 anos entre o primeiro de junho de 2000 e 30 de setembro de 2019, para cada mês e para cada tipo de estação no caso de: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima.

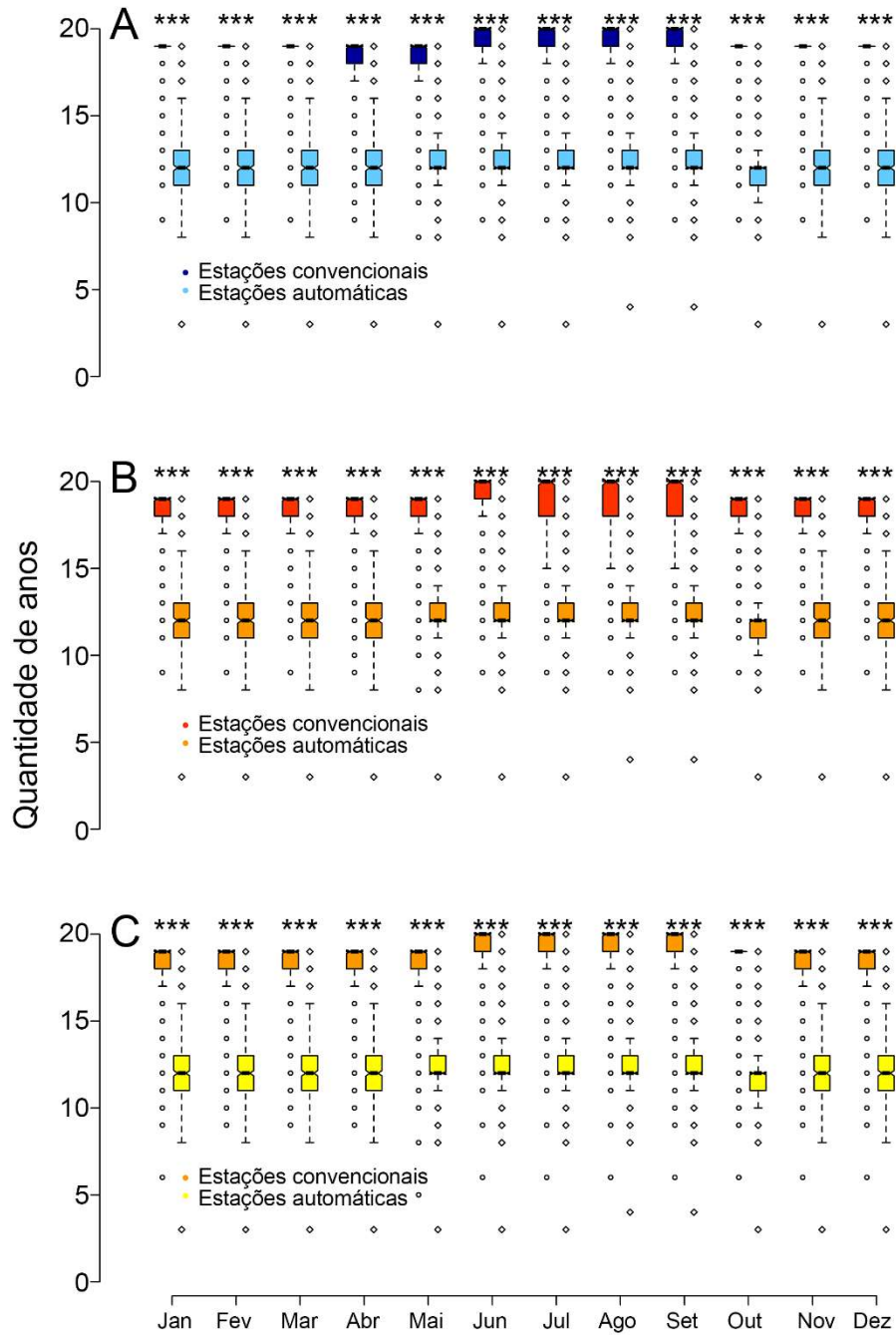


Figura 1.2 – Comparação da quantidade de anos totais efetivos, ou seja, com informação acorde aos critérios da WMO (2017), contabilizados dentro do período de 19 anos entre o primeiro de junho de 2000 e 30 de setembro de 2019, para cada mês e para cada tipo de estação no caso de: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima.

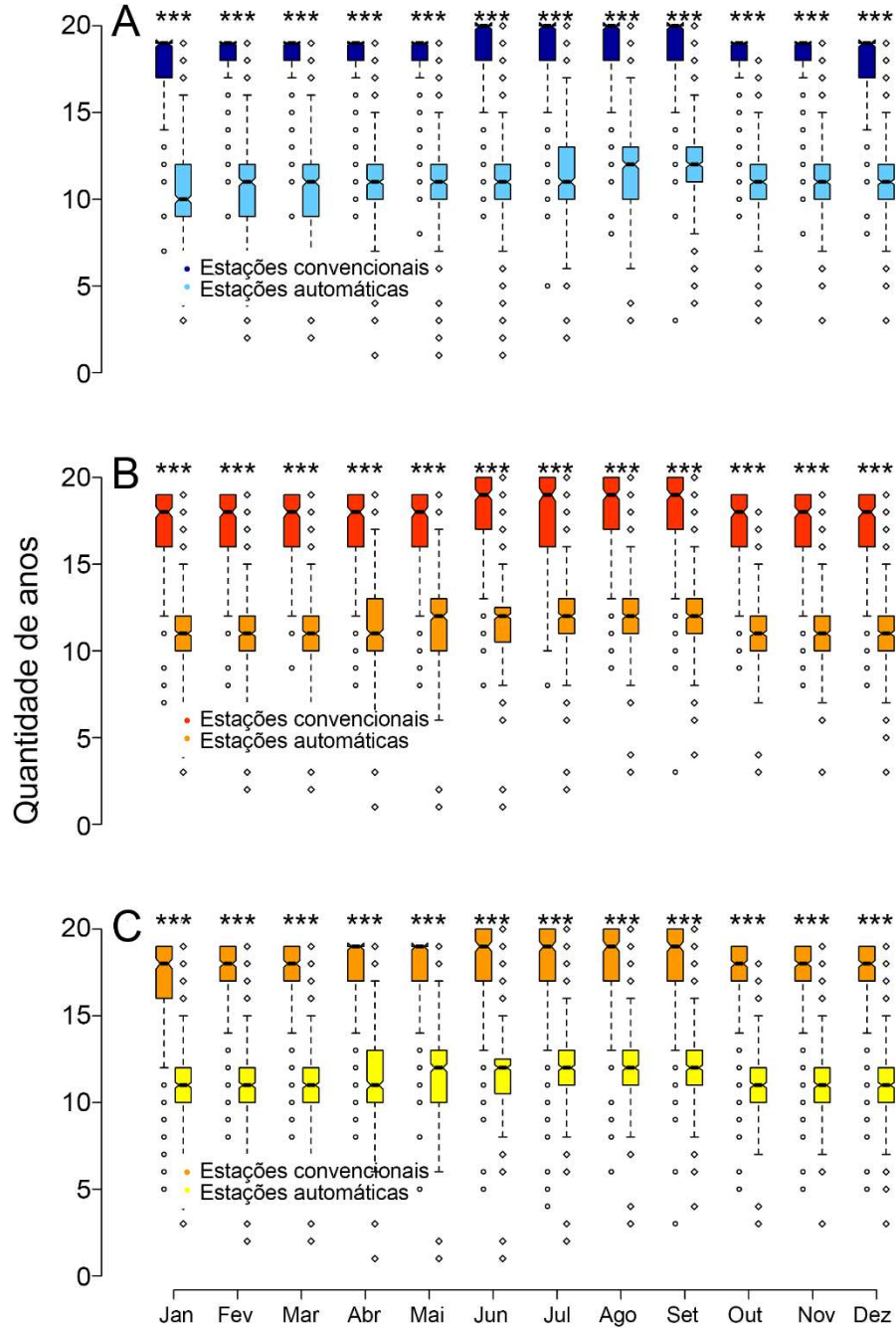
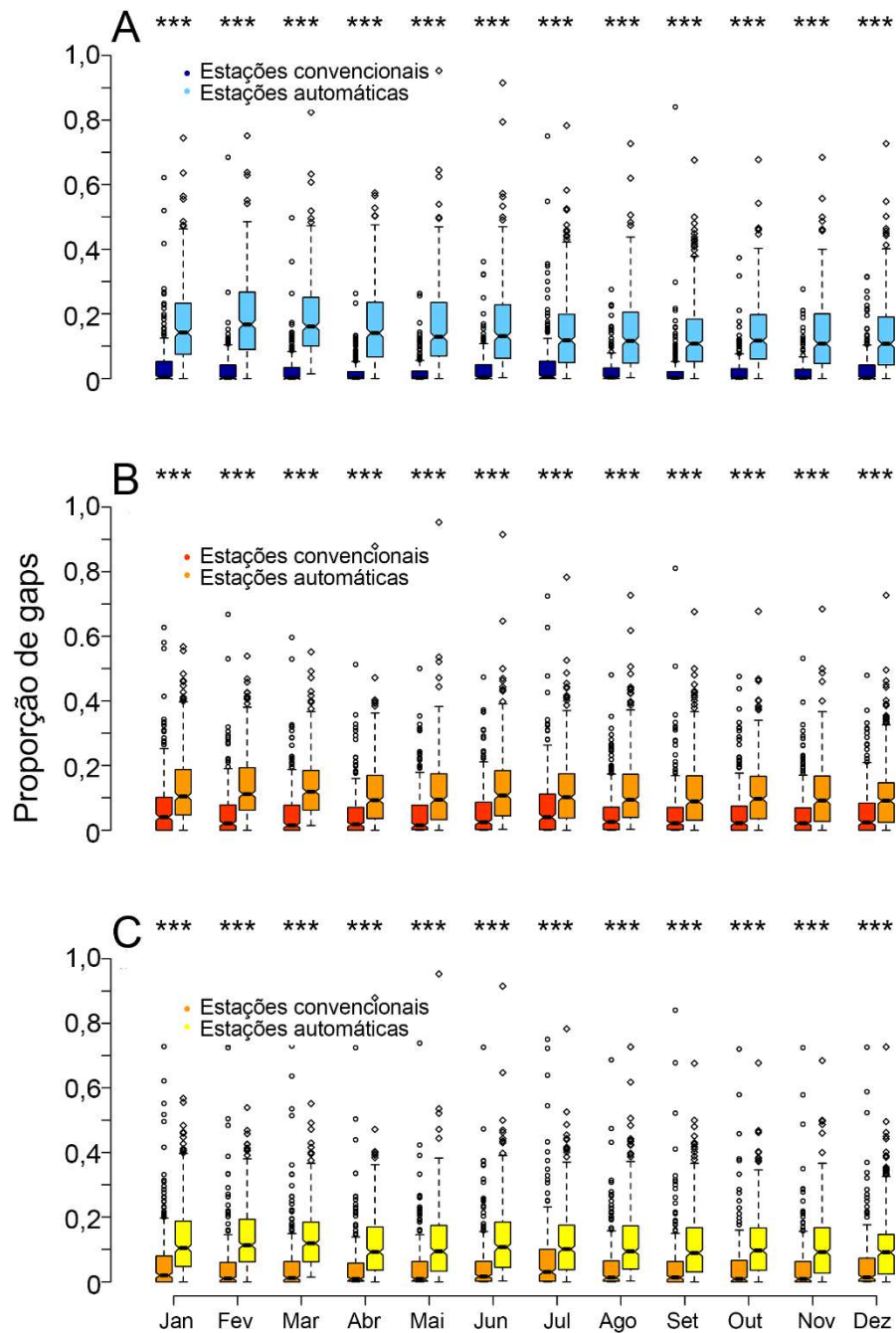


Figura 1.3 – Proporção de gaps na informação climática mensal para cada tipo de estação no caso de: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima.



325

326

327

328 Das 265 EMCs avaliadas, 257 (97,0%) forneceram dados de precipitação de acordo
 329 com os critérios da WMO para todos os meses, 255 (96,2% do total) no caso da temperatura
 330 máxima e 249 (94,0%) no caso da temperatura mínima. Em contraste, das 423 EMAs
 selecionadas, apenas 191 (45,2%) forneceram informação adequada no caso da precipitação, e

331 251 (59,3%) no caso das temperaturas. Uma lista completa, com detalhes adicionais, das
332 estações que cumpriram ou não com os critérios da WMO é fornecida na Tabela S1.

333

334 **3.3. Comparação da informação fornecida por EMCs e EMAs homônimas**

335

336 Após excluir as estações “São Simão” (EMCs WMO: 83669 - SP e EMAs WMO:
337 86773 - GO), um total de 76 pares de estações no caso da precipitação, 89 no caso da
338 temperatura máxima e 85 no caso da temperatura mínima, tiveram nomes iguais ou
339 semelhantes e coordenadas geográficas próximas, com distâncias entre coordenadas variando
340 entre 0,0 e 20,9 km (mediana 0,57 km no caso da precipitação e 0,79 km no caso das
341 temperaturas). Mais ainda, estas distâncias não necessariamente significam que ambas as
342 estações estejam de fato afastadas: pode simplesmente ser um artefato originado em erros nas
343 coordenadas ou pela baixa exatidão nas coordenadas fornecidas pelo INMET para as EMCs.
344 O previamente indicado reforça a validade das comparações a seguir.

345 Foi observado que, no caso da precipitação acumulada mensal, ainda que para todos os
346 meses, os boxplots mostram sobreposição nas distribuições e, mais ainda, nos entalhes, os
347 testes pareados *t* de Student indicam que existem diferenças estatisticamente significativas ao
348 longo de todo o ano exceto em agosto e outubro (Fig. 1.4A). Estas diferenças não seguem um
349 padrão geográfico evidente (Fig. 1.5), para as subestimativas ou sobreestimativas das EMAs
350 em relação as EMCs, pois não é possível observar agregamentos geográficos, o que poderia
351 sugerir problemas sistemáticos à nível estadual ou regional. Mais ainda, subestimativas e
352 sobreestimativas podem alternar num mesmo local de um mês para o seguinte. Os casos
353 extremos foram aqueles das estações Guaramiranga - CE (com diferenças de 63,9 mm para
354 janeiro), Cuiabá - MT (56,4 mm para fevereiro), Alto Parnaíba - MA (64,2 mm para março),
355 Fortaleza - CE (102,9 mm para abril), Turiaçu - MA (57,8 mm para maio), Recife - PE (com
356 diferenças de 74,1 mm para junho, 51,6 mm para julho, e 34,3 mm para agosto), São Joaquim
357 - SC (34,9 mm para setembro), Cruz Alta - RS (57,8 mm para outubro), Rio Verde - GO (45,6
358 mm para novembro), e Itumbiara - GO (62,6 mm para dezembro), sendo que em todas elas as
359 EMCs registraram precipitações maiores que as EMAs.

360 As temperaturas máximas médias mensais mostraram diferenças significativas nos
361 meses de dezembro até fevereiro, o que aproximadamente coincide com o verão austral, além
362 de setembro (Fig. 1.4B), quando algumas das EMAs registraram as menores temperaturas
363 máximas médias em comparação às EMCs (histogramas da Fig. 1.6). Novamente, diferenças
364 individuais podem ser encontradas ao longo de todo o ano (Fig. 1.6), ainda que de um modo

365 mais estável se comparado ao caso das precipitações: as EMAs que sobre ou subestimaram,
366 tenderam a fazer isso em todos os meses. Os casos extremos foram aqueles das estações
367 Guaramiranga - CE, aonde a automática registrou na média 3,1 °C menos que a convencional
368 para janeiro e fevereiro e 2,8 °C menos para março e, especialmente, Ibitité - MG, onde a
369 EMA registrou na média temperaturas máximas menores para abril (2,8 °C), maio (2,6 °C),
370 junho (2,3 °C), julho (2,4 °C), agosto e setembro (2,5 °C), outubro e novembro (2,9 °C), e
371 dezembro (2,8 °C).

372 As temperaturas mínimas médias não mostraram diferenças estatisticamente
373 significativas para nenhum mês (Fig. 1.4C), ainda que diferenças individuais foram detectadas
374 ao longo do ano (histogramas da Fig. 1.7), sem um padrão geográfico evidente, mas com uma
375 tendência a sobre ou subestimar as medidas em cada localidade foi ainda mais estável do que
376 no caso das temperaturas máximas médias. Por exemplo, no caso de Itaberaba (BA) a EMA
377 registrou temperaturas médias mínimas na média 2,5 °C (variação: 2,2 e 2,8 °C) superiores
378 que a EMC homônima, ao longo de todo o ano.

379

380 **3.4. Correlação com outras variáveis**

381

382 Em todos os casos, os valores de precipitação média mensal estiveram altamente
383 correlacionados, acima de 0,9, com os correspondentes valores de precipitação média mensal
384 fornecidos pelo TRMM (Tabela 1.2). Estas correlações não mostraram diferenças
385 estatisticamente significativas, exceto para outubro, quando a correlação das EMCs e a
386 precipitação média mensal do TRMM foi maior.

387 As temperaturas máximas médias mensais estiveram inversamente correlacionadas
388 com a elevação, sob -0,5 em todos os meses exceto no período desde maio até outubro,
389 coincidindo aproximadamente com o outono-inverno austral, sem diferenças estatisticamente
390 significativas entre EMCs e EMAs (Tabela 1.3). Entretanto, as temperaturas máximas médias
391 mensais mostraram uma correlação maior com a latitude, acima de 0,5 em todos os meses
392 exceto durante janeiro-março, ou seja, durante o verão austral no caso das EMCs, e fevereiro
393 durante EMAs, com diferenças estatisticamente significativas entre EMCs e EMAs no
394 período de dezembro até abril, quando as EMAs mostraram uma correlação
395 significativamente maior.

396

Figura 1.4 – Comparação da informação fornecida pelas estações convencionais e automáticas homónimas em relação a: (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima. Na base de cada gráfico encontra-se o resumo do teste bicaudal pareado *t* de Student: (N.S.) não significativo; (*) $p < 0,05$; (**) $p < 0,01$; (***) $p < 0,001$.

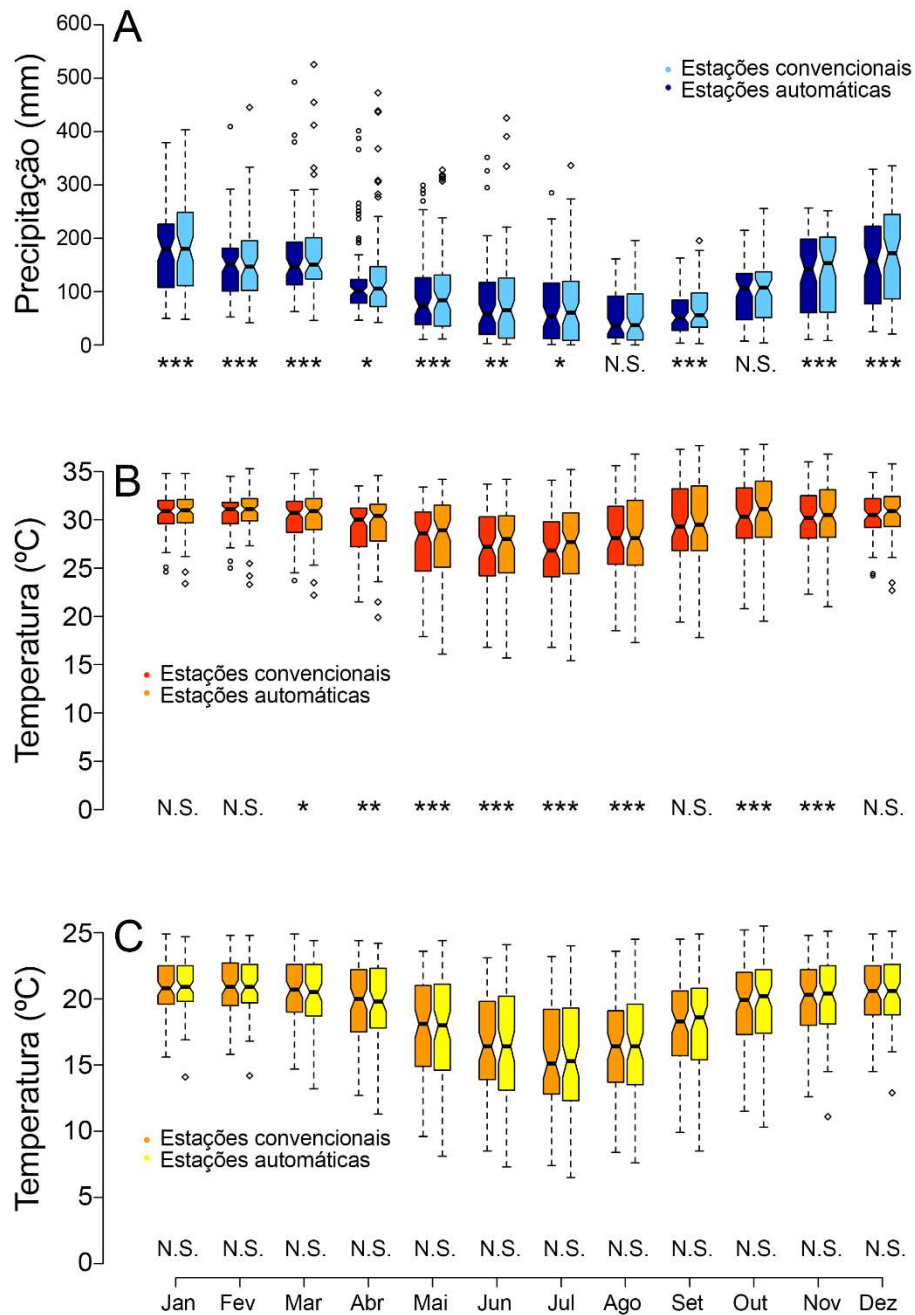


Figura 1.5 – Diferenças resultantes de subtrair a precipitação média mensal registrada por cada estação convencional menos a registrada pela automática homônima. Em azul aqueles casos em que a estação convencional $>$ estação automática; em vermelho quando estação convencional $<$ estação automática. Os histogramas mostram a frequências dessas diferenças.

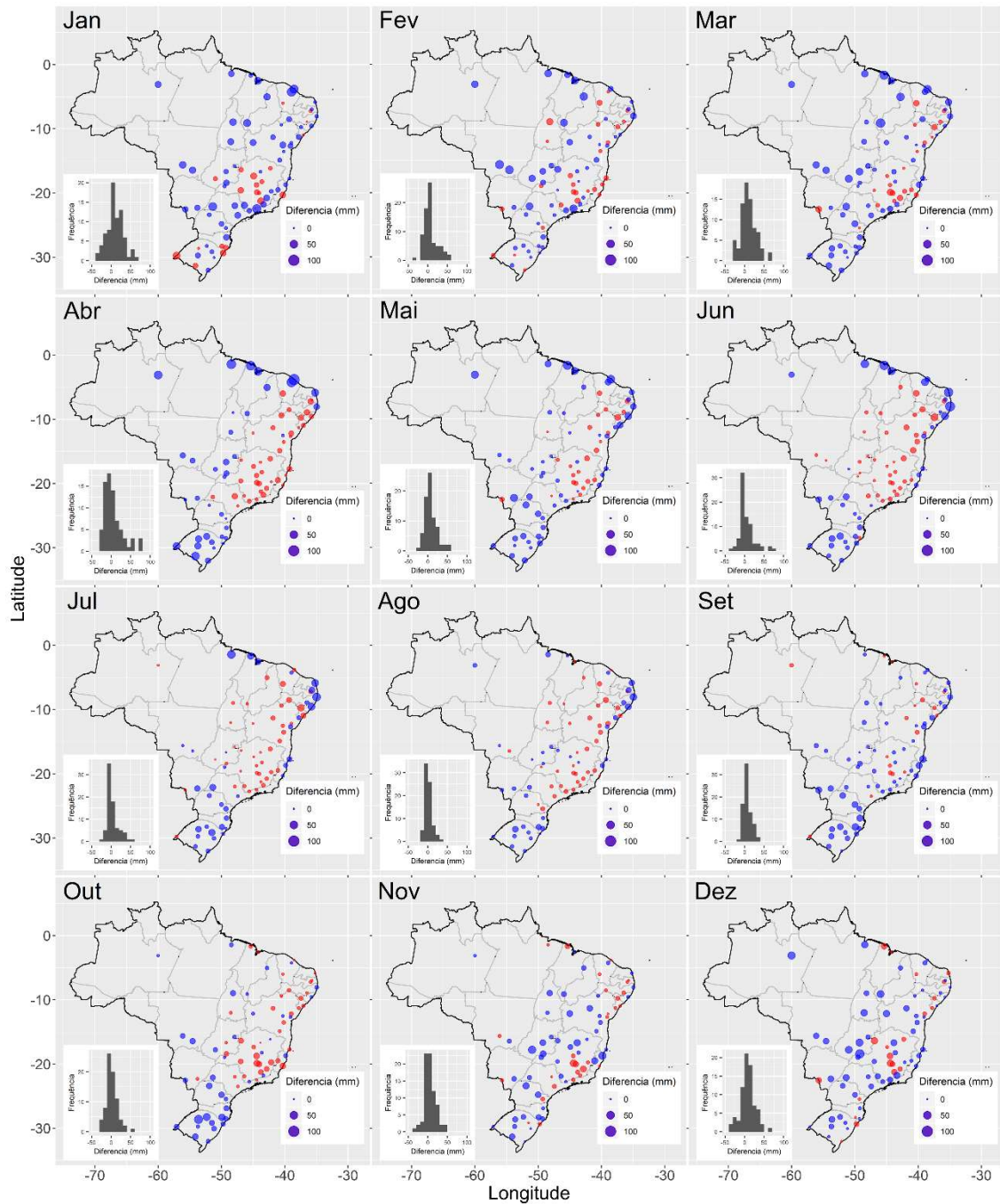
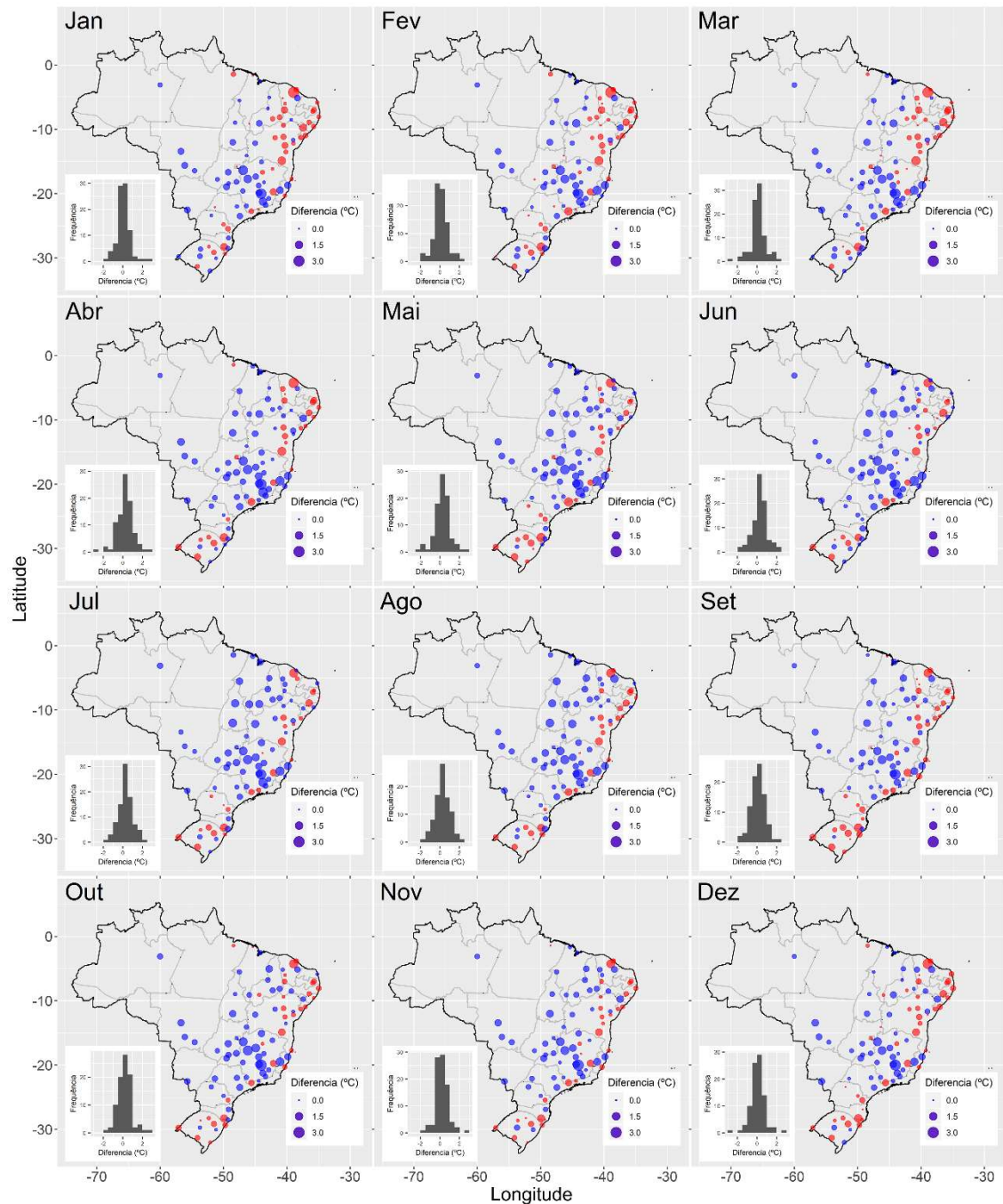


Figura 1.6 – Diferenças resultantes de subtrair a temperatura máxima média mensal registrada por cada estação convencional menos a registrada pela automática homônima. Em azul aqueles casos em que a estação convencional > estação automática; em vermelho quando estação convencional < estação automática. Os histogramas mostram a frequências dessas diferenças.



399

400

401

402

403

As temperaturas mínimas médias mensais mostraram correlações ainda mais fortes com a elevação, sob -0,6 na maioria dos casos, sem diferenças significativas entre EMCs e EMAs (Tabela 1.4). Complementarmente, as correlações das temperaturas mínimas médias

404 mensais estiveram altamente correlacionas com a latitude, acima de 0,6, no caso das EMCs e
405 0,8 no caso das EMAs, com diferenças significativas ao longo de todo o ano exceto durante
406 junho-agosto.

407

408 **3.5. Avaliação da consistência dos dados**

409

410 Ao comparar os valores das RRMSE resultantes das interpolações, os testes *t* de
411 Student não sugerem diferenças entre EMCs e EMAs no caso da precipitação (Fig. 1.8A). No
412 caso das temperaturas, esta análise sugere que, em geral, as EMAs tenderam mostrar um
413 melhor ajuste (maior concordância entre temperaturas interpoladas e observadas), com
414 diferenças estatisticamente significativas em fevereiro, no caso das temperaturas máximas, e
415 entre fevereiro e abril, no caso das temperaturas mínimas (Fig. 1.8B-C).

416

417 **4. DISCUSSÃO E CONCLUSÕES**

418

419 Do ponto de vista da abrangência geográfica, este é o maior estudo comparativo da
420 informação climática fornecida pelas EMCs e as EMAs brasileiras feito até o momento.
421 Cristaldo *et al.* (2017), concluíram que, no caso do Pantanal, embora que o número de
422 estações em operação seja satisfatório segundo o estabelecido pela WMO, a densidade da rede
423 não é satisfatória nas estações de operação por falta de manutenção, criando assim a
424 necessidade de estações adicionais. Esta conclusão pode ser extrapolada para o resto do
425 Brasil, especialmente considerando que muitas das EMAs ainda não estão fornecendo
426 informação concordante aos critérios da WMO.

427 Em geral, os resultados deste estudo mostram o que já tem sido relatado na literatura
428 para casos particulares: que o grau de semelhança da informação fornecida por ambos os tipos
429 de estações meteorológicas pode ser muito congruente. Efetivamente, enquanto alguns autores
430 já indicaram haver boa concordância entre a informação meteorológica proveniente de ambos
431 os tipos de estações (Souza *et al.*, 2003; Pereira *et al.*, 2008; Strassburger *et al.*, 2011; Funari
432 e Pereira, 2012; Almeida e Hermenegidio, 2013), outros apontaram discrepâncias, em
433 ocasiões marcadas, entre EMCs e EMAs tanto em outros países (Kuśmierk-Tomaszewska *et*
434 *al.*, 2012; Karatarakis *et al.*, 2013; Urban, 2015; Kaya, 2017) quanto a nível local no Brasil
435 (Pinto *et al.*, 2006; Oliveira *et al.*, 2010; Lucas *et al.*, 2010; Campos *et al.*, 2016; Ribeiro *et*
436 *al.*, 2017).

Figura 1.7 – Diferencias resultantes de subtrair a temperatura mínima média mensal registrada por cada estação convencional menos a registrada pela automática homônima. Em cinza escuro aqueles casos em que a estação convencional > estação automática; em cinza médio quando estação convencional < estação automática. Os histogramas mostram a frequências dessas diferencias.

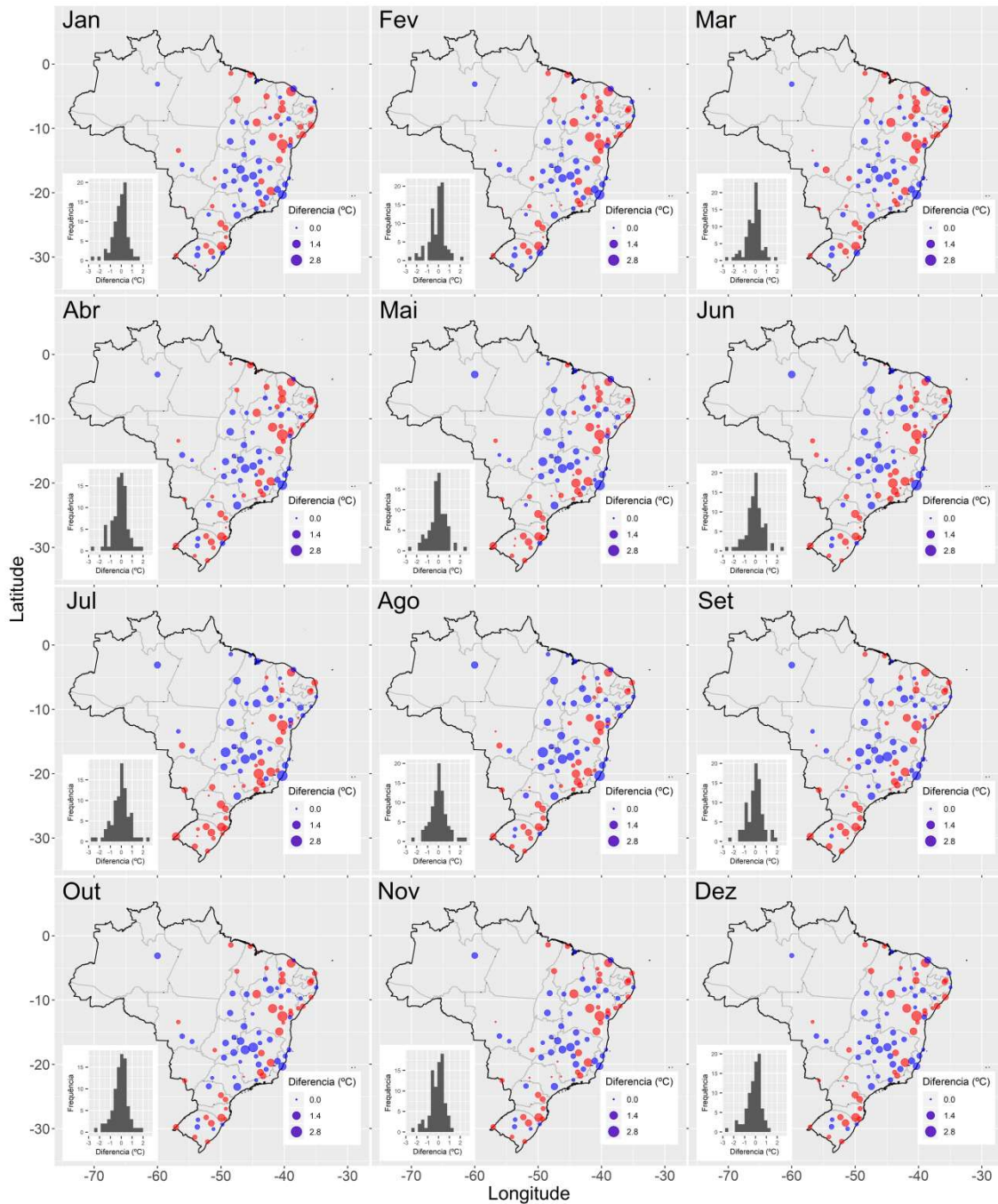


Tabela 1.2 – Correlação entre as precipitações médias mensais de cada tipo de estação climatológica e os respectivos valores de precipitação médias mensais fornecidos pelo Tropical Rainfall Measuring Mission 3B43: (r_c) correlação no caso das estações convencionais, com $p < 0,001$ para todos os casos; (r_a) correlação no caso das estações automáticas com $p < 0,001$ para todos os casos; (p_{c-a}) probabilidade resultante da comparação das correlações acima mencionadas através das transformações de Fisher (1915, 1921). Em negrito o valor estatisticamente significativo da comparação das correlações com $\alpha = 0,05$.

Mes	r_c	r_a	p_{c-a}
Jan	0,959	0,953	0,347
Fev	0,950	0,941	0,314
Mar	0,952	0,946	0,372
Abr	0,957	0,927	0,055
Mai	0,949	0,936	0,236
Jun	0,919	0,905	0,308
Jul	0,932	0,919	0,285
Ago	0,931	0,943	0,277
Set	0,969	0,961	0,261
Out	0,974	0,953	0,037
Nov	0,973	0,967	0,263
Dec	0,976	0,970	0,246

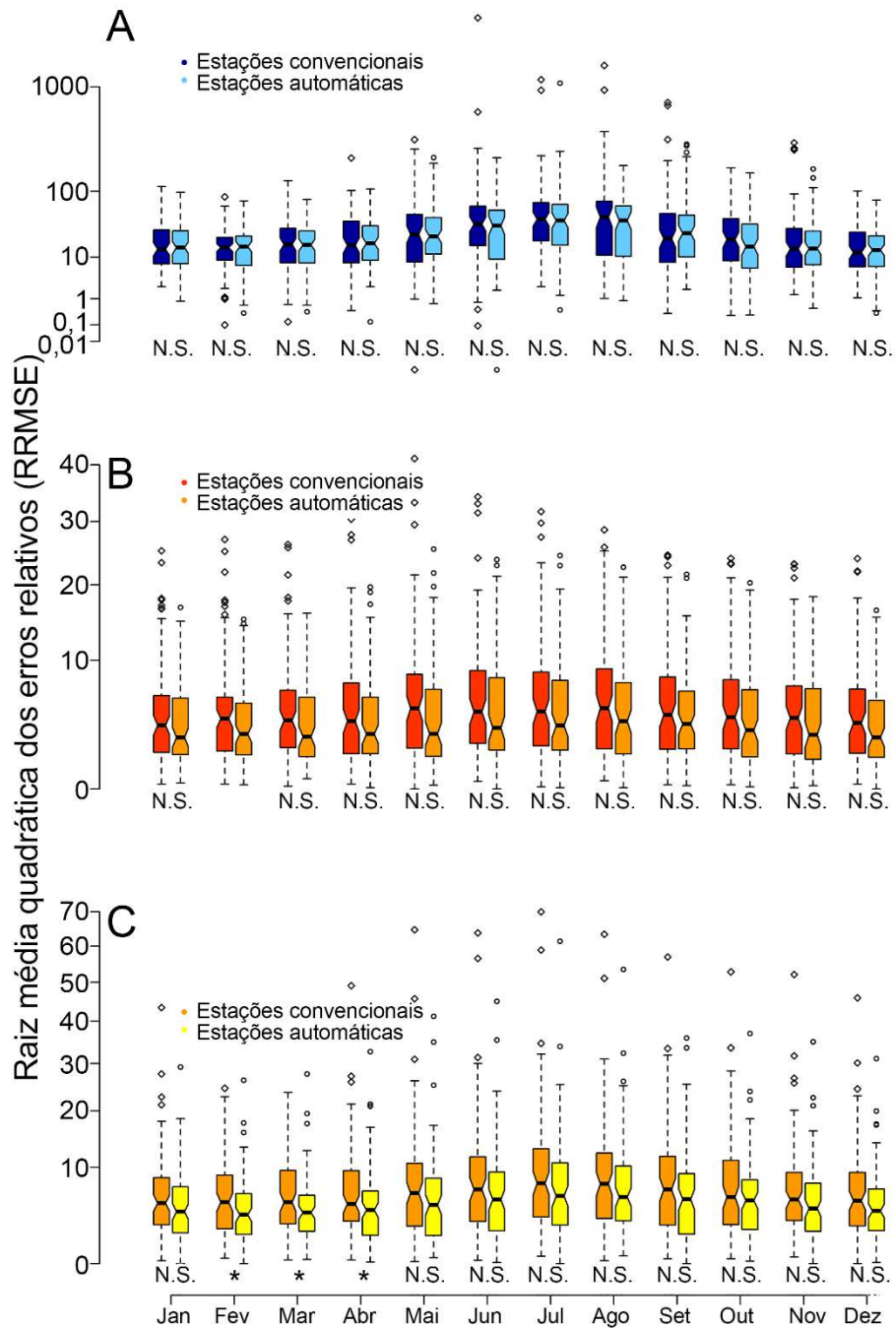
Tabela 1.3 – Correlações entre as temperaturas máximas mensais de cada tipo de estação climatológica, e as elevação e a latitude respectivas. (r_c) correlação no caso das estações convencionais; (r_a) correlação no caso das estações automáticas; (p_{c-a}) probabilidade resultante da comparação das correlações acima mencionadas através das transformações de Fisher (1915, 1921). Todas as correlações com $p < 0,001$.

Mes	Elevação			Latitude		
	r_c	r_a	p_{c-a}	r_c	r_a	p_{c-a}
Jan	-0,714	-0,733	0,398	0,353	0,570	0,035
Fev	-0,678	-0,695	0,417	0,224	0,490	0,023
Mar	-0,637	-0,629	0,466	0,406	0,633	0,020
Abr	-0,515	-0,569	0,310	0,591	0,754	0,024
Mai	-0,427	-0,459	0,398	0,769	0,851	0,058
Jun	-0,348	-0,340	0,477	0,753	0,834	0,076
Jul	-0,279	-0,276	0,492	0,735	0,807	0,123
Ago	-0,232	-0,204	0,424	0,656	0,743	0,133
Set	-0,151	-0,129	0,442	0,653	0,718	0,212
Out	-0,237	-0,241	0,489	0,653	0,751	0,103
Nov	-0,557	-0,592	0,367	0,689	0,789	0,074
Dec	-0,682	-0,727	0,281	0,561	0,717	0,042

Tabela 1.4 – Correlações parciais entre as temperaturas mínimas mensais de cada tipo de estação climatológica, e a elevação e a latitude correspondentes. (r_c) correlação no caso das estações convencionais; (p_c) probabilidade da correlação das estações convencionais; (r_a) correlação no caso das estações automáticas; (p_a) probabilidade da correlação das estações automáticas; (p_{c-a}) probabilidade resultante da comparação das correlações acima mencionadas através das transformações de Fisher (1915, 1921). Todas as correlações com $p < 0,001$.

Mes	Elevação			Latitude		
	r_c	r_a	p_{c-a}	r_c	r_a	p_{c-a}
Jan	-0,844	-0,893	0,101	0,681	0,851	0,003
Fev	-0,854	-0,854	0,133	0,686	0,872	0,001
Mar	-0,793	-0,835	0,215	0,769	0,899	0,002
Abr	-0,740	-0,792	0,212	0,858	0,934	0,005
Mai	-0,783	-0,807	0,440	0,937	0,961	0,060
Jun	-0,739	-0,725	0,424	0,933	0,954	0,111
Jul	-0,699	-0,687	0,442	0,932	0,947	0,208
Ago	-0,662	-0,687	0,386	0,911	0,945	0,057
Set	-0,558	-0,603	0,334	0,873	0,925	0,040
Out	-0,562	-0,621	0,283	0,823	0,911	0,010
Nov	-0,685	-0,752	0,190	0,835	0,919	0,008
Dec	-0,785	-0,848	0,114	0,781	0,895	0,005

Figura 1.8 – Comparação entre as estações convencionais e as automáticas homônimas dos valores mensais da raiz média quadrática dos erros relativos (RRMSE pelas suas siglas no inglês), calculados com os valores observados e interpolados para cada estação. (A) precipitação; (B) temperatura máxima; e (C) temperatura mínima. Na base de cada gráfica encontra-se o resumo do teste bicaudal pareado *t* de Student: (N.S.) não significativo; (*) $p < 0,05$; (**) $p < 0,01$; (***) $p < 0,001$. Os eixos das ordenadas estão em escada logarítmica.



441

442

443 Similarmente, neste estudo, as distribuições centradas em zero e marcadamente
444 leptocúrticas dos histogramas correspondentes as diferenças simples (calculadas como os
445 valores das EMCs menos os valores das EMAs), representados nas Figs. 1.5-7, além de
446 comparações resumidas nos boxplots das Figs. 1.1 e 1.8, indicam boa concordância na
447 informação climática fornecida por ambos os tipos de estações. Mas também encontramos
448 evidências sólidas de divergências, que em alguns casos podem ser dramáticas. De acordo
449 com Linacre (2005), a existência destas incongruências é própria de toda medição
450 meteorológica e pode ter diferentes origens. Primeiro, a forma de observação, sendo um
451 exemplo o fator humano, ou os erros tanto na leitura das medidas oferecidas pelos
452 instrumentos como na sua transcrição por parte dos observadores no caso das EMCs.
453 Segundo, os erros dos próprios dos instrumentos, os quais podem ser tanto sistemáticos como
454 ao acaso. Os erros sistemáticos (ou viés) podem ser consequência de problemas na calibração
455 ou na velocidade de resposta do instrumento às mudanças rápidas do elemento meteorológico
456 medido: as EMCs estão normalmente dotadas de sensores analógicos cuja velocidade de
457 resposta é menor que aquela dos sensores eletrônicos das EMAs. Estas diferenças sistemáticas
458 tendem ser consistentes e podem ser compensadas aplicando fatores de correção. Os erros ao
459 acaso são produzidos por defeitos nos instrumentos, ou pelo efeito combinado de fatores
460 seguindo um padrão quase aleatório. Por exemplo, Sevruk e Chvíla (2005) encontraram erros
461 na medição da precipitação por parte das EMAs na Eslováquia, especialmente quando as
462 quantidades de precipitação líquida foram sob os 0,05 mm, quando os intervalos de medição
463 foram maiores que três minutos, quando as temperaturas estavam acima dos 15 °C, e por
464 problemas do software usado para corrigir o efeito das batidas pelo vento, vibrações e
465 mudanças súbitas do peso no detector.

466 Fatores semelhantes poderiam explicar o desempenho ligeiramente melhor das EMCs
467 na medição da precipitação, segundo os indicadores considerados neste estudo. Terceiro, o
468 local exato da estação meteorológica, ou "exposição", tem que garantir a representatividade
469 do elemento meteorológico que está sendo medido na área onde a estação foi instalada, já que
470 mudanças menores nas condições exatas do local das estações podem resultar em
471 discrepâncias entre as medições. E, finalmente, as diferenças no produto da amostragem
472 também podem resultar em divergências nas medições, que no presente estudo podem incluir
473 diferenças na quantidade e sobreposição temporal das séries de dados usados.

474 Aqueles casos em que as EMAs consistentemente registraram médias mensais
475 substancialmente diferentes do que as EMCs, merecem estudos mais profundos a fim de
476 detectar as causas, que neste caso há existência de uns erros sistemáticos que poderiam ser

477 facilmente retificáveis através da aplicação de algum fator de correção. E essa correção é
478 necessária, porque tais diferenças poderiam ser interpretadas como mudanças decorrentes de
479 variações climáticas que, de fato, não ocorreram (Lucas *et al.*, 2010), e que poderiam ter
480 impactos substanciais em políticas públicas, planos de manejo, etc. Por exemplo, voltando
481 para o caso das estações localizadas na Itaberaba (BA), aonde a EMA registrou temperaturas
482 médias mínimas de 2,5 °C acima da EMC homônima, estes registros poderiam erroneamente
483 indicar um incremento substancial das temperaturas, provavelmente acima das metas
484 propostas no Acordo de Paris, que propôs reduzir as emissões de gases de efeito estufa para
485 manter o aumento da temperatura global bem abaixo de 2,0 °C e buscar esforços para limitar
486 o aquecimento a 1,5 °C acima dos níveis pré-industriais (CQNUMC, 2015). Infelizmente,
487 estas correções em muitos casos vão se limitar para aquelas localidades onde existam os dois
488 tipos de estações, pelo que seria desejável que a instalação das EMAs passe por um período
489 de calibração com EMCs, no mesmo local, seguido de testes de calibração periódicos.

490 Não obstante, a maioria dos erros detectados são aleatórios, com algumas EMAs em
491 alguns meses subestimando, e outros sobrestimando, os valores das EMCs. A origem desses
492 erros é mais difícil de determinar e ainda mais difícil de corrigir. Por exemplo, no caso da
493 precipitação estas discrepâncias podem estar associadas à intensidade das chuvas, flutuações
494 sazonais e lacunas nos registros para determinados períodos, ou ao tipo de medidores
495 utilizados e as suas características, tais como a área de captação e facilidade de entupimento
496 (Tanner, 1990; Torre, 1995; Lucas *et al.*, 2010), além do fato que grande parte da precipitação
497 no Brasil é de origem convectiva, ou seja, localizada (Reboita *et al.*, 2010) e que o TRMM faz
498 a avaliação em áreas e não em pontos como os pluviômetros em superfície.

499 Assim, o Brasil conta com dois conjuntos de dados climáticos cada um dos quais tem
500 alguns problemas. Por exemplo, as EMCs são menos numerosas, as suas coordenadas
501 geográficas são menos exatas e precisas (veja o problema que já foi mencionado por Vianna
502 *et al.* (2017) para Santa Catarina), e provavelmente, tem erros provenientes do fator humano.
503 As EMAs são mais numerosas, mas muitas ainda não fornecem informação congruente aos
504 critérios da WMO, e podem estar afetadas por problemas na calibração dos instrumentos, e
505 muitas delas estão instaladas em locais novos, onde não existe informação prévia que possa
506 servir de base de comparação para quantificar a confiabilidade da informação fornecida.
507 Assim, é natural que surge a pergunta: Diante de um cenário de caracterização climática em
508 escala macro, qual fonte de informação tem que ser usada preferencialmente?

509 Mesmo sendo um procedimento que tem recebido críticas (Stegenga e Menon, 2017),
510 é comum considerar que quando uma variedade de evidências independentes sustenta uma

511 hipótese, essa hipótese é mais provável que seja verdadeira, o que às vezes chamamos de
512 "robustez". Isso pode ser usado como apoio para várias formas de distinguir artefatos de
513 entidades reais para nos ajudar na nossa procura de objetividade. Assim, as recorrentes
514 correlações maiores entre precipitação e TRMM (apesar do grau de circularidade, pois no
515 TRMM as estações são consideradas como dados de referência), no caso das EMCs, que
516 coincide com comparações mais locais feitas no Brasil (Collischonn *et al.*, 2007), e das
517 temperaturas com a latitude no caso das EMAs, sugere que a informação de cada tipo de
518 estação tem um grau de confiabilidade um pouco maior segundo a parâmetro a serem
519 explorado. Isso, apesar dos problemas já comentados sobre as coordenadas geográficas e que
520 vários autores afirmarem que em alguns casos o TRMM pode sobrestimar a precipitação no
521 Brasil (Franchito *et al.*, 2009; Pereira *et al.*, 2013). Uma possível explicação é que a
522 topografia do Brasil não é tão abrupta, por conseguinte carente daqueles gradientes climáticos
523 altamente contrastantes como pode ser observar por exemplo nos Andes (Ramon Perazzi,
524 2016), e as imprecisões nas localizações das estações não tem um impacto significativo.
525 Ainda assim, todos os dados, tanto das EMCs como das EMAs, exigem um controle de
526 qualidade prévio para serem usados.

527 Recomenda-se realizar análises similares no futuro, quando as EMAs tenham mais
528 tempo operativas e incrementando o número de variáveis a serem consideradas.
529 Adicionalmente, o Brasil conta com um conjunto ainda maior de estações meteorológicas,
530 como indicado no portal do sistema de monitoramento agrometeorológico, muitas das quais
531 estão sob a responsabilidade de outras instituições diferentes ao INMET, e inclusive por parte
532 de particulares. Tal heterogeneidade, potencialmente acrescenta outras fontes de discrepância
533 nas medições, e futuros esforços teriam que incluir a avaliação destas estações também.

534

535 **Agradecimentos**

536

537 Os dados das estações automáticas foram gentilmente fornecidos pela Seção de
538 Armazenamento de Dados Meteorológicos do INMET. PRP agradece a bolsa da Organização
539 dos Estados Americanos por meio de seu Programa de Alianças para a Educação e a
540 Capacitação (Bolsas Brasil - PAEC OEA-GCUB).

541

542 **Bibliografia**

543

- 544 AHMAD, L.; HABIB KANTH, R.; PARVAZE, S.; SHERAZ MAHDI, S. Automatic weather station. In:
545 **Experimental Agrometeorology: A Practical Manual**. Cham: Springer, p. 83-87, 2017.
- 546 ALMEIDA, H.A. DE.; HERMENEGIDIO, G.M. DOS S. Comparação de dados meteorológicos obtidos por
547 estações meteorológicas convencional e automática. **Revista Brasileira de Climatologia**, v. 12, p. 33-47, 2013.
- 548 ARVOR, D.; DUBREUIL, V.; RONCHAIL, J.; SIMÕES, M.; FUNATSU, B.M. Spatial patterns of rainfall
549 regimes related to levels of double cropping agriculture systems in Mato Grosso (Brazil). **International Journal
550 of Climatology**, v. 34, n. 8, p. 2622-2633, 2014.
- 551 BAGHERI, F. Mapping drought hazard using SPI index and GIS (A case study: Fars province, Iran).
552 **International Journal of Environment and Geoinformatics**, v. 3, n. 1, p. 22-28, 2016.
- 553 BROCKWELL, P.J.; DAVIS, R.A. **Time Series: Theory and Methods**. Nova York: Springer-Verlag, p. 580,
554 1991.
- 555 CAMPOS, R.B.; SAKIYAMA, N.R.M.; OLIVEIRA, C.C. Análise de dados climáticos das estações
556 convencional e automática em Teófilo Otoni - MG - Estratégias bioclimáticas para a construção civil. **Revista
557 Científica Multidisciplinar Núcleo do conhecimento**, v. 1, n. 9, p. 914-937, 2016.
- 558 CAO, Y.; ZHANG, W.; WANG, W. Evaluation of TRMM 3B43 data over the Yangtze River Delta of China.
559 **Scientific Reports**, v. 8, p. 5290, 2018.
- 560 CARREGA, P. Basics of climatological and meteorological observations for GIS applications. In: **Geographical
561 Information and Climatology**. Londres: ISTE Group, p. 1-27, 2013.
- 562 CASHION, J.; LAKSHMI, V.; BOSCH, D.; JACKSON, T.J. Microwave remote sensing of soil moisture:
563 evaluation of the TRMM microwave imager (TMI) satellite for the Little River Watershed Tifton, Georgia.
564 **Journal of Hydrology**, v. 307, n. 1-4, p. 242-253, 2005.
- 565 CHAMBERS, J.M.; CLEVELAND, W.S.; KLEINER, B.; TUKEY, P.A. **Graphical Methods for Data
566 Analysis**. Boca Ratón: Chapman and Hall/Cole Publishing Company, p. 395, 1983.
- 567 CHEN, F.W.; LIU, C.W. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in
568 the Middle of Taiwan. **Paddy and Water Environment**, v. 10, p. 209-222, 2012.
- 569 COLLIER, C.G. **Hydrometeorology**. Oxford: John Wiley & Sons, Ltd, p. 376, 2016.
- 570 COLLISCHONN, B.; ALLASIA, D.; COLLISCHONN, W.; TUCCI, C.E.M. Desempenho do satélite TRMM na
571 estimativa de precipitação sobre a Bacia do Paraguai Superior. **Revista Brasileira de Cartografia**, v. 59, n. 1, p.
572 93-99, 2007.
- 573 CRISTALDO, M.F.; SOUZA, C.C. DE; LEANDRO, J. DE; PADOVANI, C.R. *et al.* Analysis and distribution
574 of the rainfall monitoring network in a Brazilian Pantanal region. **Revista Brasileira de Meteorologia**, v. 32, n.
575 2, p. 199-205, 2017.
- 576 CQNUMC - Convenção-Quadro das Nações Unidas sobre a Mudança do Clima. **Adoption of the Paris
577 Agreement**. Paris: Conference of the Parties, p. 32, 2015.
- 578 DALY, C.; HALBLEIB, M.; SMITH, J.I.; GIBSON, W.P.; DOGGETT, M.K. *et al.* Physiographically sensitive
579 mapping of climatological temperature and precipitation across the conterminous United States. **International
580 Journal of Climatology**, v. 28, p. 2031-2064, 2008.
- 581 DALY, C.; NEILSON, R.P.; PHILLIPS, D.L. A statistical-topographic model for mapping climatological
582 precipitation over mountainous terrain. **Journal of Applied Meteorology**, v. 33, n. 2, p. 140-158, 1994.
- 583 DANIELSON, J.J.; GESCH, D.B. **Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010)**.
584 Sioux Falls: U.S. Geological Survey Earth Resources Observation and Science, p. 26, 2011.

- 585 DAS, H.P.; ADAMENKO, T.I.; ANAMAN, K.A.; GOMMES, R.G.; JOHNSON, G. **Agrometeorology Related**
586 **to Extreme Events**. Geneva: Secretariat of the World Meteorological Organization No. 943, p.138, 2003.
- 587 DINKU, T.; CECCATO, P.; GROVER-KOPEC, E.; LEMMA, M.; CONNOR, S.J. *et al.* Validation of satellite
588 rainfall products over East Africa's complex topography. **International Journal of Remote Sensing**, v. 28, n. 7,
589 p. 1503-1526, 2007.
- 590 EFRON, B. **The Jackknife, the Bootstrap and Other Resampling Plans**. **CBMS-NSF Regional Conference**
591 **Series in Applied Mathematics**. Philadelphia: Society for Industrial and Applied Mathematics, p. 100, 1982.
- 592 FISHER, R.A. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely
593 large population. **Biometrika**, v.10, n. 4, p. 507-521, 1915.
- 594 FISHER, R.A. On the 'probable error' of a coefficient of correlation deduced from a small sample. **Metron**, v. 1,
595 p. 3-32, 1921.
- 596 FRANCHITO, S.H.; RAO, V.B.; VASQUES, A.C.; SANTO, C.M.E.; CONFORTE, J.C. Validation of TRMM
597 precipitation radar monthly rainfall estimates over Brazil. **Journal of Geophysical Research: Atmospheres**, v.
598 114, n. D2, p. 1-9, 2009.
- 599 FUNARI, F.L.; PEREIRA, A.J. FILHO. 2012. Análise comparativa de medições de variáveis meteorológicas
600 realizadas por estações meteorológicas convencional e automática instaladas no Parque Estadual das Fontes do
601 Ipiranga - São Paulo -SP. **Revista Brasileira de Climatologia**, v. 7, p. 159-168, 2010.
- 602 GRÄLER, B.; PEBESMA, E.; HEUVELINK, G. Spatio-temporal interpolation using gstat. **The R Journal**, v. 8,
603 n. 1, p. 204-218, 2016.
- 604 GRANGER, O.E. Precipitation distribution. In: OLIVER, J.E. **Encyclopedia of World Climatology**. Dordrecht:
605 Springer Science & Business Media, p. 576-582, 2008.
- 606 HARTMANN, D. L. **Global Physical Climatology**. Amsterdam: Elsevier Inc., p. 498, 2016.
- 607 IBGE - Instituto Brasileiro de Geografia e Estatística. **Projeto Mudança do Referencial Geodésico - PMRG**.
608 Disponível em: <https://www.ibge.gov.br/geociencias/informacoes-sobre-posicionamento-geodesico/sirgas/16691-projeto-mudanca-do-referencial-geodesico-pmrg.html?=&t=o-que-e#3>. Acessado em: 29
609 mar. 2020, 2005.
610
- 611 IBGE - Instituto Brasileiro de Geografia e Estatística. **Malha Municipal**. Disponível em:
612 <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15774-malhas.html>. Acessado
613 em: 20 mar. 2020, 2018.
- 614 JAVANMARD, S.; YATAGAI, A.; NODZU, M.; BODAGHJAMALI, J.; KAWAMOTO, H. Comparing high-
615 resolution gridded precipitation data with satellite rainfall estimates of TRMM_3B42 over Iran. **Advances in**
616 **Geosciences**, v. 25, p. 119-125, 2010.
- 617 JOYCE, R.J.; XIE, P. Kalman filter-based CMORPH. **Journal of Hydrometeorology**, v.12, n. 6, p. 1547-1563,
618 2011.
- 619 KÁLMÁN, R. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, v. 82,
620 n. 1, p. 35-45, 1960.
- 621 KARATARAKIS, N.; SARANTOPOULOS, A.; CHARANTONIS, T.; NASTOS, P.T.; LIANOS, K. *et al.* A
622 first comparative analysis of temperature data collected from automatic and conventional weather stations in
623 Greece. In: HELMIS C.; NASTOS P. **Advances in Meteorology, Climatology and Atmospheric Physics**.
624 Berlin: Springer, p. 169-174, 2013.
- 625 KAYA, S. Konvansiyonel ve otomatik meteoroloji istasyon verilerinden hesaplanan referans evapotranspirasyon
626 değerlerinin kıyaslanması. **Türk Doğa ve Fen Dergisi**, v. 6, n. 2, p. 83-87, 2017.

- 627 KIM, S. ppcor: An R Package for a fast calculation to semi-partial correlation coefficients. **Communications for**
628 **Statistical Applications and Methods**, v. 22, n. 6, p. 665-674, 2015.
- 629 KÖPPEN, W. Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur
630 Pflanzenwelt. **Geographische Zeitschrift**, v. 6, p. 593-611, 1900.
- 631 KRESSE, W.; DANKO D.M. **Springer Handbook of Geographic Information**. Dordrecht: Springer, p. 619,
632 2012.
- 633 KUMMEROW, C.; SIMPSON, J.; THIELE, O.; BARNES, W.; CHANG, A.T.C. *et al.* The status of the
634 Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. **Journal of Applied Meteorology**, v. 39,
635 n. 12, p. 1965-1982, 2000.
- 636 KUŚMIEREK-TOMASZEWSKA, R.; ŻARSKI, J.; DUDEK, S. Meteorological automated weather station data
637 application for plant water requirements estimation. **Computers and Electronics in Agriculture**, v. 88, p. 44-
638 51, 2012.
- 639 LEVIZZANI, V. Satellite rainfall estimates: a look back and a perspective. In: PROCEEDINGS OF THE
640 EUMETSAT METEOROLOGICAL SATELLITE DATA USERS CONFERENCE, Italia. **Resumos**
641 **Expandidos**: Bolonha, European Organisation for the Exploitation of Meteorological Satellites, 2000.
- 642 LI, J.; HEAP, A.D. A review of comparative studies of spatial interpolation methods in environmental sciences:
643 performance and impact factors. **Ecological Informatics**, v. 6, n. 3-4, p. 228-241, 2011.
- 644 LINACRE, E. **Climate Data and Resources: A Reference and Guide**. Londres: Taylor & Francis, p. 366,
645 2005.
- 646 LUCAS, E.W.M.; RODRIGUES, J.E.; REZENDE, J.M.; ABREU, S.F.; BRAGA A.P. Comparative analysis of
647 observed meteorological data in the conventional and surface automatic station at Brazilian National Institute of
648 Meteorology. In: WMO TECHNICAL CONFERENCE ON METEOROLOGICAL AND ENVIRONMENTAL
649 INSTRUMENTS AND METHODS OF OBSERVATION, Helsinki. **Resumos Expandidos**. Helsinki:
650 Organização Meteorológica Mundial, 2010.
- 651 MANTAS, V.; LIU, Z.; CARO, C.; PEREIRA, A. Validation of TRMM multi-satellite precipitation analysis
652 (TMPA) products in the Peruvian Andes. **Atmospheric Research**, v. 163, p. 132-145, 2015.
- 653 MENG, J.; LI, L.; HAO, Z.; WANG, J.; SHAO, Q. Suitability of TRMM satellite rainfall in driving a distributed
654 hydrological model in the source region of Yellow River. **Journal of Hydrology**, v. 509, p. 320-332, 2014.
- 655 MORITZ, S.; BARTZ-BEIELSTEIN, T. imputeTS: time series missing value imputation in R. **The R Journal**,
656 v. 9, n. 1, p. 207-218, 2017.
- 657 OLIVEIRA, A.D. DE; ALMEIDA, B.M. DE; CAVALCANTE, E.G.; SOBRINHO, J.E.; VIEIRA, R.Y.M.
658 Comparação de dados meteorológicos obtidos por estação convencional e automática em Jaboticabal - SP.
659 **Revista Caatinga**, v. 23, n. 4, p. 108-114, 2010.
- 660 PEBESMA, E.J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, n. 7, p.
661 683-691, 2004.
- 662 PEREIRA, G.; SILVA, M.E.S.; MORAES, E.C.; CARDOZO, F.S. Avaliação dos dados de precipitação
663 estimados pelo satélite TRMM para o Brasil. **Revista Brasileira de Recursos Hídricos**, v. 18, n. 3, p. 139-148,
664 2013.
- 665 PEREIRA, L.M.P.; CARAMORI, P.H.; RICCE W. DA S.; CAVIGLIONE, J.H. Análise comparativa de dados
666 meteorológicos obtidos por estação convencional e automática em Londrina - PR. **Semina: Ciências Agrárias**,
667 v. 29, n. 2, p. 299-305, 2008.
- 668 PINTO, H.S.; PELLEGRINO, G.Q.; FONSECHI, D.B.; CORAL, G.; CARAMORI, P.H. *et al.* Comparison
669 between daily meteorological data collected by automatic and conventional stations. In: INTERNATIONAL

- 670 CONFERENCE ON EXPERIENCES WITH AUTOMATIC WEATHER STATIONS, 4, 2006, Lisboa.
671 **Resumos Expandidos**. Lisboa: Instituto de Meteorologia, p. 1-9, 2006.
- 672 PLANTON, S. **Annex III. Glossary: IPCC - Intergovernmental Panel on Climate Change**. Cambridge:
673 IPCC Fifth Assessment Report, p.1465, 2013.
- 674 R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Versão 3.6.1. Vienna: R
675 Foundation for Statistical Computing, 2019.
- 676 Raju, N. J. **Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the**
677 **Environment**. Cham: Springer, p. 969, 2016.
- 678 RAMONI PERAZZI, P. **Modelado de la Distribución Espacial de la Precipitación Media Mensual de los**
679 **Andes Venezolanos**. Mérida. Tese (Mestrado em modelagem e simulação) - Facultad de Ingeniería, Universidad
680 de Los Andes, p. 76, 2016.
- 681 RAMONI-PERAZZI, P.; ABLAN-BORTONE, M.; THIELEN, D.R. A geographic analysis on the mean
682 monthly precipitation information available for the Venezuelan Andes. **Revista Geográfica Venezolana**, v. 58,
683 n. 1, p. 86-10, 2016.
- 684 REBOITA, M.S.; GAN, M.A.; ROCHA, R.P. DA; AMBRIZZI, T. Regimes de precipitação na América do Sul:
685 uma revisão bibliográfica. **Revista Brasileira de Meteorologia**, v.25, n. 2, p.185-204, 2010.
- 686 RIBEIRO, A. DE A.; ANDRADE, A.S. DE; SILVA, E.M. DA; SIMEÃO, M.; BASTOS, E.A. Comparação
687 entre dados meteorológicos obtidos por estações convencionais e automáticas no Estado do Piauí, Brasil. **Irriga**,
688 v. 22, n. 2, p. 220-235, 2017.
- 689 SAHOO, A.K.; SHEFFIELD, J.; PAN, M.; WOOD, E.F. Evaluation of the tropical rainfall measuring mission
690 multi-satellite precipitation analysis (TMPA) for assessment of large-scale meteorological drought. **Remote**
691 **Sensing of Environment**, v. 159, p. 181-193, 2015.
- 692 SEVRUK, B.; CHVÍLA, B. Error sources of precipitation measurements using electronic weight systems.
693 **Atmospheric Research**, v. 77, n. 1-4, p. 39-47, 2005.
- 694 SHEPARD, D. A two-dimensional interpolation function for irregularly-spaced data. In: PROCEEDINGS OF
695 THE 1968 ACM NATIONAL CONFERENCE, Nova York. **Resumos Expandidos**. Nova York:
696 Association for Computing Machinery, p. 517-524, 1968.
- 697 SHRIVASTAVA, R.; DASH, S.; HEGDE, M.; PRADEEPKUMAR, K.; SHARMA, D. Validation of the
698 TRMM Multi Satellite Rainfall Product 3B42 and estimation of scavenging coefficients for 131 I and 137 Cs
699 using TRMM 3B42 rainfall data. **Journal of Environmental Radioactivity**, v. 138, p. 132-136, 2014.
- 700 SIVAGAMI, M.; VAISHALI, A.; RAMAKRISHNAN, R.; SUBASINI, A. Weather prediction model using
701 Savitzky-Golay and Kalman Filters. **Procedia Computer Science**, v. 165, p. 449-455, 2019.
- 702 SOUZA, I. DE A.; GALVANI, E.; ASSUNÇÃO, H.F. DA. Estudo comparativo entre elementos meteorológicos
703 monitorados por estações convencional e automática na região de Maringá, Estado do Paraná. **Acta Scientiarum**
704 **Technology**, v. 25, n. 2, p. 203-207, 2003.
- 705 STEGENGA, J.; MENON, T. Robustness and independent evidence. **Philosophy of Science**, v. 84, n. 3, p. 414-
706 435, 2017.
- 707 STRASSBURGER, A.S.; MENEZES, A.J.E.A. DE; PERLEBERG, T.D.; EICHOLZ, E.D.; MENDEZ, M.E.G.
708 *et al.* Comparação da temperatura do ar obtida por estação meteorológica convencional e automática. **Revista**
709 **Brasileira de Meteorologia**, v. 26, n. 2, p. 273-278, 2011.
- 710 TANNER, B.D. Automated weather station. **Remote Sensing Reviews**, v. 5, n. 1, p. 73-98, 1990.

- 711 TORRE, A. NETO. **Estudo de Implementação de um Sistema de Monitoramento Remoto de Variáveis**
 712 **Edafo-Ambientais**. São Carlos, Tese (Doutorado) - Instituto de Física e Química, Universidade de São Carlos.
 713 p. 146, 1995.
- 714 TRAPASSO, L.M. Temperature distribution. In: OLIVER, J.E. **Encyclopedia of World Climatology**.
 715 Dordrecht: Springer Science & Business Media, p. 711-715, 2008.
- 716 URBAN, G. An attempt to assess the results of air temperature measurements from automatic weather stations in
 717 comparison to glass thermometer measurements in the context of weather types, **Bulletin of Geography.**
 718 **Physical Geography Series**, v. 9, n. 1, p. 67-79, 2015.
- 719 VIANNA, L.F. DE N.; PERIN, E.B.; RICCE, W. DA S.; MASSIGNAN, A.M.; PANDOLFO, C. Bancos de
 720 dados meteorológicos: análise dos metadados das estações meteorológicas no Estado de Santa Catarina, Brasil.
 721 **Revista Brasileira de Meteorologia**, v. 32, n. 1, p. 53-64, 2017.
- 722 WMO - WORLD METEOROLOGICAL ORGANIZATION. **Calculation of Monthly and Annual 30-Year**
 723 **Standard Normals**. Genebra: WMO - Organização Meteorológica Mundial, n. 341, 12 p., 1989.
- 724 WMO - WORLD METEOROLOGICAL ORGANIZATION. **Guide to Climatological Practices**. Genebra:
 725 WMO - Organização Meteorológica Mundial, n. 100, p. 12, 2011.
- 726 WMO - WORLD METEOROLOGICAL ORGANIZATION. **Guide to Meteorological Instruments and**
 727 **Methods of Observation**. Genebra: Organização Meteorológica Mundial, n. 8, p. 548, 2018.
- 728 ZHOU, C.; SU, F.; HARVEY, F.; XU, J. **Spatial Data Handling in Big Data Era**. Singapura: Select papers
 729 from the 17th IGU Spatial Data Handling Symposium, p.252, 2017.

730

731 **7. Material Suplementar**

732

733 Tabela S1. Lista das estações meteorológicas do Instituto Nacional de Meteorologia
 734 analisadas neste estudo.

735

736 **8. Endereços de internet**

737

738 IBGE:

739 [ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/mun](ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2018/Brasil/BR/br_municipios.zip)
 740 [icipio_2018/Brasil/BR/br_municipios.zip](ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2018/Brasil/BR/br_municipios.zip)

741 INMET: <http://www.inmet.gov.br/>742 <http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesConvencionais>743 Monitoramento agrometeorológico: <https://www.agritempo.gov.br/>

744 Tropical Rainfall Measuring Mission TRMM: [https://gpm.nasa.gov/data-](https://gpm.nasa.gov/data-access/downloads/trmm)
 745 [access/downloads/trmm](https://gpm.nasa.gov/data-access/downloads/trmm)

1 **ARTIGO 2 – BRAZILCLIM: THE OVERCOMING OF LIMITATIONS OF**
 2 **PREEXISTING BIOCLIMATE DATA**

3
 4 Short title: **BrazilClim**

5
 6 Paolo Ramoni-Perazzi^{1,2}, Marcelo Passamani¹, Dirk Thielen³, Carlos Padovani⁴ & Marco
 7 Aurelio Arizapana⁵

8
 9 ¹Departamento de Ecologia e Conservação, Instituto de Ciências Naturais, Universidade
 10 Federal de Lavras, Lavras (MG), Brasil, CEP 37.200-900. rpaolo1967@gmail.com. +55 35
 11 99908 08 65

12 ²Modeling and Simulation Centre (CESIMO), Faculty of Engineering, University of the
 13 Andes, Merida 5101, Venezuela.

14 ³Laboratory of Landscape Ecology and Climate, Venezuelan Institute for Scientific Research
 15 (IVIC), Caracas, Venezuela.

16 ⁴Geoprocessing Lab. Brazilian Agricultural Research Corporation (EMBRAPA - Pantanal),
 17 Corumba, MS 79320-900, Brazil

18 ⁵Research Group on Remote Sensing and Mountain Ecology, School of Engineering and
 19 Environmental Management, National Autonomous University of Huanta, Ayacucho, Peru.

20
 21 PRP ORCID: 0000-0001-9182-3798

22 MP ORCID: 0000-0002-0940-4074

23 DT ORCID: 0000-0003-4102-1927

24 CP: ORCID: 0000-0002-9067-816X

25 MAA: ORCID: 0000-0001-5923-0300

26
 27 (Artigo publicado no *International Journal of Climatology*, <https://doi.org/10.1002/joc.7325>)

28
 29 **Abstract**

30
 31 Species distribution modelling has become instrumental in assessing the influence of
 32 environmental conditions on the occurrence or abundance of taxa. The set of environmental
 33 layers used for this purpose is a crucial aspect, for which different climate-based (bioclimatic)

34 datasets have been recently developed. These bioclimatic variables result from combinations
35 of precipitation and temperatures surfaces. Here, we explored both the performance and
36 possibility of improving some of the currently available bioclimatic databases, through an
37 evaluation of the precipitation and temperatures surfaces used to generate them. For this
38 purpose, we used a combination of statistic and graphic approaches. We focused on Brazil,
39 not only due to its natural megadiversity, but also due to its continental size and orographic
40 heterogeneity: an excellent ground for refining methods replicable elsewhere. We found a
41 better match between the climatic data measured on-field and Tropical Rainfall Measuring
42 Mission (TRMM 3B43 v7) in the case of precipitation, and the surfaces provided by the
43 National Oceanic and Atmospheric Administration (NOAA) in the case of temperatures,
44 sources uncommonly used for species niche modelling. We gauge-calibrated the best
45 performing surfaces using machine-learning algorithms and generated corrected surfaces that
46 allowed us to create BrazilClim: a database of bioclimatic variables, based on improved
47 primary surfaces, which will result in more assertive predicted distributions and more actual
48 pictures of the species' ecological requirements for megadiverse Brazil, an approach
49 replicable elsewhere. All primary and bioclimatic surfaces generated for this study may be
50 freely downloaded.

51

52 **Keywords:** Brazil, bioclimatic covariates; machine learning; spatial interpolation; TRMM
53 3B43 v7; NOAA; land surface temperature; precipitation

54

55 1. INTRODUCTION

56

57 The notion that environmental conditions shape the spatial distribution of living beings
58 has been known since the antiquity, but was quantitatively evaluated only after the
59 development of spatial estimation techniques which allowed the creation of bioclimatic layers
60 since the 1980s (Booth et al., 2014). Thenceforth, species distribution modelling (hereafter
61 SDM) and other niche-related studies (Elith and Leathwick, 2009) have increasingly become
62 instrumental in diverse fields within the biological and ecological subjects dealing with the
63 influence of environmental conditions on the occurrence or abundance of taxa (Franklin and
64 Miller, 2010).

65

66 The extent to which the predictions match the actual distribution of the taxa depends
67 on several factors: (1) the amount and accuracy of the species' occurrence data as model input
(Soulтан and Safi, 2017); (2) intrinsic characteristics of the target species (Hernandez et al.,

68 2006); (3) the existence of dispersal barriers (Aliaga-Samanez et al., 2020); (4) the effect of
69 biotic interactions (Freeman and Mason, 2015); (5) strengths and weaknesses of the
70 mathematical methods used (Qiao et al., 2015); and (6) the environmental layers used (Röd-
71 der and Lötters, 2010); among others.

72 The first dataset of environmental layers available for SMDs, BIOCLIM, consisted of
73 interpolated grids of monthly mean values for maximum and minimum temperatures
74 (henceforth referred respectively as T_{\max} and T_{\min}), precipitation (Ppt), as well as 12 derived
75 bioclimatic covariates for Australia, with a resolution of 0.5° (Booth et al., 2014). The term
76 “bioclimatic” highlights the fact that variables are more meaningful from the biological point
77 of view, being derived from the monthly temperatures and precipitation, depicting annual
78 trends, seasonality, or limiting conditions (e.g., precipitation of the warmest trimester).
79 Subsequent efforts were dedicated towards both the increment of the geographic extent, and
80 the number of bioclimatic covariates provided. For example, WorldClim (Hijmans et al.,
81 2005), the most widely used source of bioclimate data for terrestrial SDM studies, follows the
82 BIOCLIM methods, but spanning all the terrestrial surfaces and including 19 bioclimatic
83 covariates.

84 The actual relationship between the bioclimatic covariates and the respective species
85 environmental tolerances, with its concomitant effect on the geographic distributions, is
86 unknown for most taxa (Alvarado-Serrano and Knowles, 2014), making the selection of the
87 environmental variables a pivotal process within SDMs. This situation has prompted the
88 development of additional bioclimatic covariates in the recent years, for example by including
89 complex water balance calculations (Xu and Hutchinson, 2013), or their relation to
90 topographic traits (Title and Bemmels, 2018). Moreover, other covariates such as soil,
91 geology, bathymetry, distance to water and habitat patch characteristics have been made
92 available and used. Withal, precipitation and temperature are the most frequently used, being
93 very often the most important predictors (Bradie and Leung, 2017).

94 Many of the bioclimatic databases have been built on ground-based climate gauge
95 networks, which are capable of providing reliable observations with high accuracy. However,
96 the uncertainty increases when such observations are estimated in unsampled areas (Borges et
97 al., 2016), especially over montane regions, extreme climatic conditions, or where gauges are
98 distributed sparsely (Tozer et al., 2012; Miao et al., 2015; Chubb et al., 2016), that is, where
99 the biodiversity is usually higher (Körner, 2002). Concomitantly, the use of bioclimatic
100 covariates derived from interpolations of data from climate gauge networks has proven to
101 have some limitations (Deblauwe et al., 2016). As an example, Ramoni-Perazzi (2016) found

102 that the precipitation layers provided by WorldClim 1.4 (Hijmans et al., 2005) completely
103 overlook important semiarid pockets within the Cordillera de Mérida, the northernmost
104 portion of the Tropical Andes, a region which leads the list of worldwide hotspots for species
105 richness/area ratio and endemism (Myers et al., 2000). Such a situation has prompted the
106 development of datasets based on remotely sensed climate data (Waltari et al., 2014; Lima-
107 Ribeiro et al., 2015; Deblauwe et al., 2016; Karger et al., 2017; Vega et al., 2017; Abatzoglou
108 et al., 2018), in an effort to search for accuracy.

109 Although satellite climate estimates have been demonstrated by multiple studies to be
110 accurate, others have revealed flaws and weakness. As an example, the Tropical Rainfall
111 Measuring Mission (TRMM) Precipitation Radar (Kummerow et al., 2000; TRMM, 2018),
112 yielded one of the longest (1998-2014), single-sensor, high-resolution databases, and multiple
113 studies have demonstrated the accuracy of its estimates for mean precipitations (Dinku et al.,
114 2007; Chen et al., 2011; Fleming et al., 2011; Zhao et al., 2015; Wang et al., 2017). However,
115 in spite of the high correlations observed, the goodness-of-fit may vary along the year (Tao et
116 al., 2016; Cao et al., 2018), among regions (Brasil Neto et al., 2020), or under extreme rain
117 events (Wolff and Fisher, 2008), resulting in overestimations of the precipitation (Curtarelli et
118 al., 2014; Nastos et al., 2016). Thus, these estimates are prone of errors, highly correlated to
119 resolution, rain rate and sampling frequency (Nesbitt and Anders, 2009). That is,
120 biogeographers and allies can currently count on several bioclimatic database options,
121 obtained following different methodologies and criteria, but all of them are prone of some
122 source of error.

123 In this study, we explore two questions: (1) Are there significant differences among
124 the bioclimatic databases when compared to on-field measurements? and (2) How can any
125 eventual incongruence be corrected? To answer these questions, we focused on megadiverse
126 Brazil, which harbors the largest, and largely undiscovered, continental biota (Brandon et al.,
127 2005), jeopardized with high rates of habitat degradation and loss, as well as the effects of
128 climate change (Brodie et al., 2012; Lapola et al., 2014; Molotoks et al., 2018).

129

130 **2. MATERIALS AND METHODS**

131 **2.1. The study area**

132

133 Brazil is a continent-sized country, not only megadiverse from a biological point of
134 view (Mittermeier et al., 1997), but also from a climatic one. This is attributed to the fact that

135 the country possesses a great variety of topographic features, exemplified by the disparity of
136 plain landscapes associated to the Amazon basin to the north and the mass of ridges and
137 mountain ranges toward the southeast. Besides, a latitudinal extent of about 45°, unique in
138 being crossed by both the Equator and the Tropic of Capricorn, creates a gradient from
139 tropical to temperate conditions. Concomitantly, Brazil hosts three main climatic types
140 according to the Köppen Classification system (Alvares et al., 2013): tropical rainy (with four
141 subtypes), dry (one subtype), and humid subtropical (with seven subtypes).

142 Such climatic diversity results in diverging environments generally grouped into six
143 major biomes. First, an equatorial climate, with average temperatures above 18 °C, and
144 variable precipitation although no actual dry season defines the Amazonia to the North
145 (Figure 2.1). Second, the Caatinga to the Northeast is characterized by a semiarid climate with
146 a prolonged dry season, that lasts for seven or more months, and average temperatures above
147 18 °C. Third, the Atlantic Forest, stretching along the Atlantic coast from the Rio Grande do
148 Norte state in the northeast to the Rio Grande do Sul state in the south. This area comprises
149 variable conditions, ranging from equatorial climates with monthly temperatures above 18 °C
150 and lacking a real dry season to humid subtropical climates characterized by hot and humid
151 summers, and cold to mild winters. Fourth, these humid subtropical climates also define the
152 Pampa biome to the South of the Country. Finally, a marked precipitation seasonality, typical
153 to savanna climates, characterizes central Brazil, where the Pantanal occupies the floodable
154 areas, while the surrounding highlands are covered by the Cerrado. Thus, Brazil is suited for
155 analyses like the ones in the present study.

156

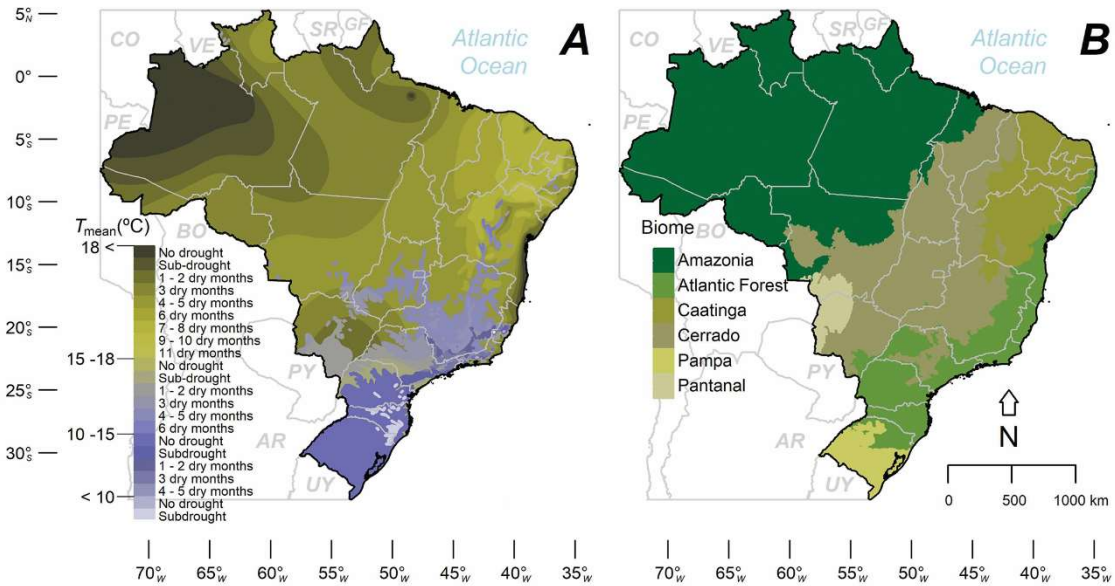
157 **2.2. Comparing observed and estimated measurements**

158

159 The most widely used bioclimatic covariates are the 19 provided, for example, by WorldClim
160 1.4 (Hijmans et al., 2005), which are derived from T_{\max} , T_{\min} , and Ppt . Ergo, in the present
161 study and for these three primary climatic variables, we focused on the comparison of data
162 measured on-field against those provided by databases listed in Table 2.1, which includes
163 sources as diverse as traditional bioclimatic databases based on interpolations such as
164 WorldClim 1.4 (Hijmans et al., 2005), and those based on remotely sensed information
165 nontraditionally used for SDMs such as TRMM 3B43 v7 (Kummerow et al., 2000; TRMM,
166 2018) or those provided by the National Oceanic and Atmospheric Administration (NOAA).

167

Figure 2.1 – Distribution of (A) climate and (B) biomes in mainland Brazil, according to the Brazilian Institute of Geography and Statistics (IBGE, 2021).



168

169

170

171

172

173

174

175

176

177

178

We performed all the analyses on a monthly basis, using the R version 3.6.3 "Holding the Windsock" (R Core Team, 2020), and chose a significance level of $\alpha = 0.05$. We obtained all climatic data directly from the respective websites of the databases, projecting them in a geographic (lon/lat) coordinate system using the WGS84 datum and resampling them to a 0.5° resolution when required. This change in spatial resolution does not represent an actual improvement of those layers delivered at lower resolution, but allows to homogenize it across databases keeping all the information of those provided at highest resolution and improving the workflow.

Table 2.1 – Basic information of the bioclimatic datasets used in this study. *Ppt*: mean monthly accumulated precipitation; T_{max} : mean monthly maximum temperature; T_{min} : mean monthly minimum temperature. “Time span” refers to the temporal series used for determining the climatology.

Database	Subsets	Variables	Resolution	Time span	Method	URL	Accessed	Citation
CRU CL 2.0		Ppt, T_{max}, T_{min}	10'	1961 - 1990	Splines interpolation among stations	https://crudata.uea.ac.uk/cru/data/hrg/tmc/	Jan/25/2020	New et al. (2002)
WorldClim 1.4		Ppt, T_{max}, T_{min}	0.5'	1960 - 1990	Splines interpolation among stations	https://www.worldclim.org/data/v1.4/worldclim14.html	Dec/05/2019	Hijmans et al. (2005)
NOAA ¹		T_{max}, T_{min}	30'	1981 - 2010	Remote sensing	https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html	Jul/28/2020	
Climond	WorldClim 1.4	Ppt	10'	1961 - 1990	Combination of CRU and WorldClim 1.4	https://www.climond.org/ClimateData.aspx	Jul/09/2020	Kriticos et al. (2012)
	CRU CL	T_{max}, T_{min}	10'	1961 - 1990				
ecoClimate	CCSM	Ppt, T_{max}, T_{min}	0.5'	1950 - 1999	Simulations for eight coupled atmosphere-ocean global climate models (AOGCMs)	https://www.ecoclimate.org/downloads/	Jul/27/2020	Lima-Ribeiro et al. (2015, 2020)
	CNRM	Ppt, T_{max}, T_{min}	0.5'					
	FGOALS	Ppt, T_{max}, T_{min}	0.5'					
	GISS	Ppt, T_{max}, T_{min}	0.5'					
	IPSL	Ppt, T_{max}, T_{min}	0.5'					
	MIROC	Ppt, T_{max}, T_{min}	0.5'					
	MPI	Ppt, T_{max}, T_{min}	0.5'					
MRI	Ppt, T_{max}, T_{min}	0.5'						
Deblauwe's	CHIRPS v. 2.0	Ppt	3'	1981 - 2013	Reuse of data from MODIS and CHIRPS	https://vdeblauwe.wordpress.com/download/	Jul/10/2020	Deblauwe et al. (2016)
	MOD11C3 v. 5.0	T_{max}, T_{min}	3'	2001 - 2013				
CHELSA 1.2		Ppt, T_{max}, T_{min}	0.5'	1979 - 2013	Reanalysis of ERA-interim	https://chelsa-climate.org/downloads/	Jul/27/2019	Karger et al. (2017, 2018)
WorldClim 2.1		Ppt, T_{max}, T_{min}	0.5'	1970 - 2000	Splines interpolation among stations	https://www.worldclim.org/data/worldclim21.html	Aug/03/2020	Fick and Hijmans (2017)
MERRAclim		T_{max}, T_{min}	2.5'	1980 - 2000	Reanalysis of satellite information (MERRA)	https://datadryad.org/stash/dataset/doi:10.5061/dryad.s2v81	Jul/19/2020	Vega et al. (2017, 2018)
Terraclimate		Ppt, T_{max}, T_{min}	0.4'	1961 - 1990	Climatically aided interpolation, combining climatological normals from the WorldClim, CRU Ts4.0 and the Japanese 55-year Reanalysis (JRA55)	http://www.climatologylab.org/terraclimate.html	Jul/10/2020	Abatzoglou et al. (2018)
TRMM_3B43 v 7		Ppt	15'	1999 - 2019	Remote sensing	https://disc.gsfc.nasa.gov/datasets/TRMM_3B43_7/summary	Jan/16/2020	Kummerow et al. (2000), TRMM (2018)

¹ CPC Global Temperature data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA

187 The climatic information measured by meteorological (i.e., on-field measurements)
188 gauges is not exempted of inexactitudes or errors. Thus, before using the information
189 provided by the Brazilian official meteorological network managed by the National Institute
190 of Meteorology (INMET), considerations from Ramoni-Perazzi et al. (*in press*) were taken
191 into account: We appraised the climatic information gathered by both conventional (hereafter
192 *CMG*) and automatic (*AMG*) meteorological gauges, with at least 10 complete years of data
193 within the period 2000-2019.

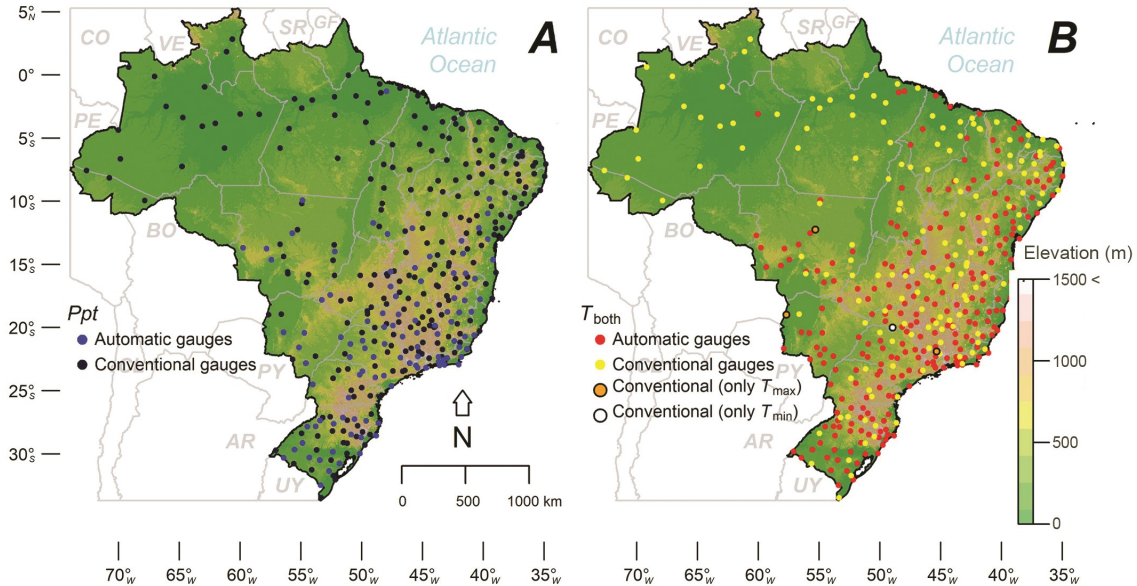
194 In the case of *Ppt*, we created a database consisting of information from 257 *CMGs*,
195 since they performed slightly better, complemented with 107 *AMGs* separated by at least 100
196 km from the nearest *CMG*, resulting in a final database of 364 gauges (Supplementary
197 Material A). In the case of T_{\max} and T_{\min} , we created a final database using the information
198 from 251 *AMGs*, which performed slightly better with regards to the temperatures,
199 complemented respectively with 158 and 156 *CMGs* separated by at least 100 km from the
200 nearest *AMG*. That is, the databases for measured T_{\max} and T_{\min} included a total of 409 and
201 407 gauges (Supplementary Material A), respectively. Both datasets covered unevenly the
202 Brazilian territory (Figure 2.2), being more concentrated towards the axis conformed by the
203 South-to-Northeast regions, coinciding with the highest turnover of climatic conditions and,
204 concomitantly, of terrestrial life zones (Figure 2.1; Soares-Filho et al., 2014; Overbeck et al.,
205 2015; Tres et al., 2020).

206 We compared the information measured on-field by the meteorological gauges against
207 the values from their respective conterminous pixels of the different databases following four
208 strategies: (1) visually through paired notched boxplots, whose medians differ with a 95% of
209 confidence when the notches of two boxes do not overlap (Chambers et al., 1983); (2)
210 statistically through paired *t*-Student tests; (3) spatially through animated maps in graphic
211 interchange format (.gif), representing the positive/negative differences between the values
212 measured on-field minus the estimated ones; and (4) calculating the root-mean-square error
213 (RMSE) a measure of accuracy frequently used to compare forecasting errors of different
214 models for a particular dataset (Hyndman and Koehler, 2006).

215 In the case of *Ppt*, these comparisons did not include the database MERRAclim (Vega
216 et al., 2017, 2018) which, instead of precipitation, approaches the water availability through
217 the use of air humidity information.

218

Figure 2.2 – Distribution of the meteorological stations providing climate information used in the present study. (A) Mean monthly accumulated precipitation; and (B) Mean maximum and mean minimum temperatures.



219

220

221 2.3. Creating the definitive climatic layers

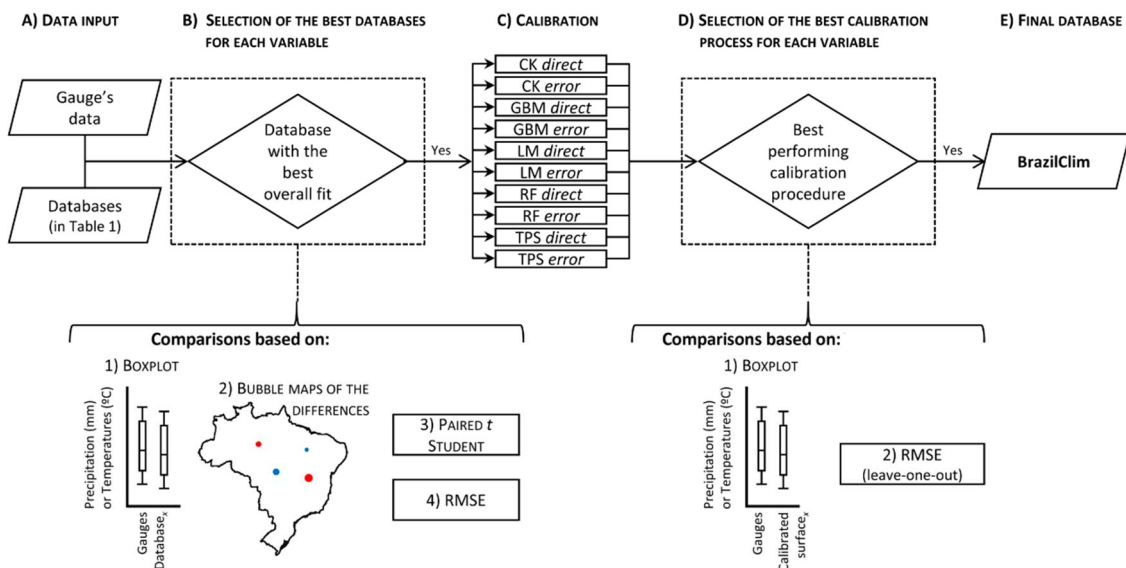
222

223 Once we determined the performance of the aforementioned databases, we kept the
 224 highest scored ones for each variable and calibrated them using the measurements obtained
 225 on-field by the weather gauges managed by the INMET. We addressed these calibrations
 226 comparing two strategies: (1) by interpolating spatially the measurements provided by
 227 INMET's gauges (i.e., direct-based method); or (2) by interpolating the differences resulting
 228 from subtracting the amounts measured on-field to the values provided in the corresponding
 229 conterminous pixels in the selected databases, and then adding such interpolated surfaces back
 230 to the corresponding surfaces of the selected databases (i.e., error-based method). All these
 231 interpolations were performed using the information from the respective selected databases as
 232 covariates and, in the case of temperatures, also including the elevation and latitude as
 233 covariates given their high correlations (Ramoni-Perazzi et al., in press). For that, we used the
 234 Global Multi-resolution Terrain Elevation Data 2010 (Danielson and Gesch, 2011).

235 Many spatial interpolation methods are currently available (Li and Heap, 2011), most
 236 of which have been implemented in free languages such as R (R Core Team, 2020); from
 237 these we selected and compared the performance of five of them. We based such a selection
 238 on the simplicity of their implementation and their comparatively low computing processing

239 requirements. We evaluated the following interpolation techniques: (1) Cokriging (CK), a
 240 multivariate geostatistical technique that takes advantage of non-exhaustive secondary
 241 information to explicitly account for the spatial cross-correlation between the primary variable
 242 and the covariates (Goovaerts, 1997), implemented in the package “gstat” (Pebesma, 2004;
 243 Gräler et al., 2016); (2) Generalized boosted regression modeling (GBM), a machine learning
 244 technique, using the package “spm” (Li, 2019); (3) Simple regression (LM), implemented in
 245 the package “stats” (R Core Team, 2020); (4) Quantile random forest (RF), another machine
 246 learning tool, which provides a non-parametric accurate way of calculating conditional
 247 quantiles for high-dimensional predictor variables, through the package “quantregForest”
 248 (Meinshausen, 2006, 2017); and (5) Thin plate splines (TPS), a type of smoothing spline for
 249 irregularly spaced data, fitted using generalized additive models, implemented in the package
 250 “fields” (Nychka et al., 2017).
 251

Figure 2.3 – Workflow for BrazilClim creation. Evaluation strategies used are listed within dashed boxes. Interpolation (calibration) acronyms: (CK) Cokriging; (GBM) Generalized boosted regression modeling; (LM) Simple regression; (RF) Quantile random forest; and (TPS) Thin plate splines.



252

253

254 We compared the performances of these interpolation techniques through two
 255 strategies. First, by calculating and comparing the RMSEs following a leave-one-out scheme:
 256 a particular cross-validation method closely related to the jackknife approach (Efron, 1982),
 257 where the number of folds equals the number of gauges in the database. Second, visually,
 258 using boxplots to contrast the distributions of the data measured on-field against those

259 provided by the databases selected against the interpolated outputs obtained through the
260 aforementioned methods.

261 After selecting the best-performing datasets of Ppt , T_{max} and T_{min} , we created
262 BrazilClim: the bioclimatic variables for the continental megadiverse Brazil. For this, we used
263 the command ‘biovars’ of the ‘dismo’ package (Hijmans et al., 2017). A diagram
264 summarizing this procedure is shown in Figure 2.3.

265

266 3. RESULTS

267

268 3.1. Comparing observed and estimated measurements

269

270 The degree of concordance between the measured on-field and the estimated values
271 greatly varied among covariates, databases, as well as from one month to another within a
272 given database. In the case of Ppt (Figure B1 in Supplementary Material B), more databases
273 match the on-field measurements during the period corresponding to the austral summer
274 (January-March) than during the austral winter (July-September). Among the evaluated
275 databases, both CHELSA v 1.2 (Karger et al., 2017, 2018) and TRMM 3B43 v7 (Kummerow
276 et al., 2000; TRMM, 2018) showed the best performances as demonstrated by the great
277 overlap of their boxplots, even in the respective notches, and the lack of statistical differences
278 in five of the 12 months, when contrasted against the precipitations measured on-field by the
279 meteorological gauges.

280 In the case of T_{max} , only the databases MERRAclim (Vega et al., 2017, 2018), or
281 GISS, IPL and MIROC from EcoClimate (Lima-Ribeiro et al., 2015, 2020) showed no
282 significant differences with respect to the values measured on-field, and only within the
283 period from May to November (Figure B2 in Supplementary Material B). Conversely, several
284 databases showed no significant differences when compared against T_{min} measured on-field
285 (Figure B3 in Supplementary Material B). From these, the one provided by Deblauwe et al.
286 (2016) outstands given its congruence in five of the 12 months, superseded only by NOAA
287 which concurs in six months.

288 All the evaluated databases showed some conspicuous discrepancies when contrasted
289 against the corresponding values reported by the meteorological gauges. In the case of the Ppt
290 (Supplementary Material C, Animated GIFs 1.a-p), these differences ranged from
291 underestimations of 474 mm in the case of the CCSM database relative to the *CMG* Colinas

292 (Maranhão), to overestimations of 588 mm for the database MPI compared to the *CMG* Soure
293 (Pará), both cases for March and both databases provided by EcoClimate (Lima-Ribeiro et al.,
294 2015, 2020).

295 In the case of T_{\max} (Supplementary Material C, Animated GIFs 2.a-p), such
296 discrepancies ranged from underestimations of 11.9 °C in October for IPSL (Lima-Ribeiro et
297 al., 2015, 2020) when set side by side to the *AMG* Campos do Jordão (São Paulo), to
298 overestimations of 10.1°C in August for MPI (Lima-Ribeiro et al., 2015, 2020) in relation to
299 the *CMG* Goiás, located in the homonymous State. Finally, in the case of T_{\min} (Supplementary
300 Material C, Animated GIFs 3.a-p), such disparities ranged from underestimations of 16.9°C
301 for NOAA for the *CMG* Santa Vitória do Palmar in July, to overestimations of 11.0°C in June
302 in the case of IPSL (Lima-Ribeiro et al., 2015, 2020) and the *CMG* São Luiz Gonzaga, both in
303 Rio Grande do Sul State.

304 In most cases, the discrepancies between the values reported by the meteorological
305 gauges and the correspondent conterminous pixels of the different covariates and evaluated
306 databases were very dynamic in terms of sign (positive or negative) and magnitude. In other
307 words, when analyzing each covariate and database separately, the values reported by any
308 given meteorological gauge can be underestimated for one month but overestimated the
309 following one. However, an attention-grabbing aspect was the tendency of these under- or
310 overestimations to be spatially clustered, with some geographic displacement trends along the
311 year, especially in the case of the *Ppt* (Supplementary Material C, Animated GIFs 1a-p).

312 In the case of *Ppt*, the RMSEs clearly indicated a better performance for the database
313 provided by Deblauwe et al., (2016), as well as TRMM 3B43 v7 (Kummerow et al., 2000;
314 TRMM, 2018), although the latter shows lower maximum and minimum RMSE values
315 (Figure 2.4A). In the case of temperatures, the RMSEs suggested a superior performance of
316 NOAA, especially in the case of the mean monthly maximum temperatures (Figure 2.4B-C).

317 Consequently, in addition to this best performance when compared to on-field
318 measurements through paired Student's t tests, for further analyses we kept TRMM 3B43 v7
319 (Kummerow et al., 2000; TRMM, 2018) in the case of *Ppt*, and NOAA in the case of
320 temperatures. Interestingly, both sources are not commonly used for species niche modelling.

321

322 **3.2. Creating the definitive climatic layers**

323

324 For *Ppt*, the great overlap of the boxplots' notches suggested that all the interpolation
325 techniques considered in the present study yielded RMSEs with similar medians (Figure 2.5).

326 However, error-based GBM, LM (both approximations) and TPS (both approximations)
 327 yielded the lowest RMSEs maxima, when compared to the remnant techniques, although the
 328 latter predicted precipitation values evidently far above those measured on-field or reported
 329 by TRMM 3B43 v7 (Kummerow et al., 2000; TRMM, 2018) in June (LM) and, especially,
 330 May (TPS; Figure B4 in Supplementary Material B). Conversely, error-based GBM
 331 interpolations were more reasonably fitted, hence, we used this procedure to gauge-calibrate
 332 the TRMM 3B43 v7 (Kummerow et al., 2000; TRMM, 2018) layers in order to produce the
 333 definitive *Ppt* surfaces (Figure 2.6; Supplementary Material C, Animated GIF 4.a). In few
 334 cases, the calibration process resulted in negative precipitation values, impossible in nature,
 335 thus we equalized them to zero prior to further calculations.

Figure 2.4 – Overall (all months) Root Mean Squared Errors contrasting all the databases evaluated in this study (listed in Table 2.1), against BrazilClim (BrCl): our gauge-calibrated dataset. (A) Mean monthly accumulated precipitation; (B) Mean maximum temperature; (C) Mean minimum temperature.

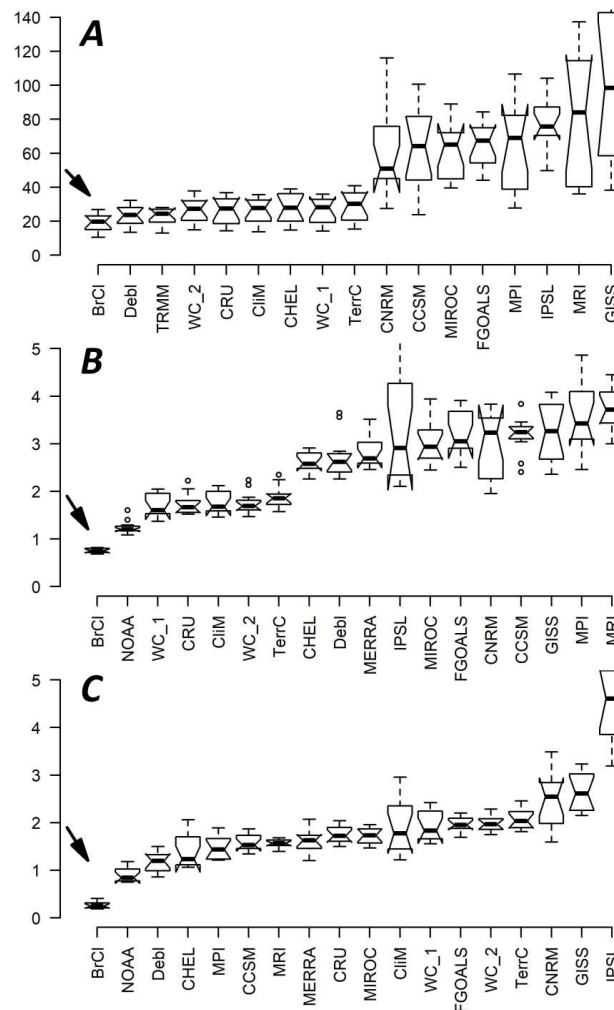


Figure 2.5 – Overall (all months) Root Mean Squared Errors calculated comparing different interpolation methods. (A) Mean monthly accumulated precipitation; (B) Mean maximum temperature; (C) Mean minimum temperature; (CK) Cokriging; (GBM) Generalized boosted regression modeling; (LM) Simple regression; (RF) Quantile random forest; and (TPS) Thin plate splines. White boxes represent the results obtained by interpolating the measurements performed on-field (“direct-based”); gray boxes represent the interpolations of the differences resulting by subtracting the amounts measured-on-field minus the value provided in the corresponding conterminous pixels by the selected databases (TRMM in the case of the precipitation and NOAA in the case of the temperatures), and then adding such interpolated surfaces to the corresponding surfaces of the selected databases (“error-based”).

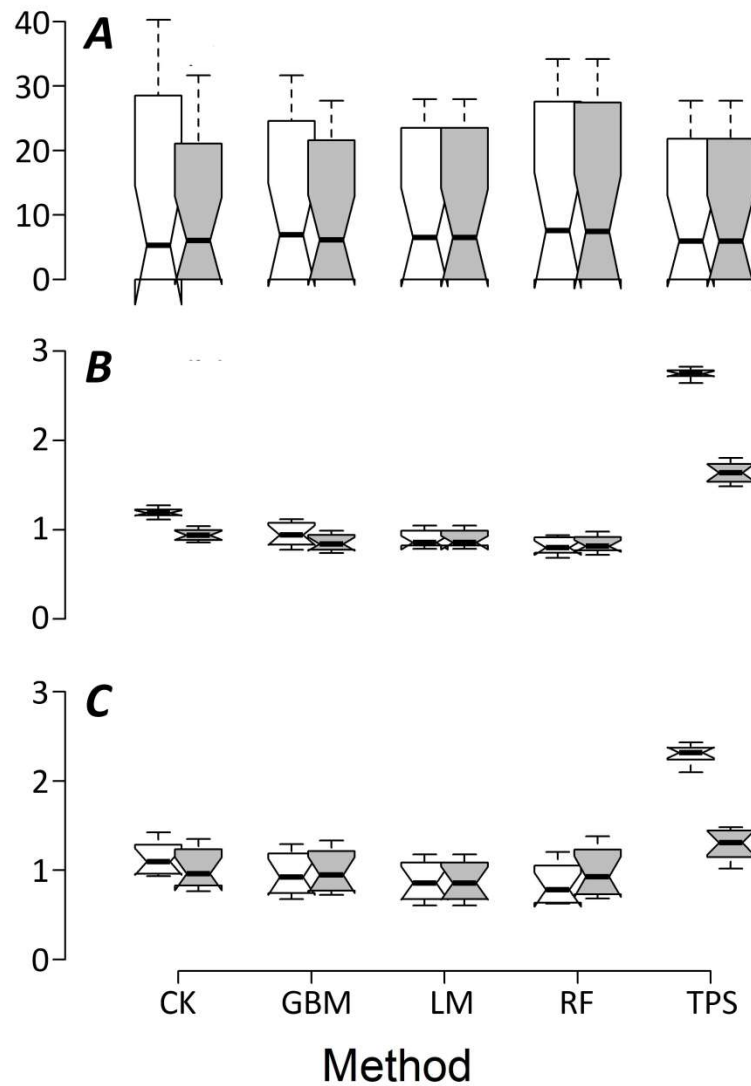
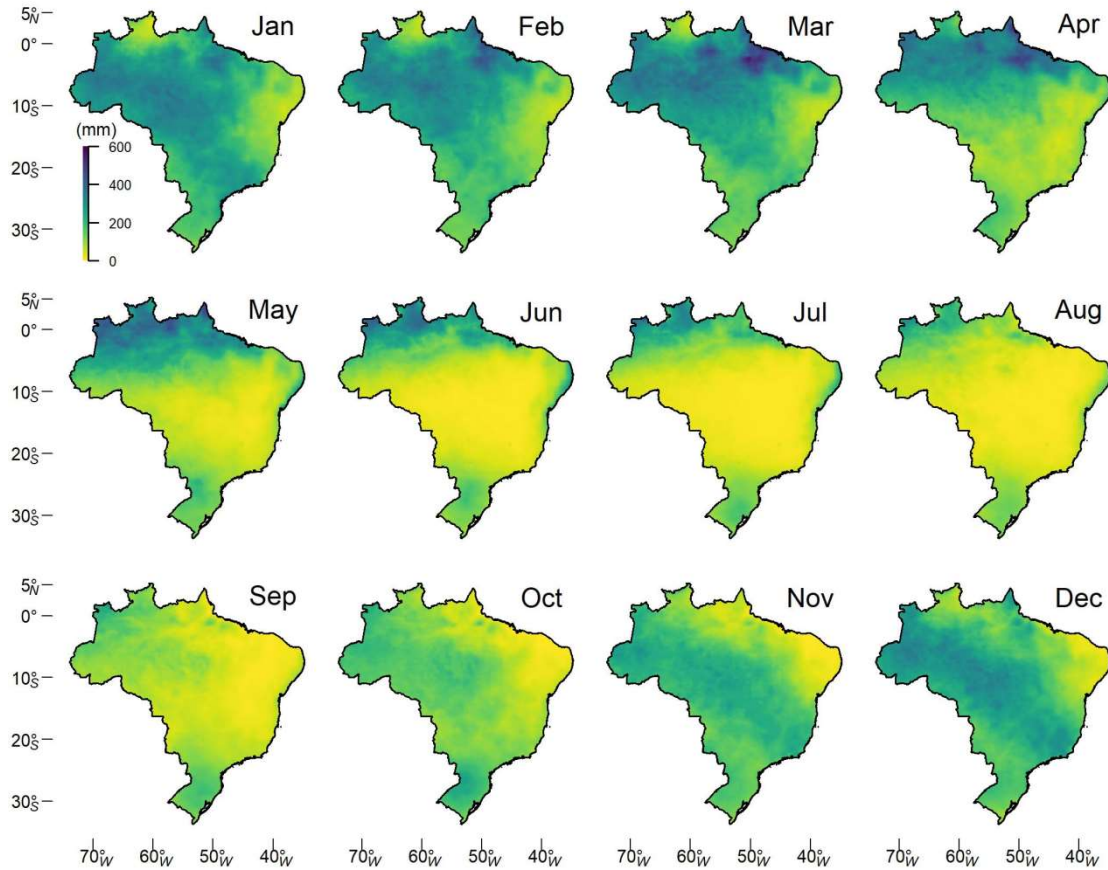


Figure 2.6 – Mean monthly accumulated precipitation surfaces of BrazilClim, resulting from gauge-calibrating information from TRMM 3B43 v7 (Kummerow et al., 2000; TRMM, 2018) through error-based interpolations with generalized boosted regression modeling (GBM).



338

339

340 Error-based GBM, along with RF (both approximations), outperformed the other
 341 interpolation methods evaluated in the case of the temperatures (Figure 2.5; Supplementary
 342 Material B). However, error-based GBM reached lowest RMSEs minima in the case of T_{\max} ,
 343 and direct-based RF in the case of T_{\min} . Hence, we used each calibration procedure to obtain
 344 the definitive T_{\max} and T_{\min} surfaces (Figures 2.7 and 2.8, respectively; Supplementary
 345 Material C, Animated GIFs 4.b-c).

346 The definitive surfaces are not exempt of discrepancies compared to the data measured
 347 on-field (Supplementary Material B, Animated GIFs 1.q, 2.r, 3.r). In the case of T_{\max} , the
 348 maximum overestimations of BrazilClim (4.4 °C) exceed those of CHELSA (1.5 °C; Karger et
 349 al., 2018), WorldClim 2.1 (2.7 °C; Fick and Hijmans, 2017), Terraclimate (3.1 °C; Abatzoglou
 350 et al., 2018) or CliMond (3.7 °C; Kriticos et al., 2012). Aside from this, BrazilClim
 351 outperforms the evaluated databases given that, for all variables, the differences with respect
 352 to the values measured on-field are largely centered at zero with narrower interquartile ranges.

353 Although these discrepancies do not follow an identifiable spatial pattern, given the
354 lack of an evident clustering or regionalization, in the case of precipitations they tend to be
355 wider outside the austral winter season, while temperatures seasonal variations seem to have
356 little or no effect (Figure B7 in Supplementary Material B).

357

358 **4. DISCUSSION**

359

360 There is a generalized tendency of modelers to use prêt-a-potter datasets.
361 Notwithstanding, given the existing incongruences between estimated and measured data, the
362 availability of free resources (such as R), and the relatively accessible computer-power
363 required to perform spatial interpolations, it is worth dedicating additional efforts to produce
364 ad hoc datasets, especially when working at subcontinental levels.

365 As we already mentioned, the bioclimatic variables are derived from the primary
366 surfaces of precipitation and temperatures. Consequently, in addition to the expected
367 uncertainties due to on-field measurement limitations (e.g., instrument precision), as
368 evidenced by discrepancies among *AMGs* and *CMGs* in the same location (Ramoni-Perazzi et
369 al., in press), and estimation (interpolation/sensing) used to create the primary surfaces, and as
370 demonstrated in the present study, the bioclimatic variables are affected by random and
371 uncorrelated errors propagated due to the combination of variables in the function. This
372 situation underlines the importance of making efforts to generate primary surfaces as reliable
373 as possible.

374 Several studies have demonstrated that satellite estimates can be significantly
375 improved through local gauge-calibration (Cheema and Bastiaanssen, 2012; Manz et al.,
376 2016; Mateus et al., 2016). In the case of South America, most of these efforts were
377 successfully carried out through deterministic or geostatistical methods (Manz et al., 2016).
378 However, our findings agree with a growing body of evidence that machine learning
379 algorithms offer a postprocessing opportunity to improve satellite-derived information by
380 correcting for measurement errors in a relatively simple way and with improved overall
381 performance from a statistical point of view (Ma et al., 2018; Just et al., 2020; Reis et al.,
382 2020).

Figure 2.7 – Mean maximum temperatures surfaces of BrazilClim, resulting from gauge-calibrating information from the National Oceanic and Atmospheric Administration (NOAA) through error-based interpolations with generalized boosted regression modeling (GBM).

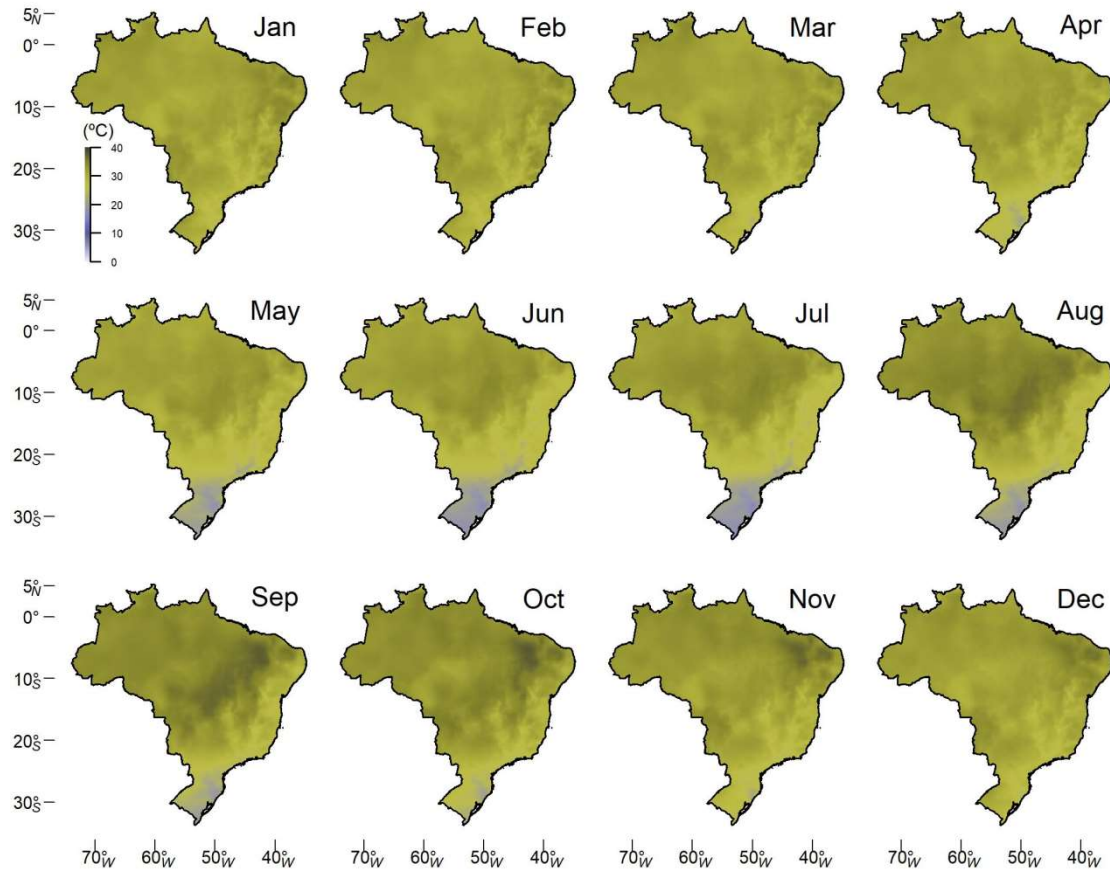
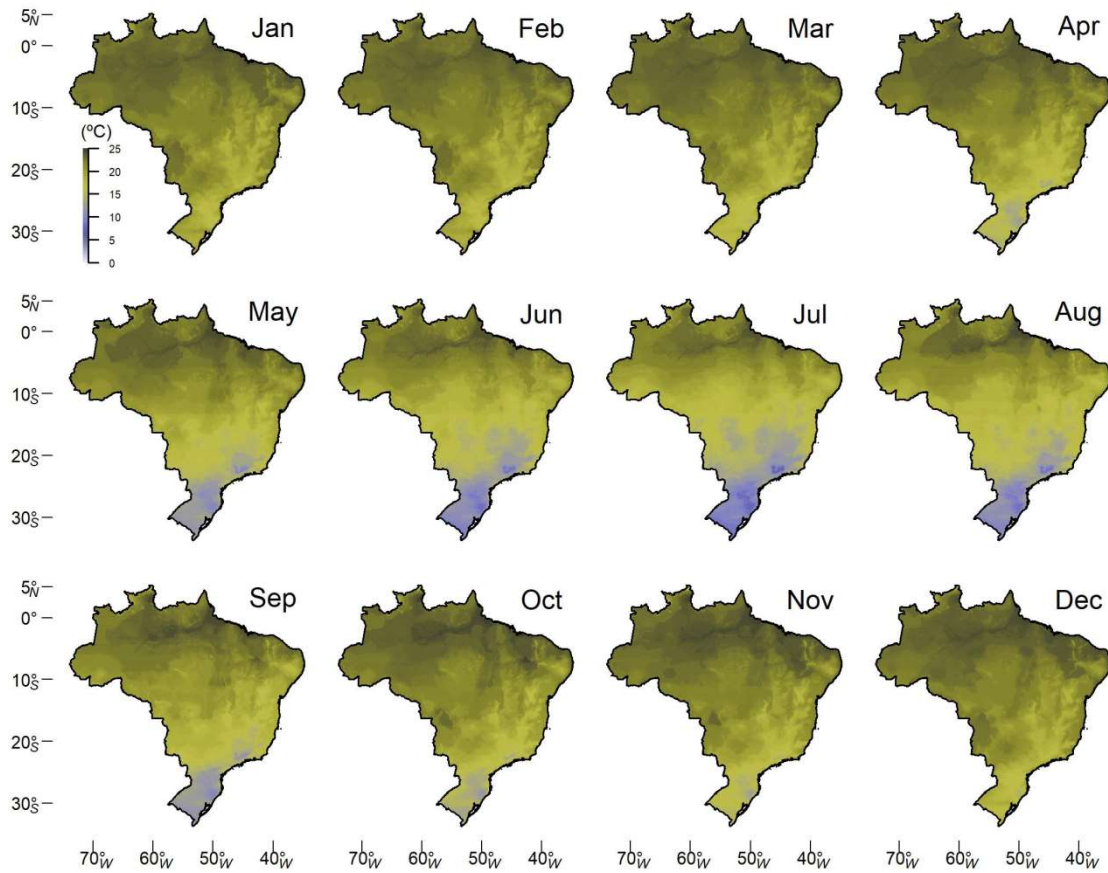


Figure 2.8 – Mean minimum temperatures surfaces of BrazilClim, resulting from gauge-calibrating information from the National Oceanic and Atmospheric Administration (NOAA) through interpolations with random forest (RF).



384

385

386

387

388

389

390

391

392

393

394 5. CONCLUDING REMARKS

395

396

397

398

Therefore, we built BrazilClim primary surfaces of precipitation and temperatures based on climate surfaces developed during the recent past, whose use and validation has been frequently referred to in the specialized literature. We calibrated these surfaces through already tested techniques, based on information measured on-field by a state agency and analyzed by us in a previous study, within a process guided by performance measures frequently used in similar studies. Consequently, it is reasonable to suggest that the robustness of the resulting primary surfaces is sufficiently validated.

These improved primary surfaces not only will result in more assertive predicted distributions, but will also provide more actual pictures of the species' ecological requirements, information valuable for fields such as ecophysiology, agriculture and others.

399 Moreover, the past climate oscillations and the concomitant habitat
400 expansions/contractions shaped the current distribution of Brazilian biota (Ledo and Colli,
401 2017; Arruda et al., 2018; Capurucho et al., 2018; Dal Vechio et al., 2018; Silveira et al.,
402 2019). On the other hand, the ongoing trends of climate change represent forthcoming
403 changes and threats to this biota (Zanin et al., 2017; Hidasi-Neto et al., 2019) and even to
404 economic activities including important ones such as agriculture (Masseti et al., 2013; Zilli et
405 al., 2020). Therefore, the logical next step would be the projection of BrazilClim to both past
406 and future climatic scenarios, which will be a valuable tool to understand the distribution of
407 the biota in a changing world.

408

409 **6. DATA ACCESSIBILITY**

410

411 The following supplementary material is available online:

412 Supplementary Material A <https://doi.org/10.4121/14924397>

413 Supplementary Material B <https://doi.org/10.4121/14925048>

414 Supplementary Material C <https://doi.org/10.4121/14932251>

415 BrazilClim primary and bioclimatic surfaces <https://doi.org/10.4121/14932947>

416 Script to gauge-calibrate the surfaces <https://doi.org/10.4121/14935791>

417

418 **7. ACKNOWLEDGEMENTS**

419

420 We acknowledge the following institution for making publicly available the respective
421 information used directly or indirectly in the present study: (1) the World Climate Research
422 Program's Working Group on Coupled Modelling, which is responsible for CMIP, and the
423 U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison,
424 which provides coordinating support and led development of software infrastructure in
425 partnership with the Global Organization for Earth System Science Portals; (2) the
426 NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, which provides the CPC Global
427 Temperature data; (3) the Climate Hazards Center of the UC Santa Barbara, for the CHIRPS v
428 2.0 data; (4) the Goddard Earth Sciences Data and Information Services Center (GES DISC),
429 for the TRMM 3B43 data; (5) the Land Processes Distributed Active Archive Center (LP
430 DAAC) of the NASA Earth Observing System, for the MOD11C3 data; (6) the Japan
431 Meteorological Agency, for the JRA55 data; and (7) the National Institute of Meteorology of

432 Brazil, for the climatic information measured on-field. The constructive comments from Dr.
 433 Stephen Barigye and three anonymous reviewers greatly improved the manuscript. PRP also
 434 thanks the scholarship from the Organization of American States through its Partnerships
 435 Program for Education and Training of the Coimbra Group of Brazilian Universities
 436 (OAS/PAEC/GCUB).

437

438 **8. BIBLIOGRAPHY**

439

440 Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A. and Hegewisch, K.C. (2018) TerraClimate, a high-resolution
 441 global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5, 170191.
 442 [https://doi.org/ 10.1038/sdata.2017.191](https://doi.org/10.1038/sdata.2017.191).

443 Aliaga-Samanez, A., Real, R., Vermeer, J. and Olivero, J. (2020) Modelling species distributions limited by
 444 geographical barriers: A case study with African and American primates. *Global Ecology and Biogeography*,
 445 29(3), 444–453. <https://doi.org/10.1111/geb.13041>.

446 Alvarado-Serrano, D.F. and Knowles, L.L. (2014) Ecological niche models in phylogeographic studies:
 447 applications, advances and precautions. *Molecular Ecology Resources*, 14(2), 233–248.
 448 <https://doi.org/10.1111/1755-0998.12184>.

449 Alvares, C.A., Stape, J.L., Sentelhas, P.C., Gonçalves, J.L. de M. and Sparovek, G. (2013) Köppen's climate
 450 classification map for Brazil. *Meteorologische Zeitschrift*, 22(6), 711–728. [https://doi.org/10.1127/0941-](https://doi.org/10.1127/0941-2948/2013/0507)
 451 [2948/2013/0507](https://doi.org/10.1127/0941-2948/2013/0507).

452 Arruda, D.M., Schaefer, C.E.G.R., Fonseca, R.S., Solar, R.R.C. and Fernandes-Filho, E.I. (2018) Vegetation
 453 cover of Brazil in the last 21 ka: New insights into the Amazonian refugia and Pleistocenic arc hypotheses.
 454 *Global Ecology and Biogeography*, 27(1), 47–56. <https://doi.org/10.1111/geb.12646>.

455 Booth, T.H., Nix, H.A., Busby, J.R. and Hutchinson, M.F. (2014) Bioclim: the first species distribution
 456 modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and*
 457 *Distributions*, 20(1), 1–9. <https://doi.org/10.1111/ddi.12144>.

458 Borges, P. de A., Franke, J., da Anunciação, Y.M.T., Weiss, H. and Bernhofer, C. (2016) Comparison of spatial
 459 interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theoretical and*
 460 *Applied Climatology*, 123, 335–348. <https://doi.org/10.1007/s00704-014-1359-9>.

461 Bradie, J. and Leung, B. (2017) A quantitative synthesis of the importance of variables used in MaxEnt species
 462 distribution models. *Journal of Biogeography*, 44(6), 1344–1361. <https://doi.org/10.1111/jbi.12894>.

463 Brandon, K., da Fonseca, G.A.B., Rylands, A.B. and da Silva, J.M.C. (2005) Special section: Brazilian
 464 conservation: Challenges and opportunities. *Conservation Biology*, 19(3), 595–600.
 465 <https://doi.org/10.1111/j.1523-1739.2005.00710.x>.

- 466 Brasil Neto, R.M., Santos, C.A., Nascimento, T.V., Silva, R.M. and dos Santos C.A. (2020) Evaluation of the
467 TRMM product for monitoring drought over Paraíba State, Northeastern Brazil: a statistical analysis. *Remote*
468 *Sensing*, 12(14), 1097. <https://doi.org/10.3390/rs12142184>.
- 469 Brodie, J., Post, E. and Laurance, W. (2012) Climate change and tropical biodiversity: A new focus. *Trends in*
470 *Ecology & Evolution*, 27(3), 145–50. <https://doi.org/10.1016/j.tree.2011.09.008>.
- 471 Cao, Y., Zhang, W. and Wang, W. (2018) Evaluation of TRMM 3B43 data over the Yangtze River Delta of
472 China. *Scientific Reports*, 8, 5290. <https://doi.org/10.1038/s41598-018-23603-z>.
- 473 Capurucho, J.M.G., Ashley, M.V., Ribas, C.C. and Bates, J.M. (2018) Connecting Amazonian, Cerrado, and
474 Atlantic forest histories: Paraphyly, old divergences, and modern population dynamics in tyrant-manakins
475 (*Neopelma/Tyrannetes*, Aves: Pipridae). *Molecular Phylogenetics and Evolution*, 127, 696–705.
476 <https://doi.org/10.1016/j.ympev.2018.06.015>.
- 477 Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A. (1983) Comparing data distributions. In: J.M.
478 Chambers, W.S. Cleveland and B. Kleiner, eds., *Graphical methods for data analysis*. Belmont: Wadsworth
479 International Group, pp. 47–74.
- 480 Cheema, M.J.M. and Bastiaanssen, W.G.M. (2012) Local calibration of remotely sensed rainfall from the
481 TRMM satellite for different periods and spatial scales in the Indus Basin. *International Journal of Remote*
482 *Sensing*, 33(8), 2603–2627. <https://doi.org/10.1080/01431161.2011.617397>.
- 483 Chen, C., Yu, Z., Li, L. and Yang, C. (2011) Adaptability evaluation of TRMM satellite rainfall and its
484 application in the Dongjiang River Basin. *Procedia Environmental Sciences*, 10(A), 396–402.
485 <https://doi.org/10.1016/j.proenv.2011.09.065>.
- 486 Chubb, T.H., Manton, M.J., Siems, S.T. and Peace, A.D. (2016) Evaluation of the AWAP daily precipitation
487 spatial analysis with an independent gauge network in the Snowy Mountains. *Journal of Southern Hemisphere*
488 *Earth Systems Science*, 66(1), 55–67. <https://doi.org/10.1071/ES16006>.
- 489 Curtarelli, M.P., Rennó, C.D. and Alcântara, E.H. (2014) Evaluation of the Tropical Rainfall Measuring Mission
490 3B43 product over an inland area in Brazil and the effects of satellite boost on rainfall estimates. *Journal of*
491 *Applied Remote Sensing*, 8(1), 083589. <https://doi.org/10.1117/1.JRS.8.083589>.
- 492 Dal Vechio, F., Prates, I., Graziotin, F.G., Zaher, H. and Rodrigues, M.T. (2018) Phylogeography and historical
493 demography of the arboreal pit viper *Bothrops bilineatus* (Serpentes, Crotalinae) reveal multiple connections
494 between Amazonian and Atlantic rain forests. *Journal of Biogeography*, 45(10), 2415–2426.
495 <https://doi.org/10.1111/jbi.13421>.
- 496 Danielson, J.J. and Gesch, D.B. (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010).
497 [online] Sioux Falls: US Geological Survey Open-File Report 2011–1073, 26 p. Available at:
498 <https://pubs.usgs.gov/of/2011/1073/> [Accessed 02 Aug. 2019]
- 499 Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J.-C., Wieringa, J.J., Ramesh,
500 B.R., Stévant, T. and Couvreur, T.L.P. (2016) Remotely sensed temperature and precipitation data improve

- 501 species distribution modelling in the tropics. *Global Ecology and Biogeography*, 25(4), 443–454.
502 <https://doi.org/10.1111/geb.12426>.
- 503 Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S.J. and Ropelewski, C.F. (2007) Validation of
504 satellite rainfall products over East Africa's complex topography. *International Journal of Remote Sensing*,
505 28(7), 1503–1526. <https://doi.org/10.1080/01431160600954688>.
- 506 Efron, B. (1982) *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: Society for industrial
507 and Applied Mathematics.
- 508 Elith, J. and Leathwick, J.R. (2009) Species distribution models: Ecological explanation and prediction across
509 space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697.
510 <https://doi.org/10.1146/annurev.ecolsys.110308.120159>.
- 511 Fick, S.E. and Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land
512 areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>.
- 513 Fleming, K., Awange, J.L., Kuhn, M. and Featherstone, W.E. (2011) Evaluating the TRMM 3B43 monthly
514 precipitation product using gridded raingauge data over Australia. *Journal of Southern Hemisphere Earth
515 Systems Science*, 61(3): 171–184.
- 516 Franklin, J. and Miller, J.A. (2010) *Mapping species distributions: Spatial inference and prediction*. Cambridge:
517 Cambridge University Press.
- 518 Freeman, B.G. and Mason, N.A. (2015) The geographic distribution of a tropical montane bird is limited by a
519 tree: acorn woodpeckers (*Melanerpes formicivorus*) and Colombian oaks (*Quercus humboldtii*) in the Northern
520 Andes. *PLOS ONE*, 10(6), 1–13. <https://doi.org/10.1371/journal.pone.0128675>.
- 521 Goovaerts, P. (1997) *Geostatistics for natural resources evaluation*. New York: Oxford University Press.
- 522 Gräler, B., Pebesma, E. and Heuvelink, G. (2016) Spatio-temporal interpolation using gstat. *The R Journal*, 8(1),
523 204–218. <https://doi.org/10.32614/RJ-2016-014>.
- 524 Hernandez, P.A., Graham, C.H., Master, L.L. and Albert, D.L. (2006) The effect of sample size and species
525 characteristics on performance of different species distribution modeling methods. *Ecography*, 29(5), 773–785.
526 <https://doi.org/10.1111/j.0906-7590.2006.04700.x>.
- 527 Hidasi-Neto, J., Joner, D.C., Resende, F., Monteiro, L. de M., Faleiro, F.V., Loyola, R.D. and Cianciaruso, M.V.
528 (2019) Climate change will drive mammal species loss and biotic homogenization in the Cerrado Biodiversity
529 Hotspot. *Perspectives in Ecology and Conservation*, 17(2), 57–63. <https://doi.org/10.1016/j.pecon.2019.02.001>.
- 530 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A. (2005) Very high resolution interpolated
531 climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978.
532 <https://doi.org/10.1002/joc.1276>.
- 533 Hijmans, R.J., Phillips, S., Leathwick, J. and Elith, J. (2017) *dismo: species distribution modeling*. R package
534 ver. 1.1-4.

- 535 Hyndman, R.J. and Koehler, A.B. (2006) Another look at measures of forecast accuracy. *International Journal of*
 536 *Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- 537 *ibge.gov.br*, (2021). Instituto Brasileiro de Geografia e Estatística. [online] Available at:
 538 <https://www.ibge.gov.br/> [Accessed 14 Jun. 2021].
- 539 Just, A.C., Liu, Y., Sorek-Hamer, M., Rush, J., Dorman, M., Chatfield, R., Wang, Y., Lyapustin, A. and Kloog,
 540 I. (2020) Gradient boosting machine learning to improve satellite-derived column water vapor measurement
 541 error. *Atmospheric Measurement Techniques*, 13, 4669–4681. <https://doi.org/10.5194/amt-13-4669-2020>.
- 542 Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P.
 543 and Kessler, M. (2017) Climatologies at high resolution for the earth’s land surface areas. *Scientific Data*, 4,
 544 170122. <https://doi.org/10.1038/sdata.2017.122>.
- 545 Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P.
 546 and Kessler, M. (2018) Data from: Climatologies at high resolution for the earth’s land surface areas. [online]
 547 Dryad Digital Repository. <http://dx.doi.org/doi:10.5061/dryad.kd1d4>. [Accessed 27 Jul. 2019]
- 548 Körner, C. (2002) Mountain biodiversity, its causes and function: An overview. In: C. Körner and E.M. Spehn,
 549 eds., *Mountain Biodiversity: A Global assessment*. London: CRC Press, pp. 3–20.
- 550 Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J. and Scott, J.K. (2012) CliMond:
 551 global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in*
 552 *Ecology and Evolution*, 3(1), 53–64. <https://doi.org/10.1111/j.2041-210X.2011.00134.x>.
- 553 Kummerow, C., Simpson, J., Thiele, O., Barnes, W., Chang, A.T.C., Stocker, E., Adler, R.F., Hou, A., Kakar,
 554 R., Wentz, F., Ashcroft, P., Kozu, T., Hong, Y., Okamoto, K., Iguchi, T., Kuroiwa, H., Im, E., Haddad, Z.,
 555 Huffman, G., Ferrier, B., Olson, W.S., Zipser, E., Smith, E.A., Wilheit, T.T., North, G., Krishnamurti, T. and
 556 Nakamura, K. (2000) The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit.
 557 *Journal of Applied Meteorology*, 39(12), 1965–1982. [https://doi.org/10.1175/1520-](https://doi.org/10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2)
 558 [0450\(2001\)040<1965:TSOTTR>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2).
- 559 Lapola, D.M., Martinelli, L.A., Peres, C.A., Ometto, J.P.H.B., Ferreira, M.E., Nobre, C.A., Aguiar, A.P.D.,
 560 Bustamante, M.M.C., Cardoso, M.F., Costa, M.H., Joly, C.A., Leite, C.C., Moutinho, P., Sampaio, G.,
 561 Strassburg, B.B.N. and Vieira, I.C.G. (2014) Pervasive transition of the Brazilian land-use system. *Nature*
 562 *Climate Change*, 4, 27–35. <https://doi.org/10.1038/nclimate2056>.
- 563 Ledo, R.M.D. and Colli, G.R. (2017) The historical connections between the Amazon and the Atlantic Forest
 564 revisited. *Journal of Biogeography*, 44(11), 2551–2563. <https://doi.org/10.1111/jbi.13049>.
- 565 Li, J. (2019) *spm: spatial predictive modeling*. R package ver. 1.2.0.
- 566 Li, J. and Heap, A.D. (2011) A review of comparative studies of spatial interpolation methods in environmental
 567 sciences: performance and impact factors. *Ecological Informatics*, 6(3–4), 228–241.
 568 <https://doi.org/10.1016/j.ecoinf.2010.12.003>.

- 569 Lima-Ribeiro, M.S., Varela, S., González-Hernández, J., Oliveira, G., Diniz-Filho, J.A.F., Peterson, A.T. and
570 Terribile, L.C. (2020) The ecoClimate database. [online] Available at: <http://ecoclimate.org> [Accessed 01 Aug.
571 2020].
- 572 Lima-Ribeiro, M.S., Varela, S., González-Hernández, J., Oliveira, G., Diniz-Filho, J.A.F. and Terribile, L.C.
573 (2015) EcoClimate: a database of climate data from multiple models for past, present, and future for
574 macroecologists and biogeographers. *Biodiversity Informatics*, 10, 1–21. <https://doi.org/10.17161/bi.v10i0.4955>.
- 575 Ma, L., Zhang, G. and Lu, E. (2018) Using the Gradient Boosting Decision Tree to improve the delineation of
576 hourly rain areas during the summer from Advanced Himawari Imager data. *Journal of Hydrometeorology*,
577 19(5), 761–776. <https://doi.org/10.1175/JHM-D-17-0109.1>.
- 578 Manz, B., Buytaert, W., Zulkafli, Z., Lavado, W., Willems, B., Robles, L.A. and Rodríguez-Sánchez J.-P. (2016)
579 High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. *Journal of*
580 *Geophysical Research: Atmospheres*, 121(3), 1190–1207. <https://doi.org/10.1002/2015JD023788>.
- 581 Massetti, E., Guiducci, R., de Oliveira, A.F., Mendelsohn, R. (2013) The impact of climate change on the
582 Brazilian Agriculture: A Ricardian study at microregion level. *SSRN Electronic Journal*, 200, 1–31.
583 <https://doi.org/10.2139/ssrn.2637287>.
- 584 Mateus, P., Borma, L.S., da Silva, R.D., Nico, G. and Catalão, J. (2016) Assessment of two techniques to merge
585 ground-based and TRMM rainfall measurements: a case study about Brazilian Amazon Rainforest. *GIScience &*
586 *Remote Sensing*, 53(6), 689–706. <https://doi.org/10.1080/15481603.2016.1228161>.
- 587 Meinshausen, N. (2006) Quantile regression forests. *Journal of Machine Learning Research*, 7(35), 983–999.
- 588 Meinshausen, N. (2017) *quantregForest: quantile regression forests*. R package ver. 1.3-7.
- 589 Miao, C., Ashouri, H., Hsu, K.-L., Sorooshian, S. and Duan, Q. (2015) Evaluation of the PERSIANN-CDR daily
590 rainfall estimates in capturing the behavior of extreme precipitation events over China. *Journal of*
591 *Hydrometeorology*, 16(3), 1387–1396. <https://doi.org/10.1175/JHM-D-14-0174.1>.
- 592 Mittermeier, R.A., Mittermeier, C.G. and Gil, P.R. (1997) Megadiversity: Earth's biologically wealthiest nations.
593 Mexico: CEMEX.
- 594 Molotoks, A., Stehfest, E., Doelman, J., Albanito, F., Fitton, N., Dawson, T.P. and Smith, P. (2018) Global
595 projections of future cropland expansion to 2050 and direct impacts on biodiversity and carbon storage. *Global*
596 *Change Biology*, 24(12), 5895–5908. <https://doi.org/10.1111/gcb.14459>.
- 597 Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. and Kent, J. (2000) Biodiversity hotspots
598 for conservation priorities. *Nature*, 403, 853–858. <https://doi.org/10.1038/35002501>.
- 599 Nastos, P.T., Kapsomenakis, J. and Philandras, K.M. (2016) Evaluation of the TRMM 3B43 gridded
600 precipitation estimates over Greece. *Atmospheric Research*, 169(B), 497–514.
601 <https://doi.org/10.1016/j.atmosres.2015.08.008>.

- 602 Nesbitt, S.W. and Anders, A.M. (2009) Very high resolution precipitation climatologies from the Tropical
603 Rainfall Measuring Mission precipitation radar. *Geophysical Research Letters*, 36(15), L15815.
604 <https://doi.org/10.1029/2009GL038026>.
- 605 New, M., Lister, D., Hulme, M. and Makin, I. (2002). A high-resolution data set of surface climate over global
606 land areas. *Climate Research*, 21(1), 1–25. <https://doi.org/10.3354/cr021001>
- 607 Nychka, D., Furrer, R., Paige, J. and Sain, S. (2017) *fields: tools for spatial data*. Boulder: University
608 Corporation for Atmospheric Research.
- 609 Overbeck, G.E., Vélez-Martin, E., Scarano, F.R., Lewinsohn, T.M., Fonseca, C.R., Meyer, S.T., Müller, S.C.,
610 Ceotto, P., Dadalt, L., Durigan, G., Ganade, G., Gossner, M.M., Guadagnin, D.L., Lorenzen, K., Jacobi, C.M.,
611 Weisser, W.W. and Pillar, V.D. (2015) Conservation in Brazil needs to include non-forest ecosystems. *Diversity*
612 *and Distributions*, 21(12), 1455–1460. <https://doi.org/10.1111/ddi.12380>.
- 613 Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7), 683–
614 691. <https://doi.org/10.1016/j.cageo.2004.03.012>.
- 615 Qiao, H., Lin, C., Jiang, Z. and Ji, L. (2015) Marble Algorithm: a solution to estimating ecological niches from
616 presence-only records. *Scientific Reports*, 5, 14232. <https://doi.org/10.1038/srep14232>.
- 617 R Core Team. (2020) *R: a language and environment for statistical computing*. Vienna: R Foundation for
618 Statistical Computing.
- 619 Ramoni-Perazzi, P. (2016) *Modelado de la distribución espacial de la precipitación media mensual de los Andes*
620 *Venezolanos*. Master. University of the Andes.
- 621 Ramoni-Perazzi, P., Passamani, M., Thielen, D. and Padovani, C. (in press) O tradicional ou o moderno? Uma
622 visão da informação da rede de estações meteorológicas brasileiras. *Revista Brasileira de Meteorologia*.
623 <https://doi.org/10.1590/0102-77863630115>
- 624 Reis, A.A.D., Werner, J.P.S., Silva, B.C., Figueiredo, G.K.D.A., Antunes, J.F.G., Esquerdo, J.C.D.M., Coutinho,
625 A., Lamparelli, R., Rocha, J. and Magalhães, P. (2020) Monitoring pasture aboveground biomass and canopy
626 height in an integrated crop-livestock system using textural information from PlanetScope imagery. *Remote*
627 *Sensing*, 12(16), 2534. <https://doi.org/10.3390/rs12162534>.
- 628 Rödder, D. and Lötters, S. (2010) Explanative power of variables used in species distribution modelling: an issue
629 of general model transferability or niche shift in the invasive Greenhouse frog (*Eleutherodactylus planirostris*).
630 *Naturwissenschaften*, 97, 781–796. <https://doi.org/10.1007/s00114-010-0694-7>.
- 631 Silveira, M.H.B., Mascarenhas, R., Cardoso, D. and Batalha-Filho, H. (2019) Pleistocene climatic instability
632 drove the historical distribution of forest islands in the northeastern Brazilian Atlantic Forest. *Palaeogeography*,
633 *Palaeoclimatology, Palaeoecology*, 527, 67–76. <https://doi.org/10.1016/j.palaeo.2019.04.028>.
- 634 Soares-Filho, B., Rajão, R., Macedo, M., Carneiro, A., Costa, W., Coe, M., Rodrigues, H. and Alencar, A.
635 (2014) Cracking Brazil's Forest Code. *Science*, 344(6182), 363–364. <https://doi.org/10.1126/science.1246663>.

- 636 Soutan, A. and Safi, K. (2017) The interplay of various sources of noise on reliability of species distribution
637 models hinges on ecological specialisation. *PLOS ONE*, 12(11), e0187906.
638 <https://doi.org/10.1371/journal.pone.0187906>.
- 639 Tao, H., Fischer, T., Zeng, Y. and Fraedrich, K. (2016) Evaluation of TRMM 3B43 precipitation data for
640 drought monitoring in Jiangsu Province, China. *Water*, 8(6), 221. <https://doi.org/10.3390/w8060221>
- 641 Title, P.O. and Bemmels, J.B. (2018) ENVIREM: an expanded set of bioclimatic and topographic variables
642 increases flexibility and improves performance of ecological niche modeling. *Ecography*, 41, 291–307.
643 <https://doi.org/10.1111/ecog.02880>.
- 644 Tozer, C.R., Kiem, A.S. and Verdon-Kidd, D.C. (2012) On the uncertainties associated with using gridded
645 rainfall data as a proxy for observed. *Hydrology and Earth System Sciences*, 16, 1481–1499.
646 <https://doi.org/10.5194/hess-16-1481-2012>.
- 647 Tres, A., Tetto, A.F., Soares, R.V., Wendling, W.T. and Santos, G.H. (2020) Ecological life zones of Brazil.
648 *Floresta*, 50, 1575–1584. <https://doi.org/10.5380/ufv.v50i3.64833>.
- 649 TRMM (2018) [Tropical Rainfall Measuring Mission]. TRMM (TMPA/3B43) Rainfall Estimate L3 1 month
650 0.25 degree x 0.25 degree V7. [online] Greenbelt: Goddard Earth Sciences Data and Information Services
651 Center.
- 652 Vega, G.C., Pertierra, L.R. and Olalla-Tárraga, M.Á. (2017) MERRAclim, a high-resolution global dataset of
653 remotely sensed bioclimatic variables for ecological modelling. *Scientific Data*, 4, 170078.
654 <https://doi.org/10.1038/sdata.2017.78>.
- 655 Vega, G.C., Pertierra, L.R. and Olalla-Tárraga, M.Á. (2018) Data from: MERRAclim, a high-resolution global
656 dataset of remotely sensed bioclimatic variables for ecological modelling. [online] Dryad Digital Repository.
657 <https://doi.org/10.5061/dryad.s2v81>. [Accessed 19 Jul. 2019]
- 658 Waltari, E., Schroeder, R., McDonald, K., Anderson, R.P. and Carnaval, A. (2014) Bioclimatic variables derived
659 from remote sensing: assessment and application for species distribution modelling. *Methods in Ecology and*
660 *Evolution*, 5(10), 1033–1042. <https://doi.org/10.1111/2041-210X.12264>.
- 661 Wang, G., Zhang, P., Liang, L. and Zhang, S. (2017) Evaluation of precipitation from CMORPH, GPCP-2,
662 TRMM 3B43, GPCC, and ITPCAS with ground-based measurements in the Qinling-Daba Mountains, China.
663 *PLOS ONE*, 12(10), e0185147. <https://doi.org/10.1371/journal.pone.0185147>.
- 664 Wolff, D.B. and Fisher, B.L. (2008) Comparisons of Instantaneous TRMM Ground Validation and Satellite
665 Rain-Rate Estimates at Different Spatial Scales. *Journal of Applied Meteorology and Climatology*, 47(8), 2215–
666 2237. <https://doi.org/10.1175/2008JAMC1875.1>.
- 667 Xu, T. and Hutchinson, M.F. (2013) New developments and applications in the ANUCLIM spatial climatic and
668 bioclimatic modelling package. *Environmental Modelling & Software*, 40, 267–279.
669 <https://doi.org/10.1016/j.envsoft.2012.10.003>.

- 670 Zanin, M., Tessarolo, G., Machado, N. and Albernaz, A. (2017) Climatically-mediated landcover change:
671 Impacts on Brazilian territory. *Anais da Academia Brasileira de Ciências*, 89(2), 939–952.
672 <https://doi.org/10.1590/0001-3765201720160226>.
- 673 Zhao, H., Yang, S., Wang, Z., Zhou, X., Luo, Y. and Wu, L. (2015) Evaluating the suitability of TRMM satellite
674 rainfall data for hydrological simulation using a distributed hydrological model in the Weihe River catchment in
675 China. *Journal of Geographical Sciences*, 25, 177–195. <https://doi.org/10.1007/s11442-015-1161-3>.
- 676 Zilli, M., Scarabello, M., Soterroni, A.C., Valin, H., Mosnier, A., Leclère, D., Havlík, P., Kraxner, F., Lopes,
677 M.A. and Ramos, F.M. (2020) The impact of climate change on Brazil's agriculture. *Science of The Total*
678 *Environment*, 740, 139384. <https://doi.org/10.1016/j.scitotenv.2020.139384>.

ARTIGO 3 – EVALUATING THE BRAZILIAN SYSTEM OF STRICTLY PROTECTED AREAS USING OWLS AS AN INDICATOR GROUP

Short running title: Brazilian protected areas and owls

Paolo Ramoni-Perazzi^{1,2}, Marcelo Passamani¹, Dirk Thielen³, Luis Fernando Landeo Atmi⁴ & Marco Aurelio Arizapana⁴

¹Departamento de Ecologia e Conservação, Instituto de Ciências Naturais, Universidade Federal de Lavras, Lavras (MG), Brasil, CEP 37.200-900.

²Modeling and Simulation Centre (CeSiMo), Faculty of Engineering, University of the Andes, Merida 5101, Venezuela.

³Laboratory of Landscape Ecology and Climate, Venezuelan Institute for Scientific Research (IVIC), Caracas, Venezuela.

⁴Research Group on Remote Sensing and Mountain Ecology, School of Engineering and Environmental Management, National Autonomous University of Huanta, Ayacucho, Peru.

(Artigo submetido para o jornal *Perspectives in Ecology and Evolution*)

PRP (corresponding author) email: rpaolo1967@gmail.com

PRP ORCID: 0000-0001-9182-3798

MP ORCID: 0000-0002-0940-4074

DT ORCID: 0000-0003-4102-1927

LFLA ORCID: 0000-0001-8393-5562

MAA: ORCID: 0000-0001-5923-0300

ABSTRACT

Conservation efforts guided by biodiversity hotspots of indicator taxa has become a common approach. This requires a crystal-clear knowledge on the taxonomy and distribution of such indicators, often a troublesome task in the Tropics, especially for rare or secretive taxa. We assessed the potential distribution of 21 species and 21 subspecies of Brazilian Strigidae through Species Distribution Modelling (SDM). We (1) generated SDMs for each species and

subspecies, (2) evaluated the niche similarity among subspecies, (3) elaborated species' richness maps, and (4) contrasted such information to the strict protection areas in Brazil. With 81% of the Brazilian species recorded, both the Atlantic Forest and the Cerrado have the highest richness, followed by the Amazonia (67%), Pampa (62%), Caatinga (57%) and Pantanal (48%). However, the comparison of the recorded and predicted richness suggests overall incomplete inventories, especially in the Pampa. On the other hand, subspecies showed marked niches divergencies, suggesting that the actual Strigidae species richness is underestimated in Brazil. Cerrado and Atlantic Forest are the most threatened biomes, with preservation areas relatively small and scattered. Thus, our study is a response to the urgent call to explore owl lineage diversification in Brazil to improve biodiversity-related conservation efforts.

Keywords: Strigidae, priority areas, species distribution models, niche equivalence, megadiversity

INTRODUCTION

Knowing where the species are found has always been important and gained relevance during the Anthropocene. However, given the faulty funding levels (Gallo-Cajiao et al., 2018) and the inadequate occurrence data available (Hortal et al., 2015) it can be troublesome, especially in the tropics. Accordingly, species distribution models (hereafter SDMs) became a reliable and more convenient tools compared to other more traditional procedures (Mainali et al., 2020). These models combine occurrence data, environmental information and computer algorithms to generate probabilistic distributions of species in space or environment while reducing both false negatives and false positives errors (Mendes et al., 2020). Therefore, generating, gathering and standardizing species distribution information, useful to determine threats to the environment or critical areas, is crucial in conservation management (Sánchez de Dios et al., 2017).

Among birds, raptors constitute a paradox since they are comparatively under-studied, given their low fertility and abundances, while exerting great appeal for financial support, given their ecological roles as tops predators, symbolism and threat levels (Donázar et al., 2016). Among raptors, true owls (Strigidae) are characterized by their global high diversity (> 220 species; Gill et al., 2021), interspecific variation in size, distribution ranges, habitat specialization, and responses to habitat structure (Barros and Cintra, 2009) and alteration

(Enríquez, 2017), making them a reliable surrogate group to identify areas of conservation priority.

However, owls are overlooked in fauna inventories given their nocturnal habits (Silva and Medellín, 2001), restricting the knowledge on their biology and, probably, their representation in the official lists of threatened fauna in countries such as Brazil (Motta-Junior et al., 2017; Motta-Junior and Braga, 2012). Even the information on their distribution is sparse, anecdotal, insufficiently detailed (Motta-Junior and Braga, 2012) and probably incomplete.

Brazil harbors some 21 recognized species of Strigidae (Gill et al., 2021), including several "near-endemics" such as the East Brazilian pygmy owl (*Glaucidium minutissimum*), the black-capped screech owl (*Megascops atricapilla*), the long-tufted screech-owl (*M. sanctaecatarinae*), the tawny-browed owl (*Pulsatrix koeniswaldiana*) and the rusty-barred owl (*Strix hylophila*) (Motta-Junior et al., 2017), and the endemic Pernambuco pygmy owl (*Glaucidium mooreorum*), critically endangered (BirdLife International, 2019) or even extinct (Pereira, 2010). However, such biodiversity is greatly underestimated, as suggested by the proposition of new owl species for Brazil based on molecular and bioacoustics traits by Dantas et al. (2021).

The need to complete our understanding of the biology, ecology and distribution of owls in Brazil, especially under scenarios imposed by the Anthropocene, places their SDMs as a challenging and urgent scientific and conservation task. Here, for mainland Brazil, we: (1) generated SDMs for each species and subspecies based on a maximum entropy approach; (2) evaluated niche similarities between conspecific subspecies; (3) created species richness maps; (4) determined the biodiversity hotspots; and (5) identified priority conservation areas contrasting them against the existing network of strictly protected areas.

METHODS

Study área

Brazil is a continent-sized country of contrasting topographic features (plain landscapes in the Amazon basin, or rugged topography in the southeast), a wide latitudinal extent and continentality. Concomitantly, there are three major climatic types as defined by Köppen (Alvares et al., 2013): humid subtropical, tropical rainy and tropical dry. This is consistent with the presence of six major terrestrial biomes: Amazonia, Atlantic Forest, Caatinga,

Cerrado, Pampa and Pantanal (IBGE, 2019). Nevertheless, and acknowledging the importance of dispersal in SDMs (Barve et al., 2011), we considered Brazil as an orographic continuum for flying species given that: (1) the lack of massive and steeply orographic barriers; (2) the system of rivers has a limited influence in shaping bird species distributions (Fluck et al., 2020; Santorelli et al., 2018).

Species and subspecies account

We followed the International Ornithological Committee v. 11.1 (Gill et al., 2021), which considers 21 valid species and 21 conspecific subspecies. To model the distributions of the subspecies separately, and given the lack of more comprehensive sources, we split the occurrences following Gill et al. (2021), and the maps from www.xeno-canto.org (Table 3.1).

Environmental covariates

Bioclimatic covariates have been successfully used in SMDs since their beginning (Booth et al., 2014) and perhaps reflect physiological constraints. Soil type and geology have also been favorably used in SDMs for Neotropical flying vertebrates, likely echoing deeper ecological and historical constraints (Ramoni-Perazzi et al., 2020). Thus, we used BrazilClim, a database of bioclimatic covariates created ad hoc for the Brazilian mainland (Ramoni-Perazzi et al., 2021), elevation information (GMTED2010; Danielson and Gesch, 2011), geological substrate (hereafter geology; Gómez Tapias et al., 2019), and soil type (Hengl et al., 2017). All variables were used at (or resampled to) 30 arc seconds resolution.

Table 3.1 – Account of the species and subspecies occurring in Brazil, according to the International Ornithological Committee v. 11.1 (Gill et al. 2021). “Distribution” refers to the range described in the aforementioned source, used jointly to biomes, major hydrographic basins and reinforced by other sources such www.xeno-canto.org, to proximately assign the occurrences to the different subspecies (our “interpretation”), to create the occurrences databases used to model the subspecies distributions.

English name	Scientific name	Distribution	Interpretation
Buff-fronted Owl	<i>Aegolius harrisii</i>		
	<i>A. h. iheringi</i>	“Paraguay, SE Brazil, Uruguay and NE Argentina”	
Striped Owl	<i>Asio clamator</i>		
	<i>A. c. clamator</i>	“Colombia and Venezuela to E Peru and C, NE Brazil”	Caatinga and Amazonia
	<i>A. c. midas</i>	“E Bolivia to Paraguay, S Brazil, Uruguay and N Argentina”	Cerrado, Atlantic Forest and Pampa
Short-eared Owl	<i>Asio flammeus</i>		
	<i>A. f. suinda</i>	“S Peru and S Brazil to Tierra del Fuego”	
Stygian Owl	<i>Asio stygius</i>		
Burrowing Owl	<i>Athene cunicularia</i>		
	<i>A. c. minor</i>	“C, SE Venezuela, S Guyana and N Brazil”	Amazonia
	<i>A. c. cunicularia</i>	“S Bolivia, Paraguay and S Brazil to Tierra del Fuego”	Pampa
	<i>A. c. grallaria</i>	“E and C Brazil”	Remnant biomes
Great Horned Owl	<i>Bubo virginianus</i>		
	<i>B. v. deserti</i>	“NE Brazil (NC Bahia)”	Caatinga
	<i>B. v. nacurutu</i>	“E Colombia through the Guianas to N, E Brazil, Argentina, Bolivia and C Peru”	Remnant biomes
Ferruginous Pygmy Owl	<i>Glaucidium brasilianum</i>		

	<i>G. b. ucayalae</i>	“Amazonia”	Amazonia
	<i>G. b. brasilianum</i>	“E Brazil to NE Argentina”	Remnant biomes
Amazonian Pygmy Owl	<i>Glaucidium hardyi</i>		
East Brazilian Pygmy Owl	<i>Glaucidium minutissimum</i>		
Pernambuco Pygmy Owl	<i>Glaucidium mooreorum</i>		
Crested Owl	<i>Lophotrix cristata</i>		
Black-capped Screech Owl	<i>Megascops atricapilla</i>		
Tropical Screech Owl	<i>Megascops choliba</i>		
	<i>M. c. choliba</i>	“S Brazil and E Paraguay”	Paraná river basin
	<i>M. c. cruciger</i>	“E Colombia and E Peru through Venezuela, the Guianas to NE Brazil”	Amazonia
	<i>M. c. uruguayii</i>	“SE Brazil, Uruguay and NE Argentina”	Pampa
	<i>M. c. decussatus</i>	“C, E Brazil”	Remnant areas
Foothill Screech Owl	<i>Megascops roraimae</i>		
Long-tufted Screech Owl	<i>Megascops sanctaecatarinae</i>		
Tawny-bellied Screech Owl	<i>Megascops watsonii</i>		
	<i>M. w. watsonii</i>	“N of the Amazon”	N of the Amazon river
	<i>M. w. usta</i>	“S of the Amazon”	S of the Amazon river
Tawny-browed Owl	<i>Pulsatrix koeniswaldiana</i>		
Spectacled Owl	<i>Pulsatrix perspicillata</i>		
	<i>P. p. perspicillata</i>	“E Colombia through the Guianas and Amazonia”	Amazonia
	<i>P. p. pulsatrix</i>	“Paraguay, E Brazil and NE Argentina”	Atlantic Forest
Black-banded Owl	<i>Strix huhula</i>		
	<i>S. h. huhula</i>	“Colombia, Venezuela and the Guianas S to E Peru, E Bolivia, NW Argentina and EC Brazil”	Amazonia and Cerrado

	<i>S. h. albomarginata</i>	“SE Brazil, Paraguay and NE Argentina”	Atlantic Forest
Mottled Owl	<i>Strix virgata</i>		
	<i>S. v. superciliaris</i>	“NC, NE Brazil”	Amazonia
	<i>S. v. borelliana</i>	“S Brazil, Paraguay and NE Argentina”	Atlantic Forest

We performed all the analyses using R 3.6.3 (R Core Team, 2020). To remove collinearity, we reduced the number of continuous (bioclimatic + elevation) covariables through a principal component analysis, keeping the first four components, whose eigenvalues were higher than one and explained 90.4% of the variance (Supplementary material A, Appendix 1, Fig. A). The first component (PC1) can be interpreted as a contrast between the temperature during the most extreme conditions and its seasonality/variability (Supplementary material A, Appendix 1, Fig. B1). Similarly, the second component (PC2) contrasts the precipitation during extreme conditions and its seasonality. The third component (PC3) involves the effect of water availability since involves temperatures under extreme conditions/seasonality and precipitation during the warmest quarter. Finally, the fourth component (PC4) has a weak negative association between Mean Diurnal Temperature Range (Bio 02) and elevation.

Spatial distribution models

We gathered information on the distribution (= occurrences) of the Strigidae in Brazil from: (1) skin specimens deposited in several museums according to the Global Biodiversity Information Facility; (2) more than 164 publications in peer-reviewed literature regarding taxonomic assessments, fauna inventories or owl biology; and (3) field records from the bioacoustics database www.xeno-canto.org. We provide the citations of these sources in Supplementary material A, Appendix 2. The quality of the geographic coordinates varied from GPS recordings until those of the nearest town listed on the specimens' labels, and we corroborated each one through different sources such as www.geonames.org.

To our knowledge, there are no records for the buff-fronted owl (*Aegolius harrisi*), and only few for the foothill screech owl (*Megascops roraimae*), in Northern Brazil. Thus, we included the foreign localities known for both species in the very vicinity of the Northern border, after reassigning them with coordinates within their respective closest Brazilian territory. The Pernambuco pygmy-owl is known from two localities (Silva et al., 2002), to which we added eight random points located within a polygon resulting from two merged circles, each centered in one of the known localities and radius equaling the distance between both, clipped by the neighbor coastline. We excluded a record of the short-eared owl (*Asio flammeus*) in the Roraima

State (wikiaves.com.br; consulted on April 10, 2021), likely belonging to the subspecies *A. f. pallidicaudus* from “Venezuela, Guyana and Suriname” (Gill et al., 2021).

The geographical and environmental clustering of field surveys, known as spatial autocorrelation, can negatively affect the performance of the SDMs (Veloz, 2009), for which some authors remove those records under the same environmental conditions within an arbitrary distance (Delgado-Jaramillo et al., 2020). Thus, we created two datasets for each taxon, one including all the records and another excluding those closer than 25 km, and computed empirical entrograms for each one using “elsa” (Naimi et al., 2019).

We used “ENMeval” (Muscarella et al., 2014), a package based on Maxent (Phillips et al., 2017), that automatically splits data into training/test subsets, performs SDMs across a range of settings, and calculates diverse evaluation metrics. For each taxon, we ran 10 models, each one after partitioning occurrences in testing and training bins using a 10-fold cross-validation scheme (Fielding and Bell, 1997). For each run, we created 10 000 pseudoabsence points distributed randomly throughout Brazil and selected the model with the lowest Akaike information criterion corrected for small samples sizes ($\Delta AICc=0$) as the best one, since it reflects both model goodness-of-fit and complexity (Warren and Seifert, 2011) and less overfitting (Muscarella et al., 2014).

Different habitat suitability thresholds may disagree in terms of suitable areas and omission errors (Liu et al., 2016). Thus, for each taxon, we plotted the extension of the predicted area (in number of 30 arcsec by 30 arcsec pixels) against the number of omissions and compared across taxa the performance of the different thresholds, keeping the one that consistently provided the lowest values for both measurements. The final binary models combined the best models ($\Delta AICc = 0$) and the threshold with the lower number of omissions within the smallest predicted area. We stacked these binary distributions to create two maps of taxa richness for (a) the 21 species evaluated (hereafter *sensu stricto* map), (b) the 12 monotypic species (including polytypic ones represented by only one subspecies in Brazil) and 21 subspecies (henceforth *sensu lato* map). We overlapped the protected areas distributions corresponding to IUCN's categories I to IV (according to Protected Planet, 2021) on each richness map.

Subspecies niche analysis

We used ‘phyloclim’ (Heibl and Calenge, 2018) to quantify the niche overlap (measurements D and I) for each pair of conspecific subspecies based on predictions (Warren et al., 2008). Both indexes report whether the average agreement of environmental requirements calculated between subspecies pairs is greater or smaller than expected if such subspecies were independent of each other, ranging from zero (no overlap) to one (identical niches). Then, we assessed the null hypothesis of niche identity, evaluating the equivalency between pairs of predictions, comparing the respective observed values of D and I against those generated through 100 pseudoreplicates, assigning the occurrence points for both subspecies to one lineage or the other at random to simulate the potential overlap of a group of points occurring across a given geographic space (Warren et al., 2008).

Moreover, we used background randomization tests to contrast the observed niche overlap values against a null distribution of 100 overlap values generated by comparing the predictions of one subspecies against the predictions created from random points from the geographic space of the other subspecies (Warren et al., 2008). We considered that: (1) there was niche conservation when the backgrounds overlapped and niche distance was not significantly different from zero; (2) there was niche divergence when the distance between both subspecies was significantly different from zero, independently of backgrounds’ overlap.

RESULTS

The occurrences

We gathered 2600 occurrence localities, after removing redundant ones (Supplementary Material B). Most occurrences were from the Atlantic Forest (46%), followed by the Amazonia (21%), Cerrado (17%), Caatinga (8%), Pampa (4%), and Pantanal (3%; Fig. 3.1; Table 3.2). The Southeastern Region, ~11% of the Brazilian territory, was the best represented with 34% of the records, while the Northern region, 45% of the Brazilian territory, only provided 25% of the records. Both the Atlantic Forest and Cerrado were the most diverse biomes harboring records for 81% of the species each, followed by the Amazonia (67%), Pampa (62%), Caatinga (57%), and Pantanal (48%).

Spatial distribution models

Excluding the case of the striped owl (*Asio clamator*), the empirical entrograms showed no substantial variations when excluding those localities closer than 25 km (Supplementary Material A, Appendix 1, Fig. C). Thus, we performed the SDMs including all the occurrences, except for the aforementioned taxa. Such a removal resulted in 2512 occurrence localities.

We provide detailed information of the models in Supplementary material A, Appendix 3, Table A. In concordance with Liu et al. (2016), we found that thresholds based on sensitivity-specificity outperformed the others (Supplementary material Appendix 1, Fig. D). For such a reason, we used the equal sensitivity-specificity threshold to create the binary maps.

In all cases, the climate-based variables were the most important predictors of suitability (Table 3.3), with the temperature-based PC1 ranking highest for percentage contribution for 34 of the taxa studied, followed by precipitation-based PC3 (three taxa), and PC2 and PC4 (two taxa each).

Figure 3.1 – Records of Brazilian Strigidae used in the modeling process. Biome colors correlates with the abundance of records (the darkest the most abundant). State and countries acronyms follow ISO 3166-2 and ISO 3166-1 alpha-2, respectively.

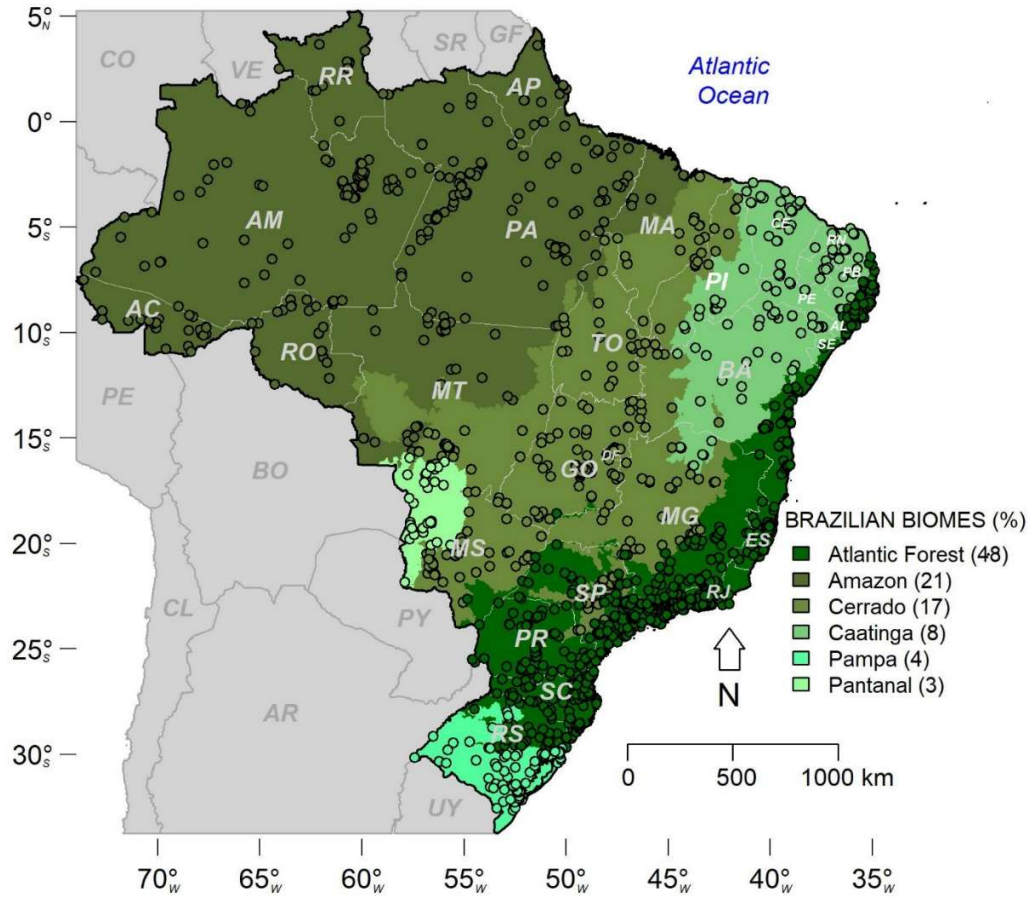


Table 3.2 – Numbers of records, omissions, number of species (sensu stricto) recorded and predicted and their differences for the different Brazilian biomes. The numbers in parentheses represent percentages, that in the case of records are based on the total of 2512 occurrences (2504 from collections/literature + eight generated for *Glaucidium mooreorum*) or the 21 species.

Biome	Records	Omissions	Richness _{rec}	Richness _{pred}	Difference
Amazonia	539 (21)	251(38)	14(67)	14(67)	0
Atlantic Forest	1158 (46)	74(11)	17(81)	18(86)	+1
Caatinga	209 (8)	101(15)	12(57)	13(62)	+1
Cerrado	434 (17)	199(30)	17(81)	18(86)	+1
Pampa	105 (4)	24(4)	13(62)	16(76)	+3
Pantanal	67 (3)	17(3)	10(48)	9(43)	+1

The predictions of the monotypic species fitted the best their traditionally reported distributions, having lower omission percentages (median 16%, ranging between 0% for *G. mooreorum*, to 38%, for *Lophotrix cristata*; Supplementary material A, Appendix 1, Fig. E) compared to the polytypic ones (median 30%, from 22% for *Asio clamator* to 42% for *Strix huhula*). These omissions usually felt outside the corresponding most represented biome (often, outside the Atlantic Forest). Besides, we found a general tendency towards fitting improvements after running models based on occurrences of the respective subspecies (median 23% of omissions, from 6% in *Megascops choliba uruguayi* and, exceptionally, 42% for *Asio clamator clamator*). For the endemic and probably extinct Pernambuco pygmy-owl (*G. mooreorum*), our models predicted a very restricted range around both known localities, but also two additional separated spots in forested areas along the rivers Jaguaribe and Gramame (João Pessoa Municipality, Paraíba State).

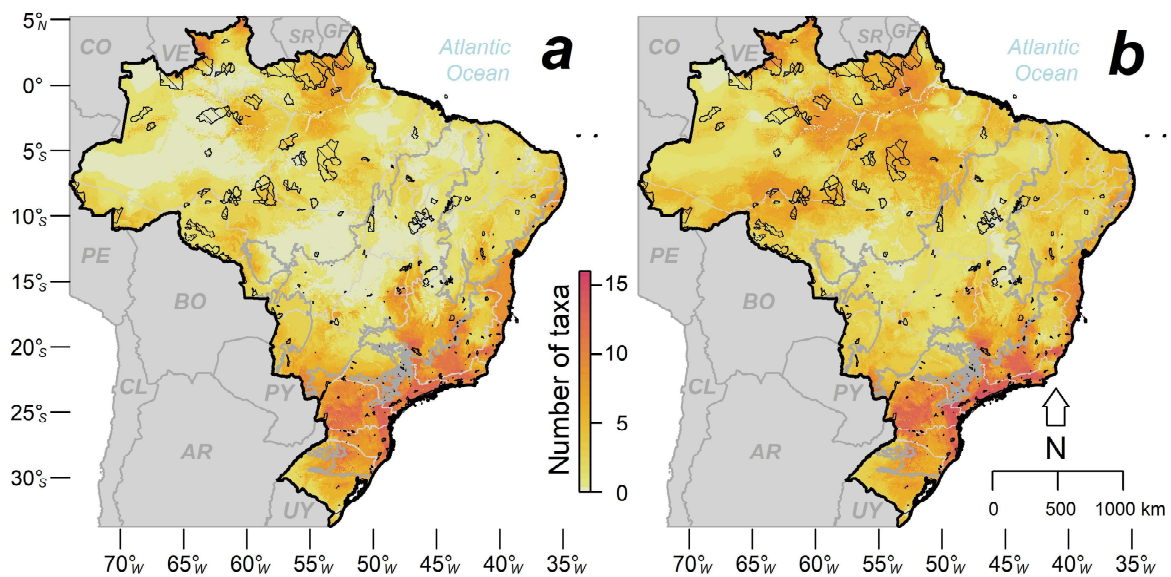
Table 3.3 – The mean \pm standard deviation of the percent of contribution and permutation importance of the explanatory variables to the SDMs of each species and subspecies. The percent of contribution is the gain of the model by including a given environmental covariate at each step of the Maxent algorithm. The permutation importance is the influence of each environmental covariate on the final model, measured through the random permutation of the values of such covariate among the training points and measuring the resulting decrease in the training area under the curve. Highest scores for each taxon in bold.

Species/subspecies	Percent of contribution						Permutation importance					
	PC1	PC2	PC3	PC4	Geology	Soil	PC1	PC2	PC3	PC4	Geology	Soil
<i>Aegolius harrisi</i>	37.3\pm1.8	10.7 \pm 0.7	5.7 \pm 0.4	33.5 \pm 1.4	1.8 \pm 0.7	11.2 \pm 0.9	40.8\pm1.9	20.0 \pm 2.6	2.9 \pm 0.6	29.2 \pm 2.8	2.4 \pm 0.8	4.8 \pm 0.8
<i>Asio clamator</i>	71.2\pm1.2	7.7 \pm 0.9	6.4 \pm 0.7	2.7 \pm 0.8	7.8 \pm 0.6	4.1 \pm 0.6	46.9\pm3.6	30.2 \pm 3.1	10.3 \pm 1.9	2.5 \pm 0.6	6.3 \pm 1	3.7 \pm 1.2
<i>A. c. clamator</i>	12.2 \pm 1.2	39.5\pm1.7	24.7 \pm 1.2	5.8 \pm 0.7	17.8 \pm 1.4	0 \pm 0	16.9 \pm 11.7	49.8\pm7.9	12.0 \pm 13.8	3.2 \pm 3.9	18.1 \pm 11.2	0 \pm 0
<i>A. c. midas</i>	71.1\pm1.7	10.3 \pm 1	4.6 \pm 1	3.1 \pm 0.5	8.1 \pm 0.5	2.8 \pm 0.7	45.5\pm3.4	33.7 \pm 3.1	9.1 \pm 1.3	4.0 \pm 0.7	5.6 \pm 0.9	2.1 \pm 0.7
<i>Asio flammeus</i>	87.5\pm1.2	1.3 \pm 0.4	0 \pm 0	2.4 \pm 0.2	1.5 \pm 0.3	7.4 \pm 1.1	91.1\pm2.2	0.1 \pm 0.2	0 \pm 0	2.1 \pm 1	1.3 \pm 1.1	5.4 \pm 1.8
<i>Asio stygius</i>	53.0\pm4.1	19.3 \pm 1.3	14.3 \pm 2	6.3 \pm 3.5	3.4 \pm 1.4	3.7 \pm 0.6	39.4\pm5.8	28 \pm 3.9	24.5 \pm 3.6	1.2 \pm 0.9	4.0 \pm 0.9	2.9 \pm 1.5
<i>Athene cucularia</i>	76.2\pm0.8	5.8 \pm 0.5	8.8 \pm 0.6	3.2 \pm 0.5	4.8 \pm 0.7	1.2 \pm 0.3	61.9\pm1.3	15.5 \pm 0.9	9.9 \pm 1.2	4.5 \pm 1.3	5.4 \pm 0.8	2.7 \pm 0.7
<i>A. c. minor</i>	37.8\pm2.0	18.8 \pm 2.4	37.7\pm1.5	5.6 \pm 1.9	0 \pm 0.1	0 \pm 0	38.8\pm29.6	21.9 \pm 12.1	34.7 \pm 17.7	4.5 \pm 2.4	0.1 \pm 0.2	0 \pm 0
<i>A. c. cucularia</i>	72.4\pm0.8	8.2 \pm 0.8	11.1 \pm 0.8	3.5 \pm 0.5	3.4 \pm 0.5	1.5 \pm 0.5	65.1\pm2.4	14.2 \pm 1.7	11.8 \pm 1.5	3.8 \pm 0.8	3.5 \pm 0.3	1.6 \pm 0.5
<i>A. c. grallaria</i>	40.2\pm1.3	27.6 \pm 1.5	0 \pm 0	32.2 \pm 1.2	0 \pm 0	0 \pm 0	34.2\pm12.4	34.2\pm8.4	0 \pm 0	31.6 \pm 15.6	0 \pm 0	0 \pm 0
<i>Bubo virginianus</i>	52.5\pm2.8	2.9 \pm 0.9	17.2 \pm 1.5	9.3 \pm 1.7	7.0 \pm 0.8	11 \pm 1.8	41.0\pm5.9	5.2 \pm 2.3	22.6 \pm 6.2	16.5 \pm 5.3	2.4 \pm 3.1	12.4 \pm 3.8
<i>B. v. deserti</i>	0 \pm 0	31.9 \pm 0.8	64.1\pm0.8	0 \pm 0	4.1 \pm 0.3	0 \pm 0	0 \pm 0	35.8 \pm 13.4	60.4 \pm 11.8	0 \pm 0	3.8 \pm 4.2	0 \pm 0
<i>B. v. nacurutu</i>	49.4\pm1.3	9.6 \pm 0.9	14.2 \pm 1	10.2 \pm 0.8	10.5 \pm 0.7	6.1 \pm 0.9	23.6 \pm 3	24.0 \pm 3.5	32.6\pm2.2	14 \pm 2.3	2.9 \pm 0.9	2.9 \pm 1.4
<i>Glaucidium brasilianum</i>	52.6\pm13.3	6.4 \pm 1.1	13.4 \pm 6.5	6.8 \pm 1.2	13.5 \pm 13.3	7.3 \pm 5.9	56.6\pm4.9	14.6 \pm 6.4	15.2 \pm 8.9	9.2 \pm 6.4	3.9 \pm 1.9	0.6 \pm 0.6
<i>G. b. brasilianum</i>	61.9\pm1.6	6.5 \pm 0.9	18.1 \pm 0.5	6.1 \pm 1.1	3.5 \pm 1	4.0 \pm 1.7	58.1\pm2.6	10.5 \pm 1.4	22.0 \pm 1.9	4.7 \pm 0.7	3.4 \pm 0.5	1.4 \pm 0.4
<i>G. b. ucayalae</i>	34.2\pm1.0	7.3 \pm 0.5	4.4 \pm 0.8	6.8 \pm 1.1	32.0 \pm 1.2	15.3 \pm 0.9	56.5\pm8.1	22.4 \pm 5.2	3.2 \pm 2.6	14.5 \pm 8.7	3.5 \pm 3.1	0 \pm 0
<i>Glaucidium hardyi</i>	35.9\pm1.7	19.5 \pm 1.4	7.0 \pm 0.7	14.0 \pm 0.7	1.8 \pm 0.7	21.7 \pm 1.2	33.8\pm6.3	8.6 \pm 2.1	10.3 \pm 2.9	15.9 \pm 4.1	0.5 \pm 0.3	30.9 \pm 5.1
<i>Glaucidium minutissimum</i>	45.5\pm1.7	28.5 \pm 1.6	4.0 \pm 0.9	6.1 \pm 1	5.6 \pm 1.4	10.3 \pm 0.8	49.5\pm2.4	30.6 \pm 3.8	1.2 \pm 0.4	6.5 \pm 1.2	1.9 \pm 0.7	10.4 \pm 2.8
<i>Glaucidium mooreorum</i>	24.2 \pm 1.4	12.9 \pm 2.2	23.4 \pm 1	38.6\pm1.1	0.7 \pm 1.0	0 \pm 0	32.0 \pm 24.6	37.5\pm21.9	16.9 \pm 7.3	13.7 \pm 5.9	0 \pm 0	0 \pm 0
<i>Lophotrix cristata</i>	53.8\pm3.7	29.9 \pm 1.0	7.8 \pm 1.3	4.9 \pm 1.6	0.1 \pm 0.1	3.5 \pm 1.2	45.4\pm7.1	24.8 \pm 6.6	15.6 \pm 4.8	6.5 \pm 3.7	0.1 \pm 0.2	7.6 \pm 2.5
<i>Megascops atricapilla</i>	57.3\pm2.1	26.7 \pm 1.4	2.7 \pm 0.4	4.9 \pm 1	3.2 \pm 1.6	5.2 \pm 0.5	47.9\pm2.4	44.1 \pm 3.3	1.0 \pm 0.7	3.7 \pm 0.4	2.2 \pm 1.1	1.1 \pm 0.5
<i>Megascops choliba</i>	52.7\pm12.5	15.9 \pm 8.7	16.1 \pm 12	8.7 \pm 5.7	5.2 \pm 2.4	1.5 \pm 1.4	47.7\pm10.2	22.1 \pm 19.7	19.3 \pm 11.9	5.4 \pm 3.4	3.7 \pm 2.4	1.7 \pm 2.3

<i>M. ch. choliba</i>	61.3±1.6	11.5±0.7	12.9±1.1	7.1±1	7.0±0.7	0.1±0.1	59.6±4.3	16±2.2	14.4±3.2	7.1±1.1	2.6±0.4	0.3±0.2
<i>M. ch. cruciger</i>	57.8±2.5	21.1±1.2	9.2±0.7	2.6±0.6	5.5±1.8	3.7±0.6	46.4±4.0	15.4±4.2	20.8±3.2	4.6±1.7	6.8±1.9	5.9±1.8
<i>M. ch. decussatus</i>	29.4±1.7	9.1±1.3	37.5±0.9	16.9±0.7	4.7±1.0	2.3±0.4	44.9±2.0	4.1±1.5	36.5±2.2	10.4±1.4	2.8±0.6	1.3±0.6
<i>M. ch. uruguayi</i>	62.3±4.5	29.7±4.5	2.4±0.9	3.3±0.6	1.5±1.0	0.8±0.5	40.2±17.3	56.6±16.9	1.3±1.2	1.1±0.7	0.7±0.4	0.1±0.1
<i>Megascops roraimae</i>	22.4±1.6	36.8±1.2	1.6±0.6	39.3±2.3	0±0	0±0	20.3±15.4	46.6±9.2	1.7±2.7	31.3±11.4	0±0	0±0
<i>Megascops sanctaecatarinae</i>	63.8±1.0	27.7±1.4	0.3±0.2	5.6±1	0±0	2.6±0.7	28.9±7.7	69.1±7.7	0.5±0.4	0.6±0.5	0±0	0.8±0.4
<i>Megascops watsonii</i>	42.6±5.2	22.9±6.2	10.5±7.5	5.5±4.1	8.7±6.2	9.8±6.8	40.6±7.8	29.9±12.8	12.2±7.0	5.0±3.9	6.6±5.1	5.8±5.1
<i>M. w. usta</i>	38.0±1.3	27.7±1.3	4.0±0.6	1.5±0.4	12.0±1.8	16.9±0.9	44.2±6.4	35.2±6.3	3.9±1.8	2.7±0.9	8.8±2.1	5.2±1.5
<i>M. w. watsonii</i>	41.1±2.1	26.6±0.6	20.8±1.0	10.4±3.0	0.3±0.5	0.9±0.3	32.5±4.6	40.7±5.5	18.5±4.8	8.1±5.3	0±0	0.2±0.2
<i>Pulsatrix koenigswaldiana</i>	64.9±1.0	22.5±1.1	2.1±0.5	6.7±0.8	0.6±0.3	3.1±0.5	55.1±3.7	32±2.9	1.7±0.4	9.3±1.2	0.9±0.5	0.9±0.3
<i>Pulsatrix perspicillata</i>	26.2±20.9	19.2±7.1	29.8±13.8	11.9±8.8	7.6±8.3	5.3±3.3	31.6±14.2	21.7±8.1	31.6±9.7	6.9±5.8	4.9±5.7	3.2±2
<i>P. p. perspicillata</i>	55.2±0.8	29.0±0.8	11.6±0.5	0.1±0.1	2.1±1.0	2.0±0.4	50.1±5.5	22.7±3.5	22.1±4.9	0.4±0.5	1.3±1.0	3.3±1.3
<i>P. p. pulsatrix</i>	10.4±2.2	14.1±1.6	43.9±1.3	20.4±1.1	1.6±0.7	9.6±1.2	21.6±5.7	28.7±8	37.1±9.9	9.2±5.9	0.7±0.9	2.7±2.8
<i>Strix huhula</i>	35.7±21.0	26.5±7.3	21.1±13.1	2.8±2.0	7.1±7.1	6.9±4.8	40.6±13.8	24.8±9.4	21.4±15.5	2.5±2.0	4.4±4.5	6.4±5.7
<i>S. h. albomarginata</i>	64.7±1.7	24.3±1.6	3.1±0.3	4.9±0.9	0.8±0.3	2.2±0.9	58.0±7.9	35.1±7.4	1.3±0.6	4.5±1.6	0.7±0.5	0.4±0.3
<i>S. h. huhula</i>	22.6±0.8	35.9±1.5	31.8±1.4	0.4±0.4	3.8±1.7	5.5±1.1	33.8±3.4	21.2±5.6	36.3±4.1	0.7±0.9	2.6±1.6	5.4±2
<i>Strix hylophila</i>	79.3±0.8	7.4±0.9	3.4±0.6	6.6±0.6	0.9±0.4	2.4±0.4	51.1±3.7	39.3±4.1	0.9±0.5	6.3±2.3	1.1±0.6	1.1±0.4
<i>Strix virgata</i>	46.5±10	28.1±3.1	9.1±4.9	8.9±5.6	2.8±2.2	4.6±3.3	40.6±12.6	40.9±8.9	8±6.2	6.3±2.7	1.3±1.3	2.9±3.1
<i>S. v. borelliana</i>	49.4±1.8	28.2±1.6	5.1±0.5	8.0±1.0	3.8±0.8	5.5±0.6	43.1±2.2	45.0±3.4	4.2±1.2	5.4±1.2	1.0±0.4	1.3±0.4
<i>S. v. supercilialis</i>	56.6±2.3	24.6±0.8	15.6±1.7	2.8±0.6	0±0	0.5±1.5	49.8±15.8	31.8±10	12.9±8.8	4.6±3.0	0±0	0.9±2.9

According to the *sensu stricto* map (Fig. 3.2a), the Atlantic Forest hosts the highest potential richness (up to 16 species per pixel), especially around the littoral and mountainous areas of Southeastern Brazil. Scattered areas in the Amazonia, associated to the Guiana Shield highlands, and areas of the Cerrado bordering the Atlantic Forests, are also potentially diverse (up to 15 species per pixel each). Conversely, wide coldspots (around zero predicted species) characterize vast portions of the Cerrado and the Amazonia, especially in the limit of both biomes or, for example, in the Solimões basin. On the other hand, the *sensu lato* map (Fig. 3.2b) keeps the same areas of high biodiversity but reducing the extension of the coldspots in the Amazonia to some areas in the Negro river basin and the Rondônia state. Thus, both approaches indicate that the Atlantic Forest harbors the highest richness, but is poorly covered by strictly protected areas since these become substantially smaller and sparser within a gradient from Northwestern to Southeastern Brazil. However, by comparing the number of species (*sensu stricto*) recorded against those predicted, we found that almost all biomes are misestimated (Table 3.2), especially the Pampa.

Figure 3.2 – Predicted Strigidae species richness patterns in Brazil based on (A) species *sensu stricto*; and (B) *sensu lato* (including subspecies). Dashed polygons represent the strict protected areas. Biome limits in dark gray lines.



Subspecies' niche analysis

The results of the similarity tests between conspecific subspecies are summarized in Supplementary material A, Appendix 1, Fig. F. We found niche overlap values significantly lower than expected from a null model distribution for all the pairwise comparisons. Consequently, we reject the null hypothesis of niche identity for all pairs.

Our background tests yielded a more complex scenario (Supplementary material A, Appendix, 1 Fig. G). The observed values of *D* and *I* usually lied below the corresponding null distributions; hence, differences are bigger than expected by chance. However, we were unable to reject the null hypothesis of the background similarity test when comparing *A. c. cunicularia* → *A. c. grallaria*, *B. v. nacurutu* → *B. v. deserti*, *M. c. choliba* → *M. c. decussatus* and *M. c. choliba* → *M. c. uruguaii*, with arrows indicating the directionality of the comparison: the first subspecies against a randomized background derived from the second one. However, the *D* and *I* metrics of the respective complementary comparisons fell below their corresponding null distributions, indicating that these were more different than expected by chance, hence leading us to reject the null hypothesis that their niches are similar.

DISCUSSION

In general, our results show that the situation of Strigidae in Brazil is complicated, involving an underestimated richness, within an inadequate framework of protected areas, in a megadiverse country characterized by high rates of habitat transformations (cover changes and fragmentation).

Sampling biases, distribution patterns and taxonomic knowledge

Brazil harbors a rich owl diversity, distributed unequally throughout its territory, ranging between zero and 16, averaging three, species per 30 arcsec * 30 arcsec pixel. Since 46% of the occurrences are from the Atlantic Forest, its highest predicted richness can partially respond to sampling biases. In the Atlantic Forest sampling efforts are favored by neighboring research centers available and major urban areas (Ribeiro et al., 2016). The Amazonia is the second most sampled biome, but replicating the clustered pattern already described for botanical and

ornithological records (Vale and Jenkins, 2012) probably prompted by local facilities or recurrent bird-oriented inventories, leaving poorly-sampled areas coinciding with the coldspots. However, all biomes are under-sampled, as suggested by the difference between the number of recorded and predicted species, especially in the Pampa. Hence, our results agree with Fernández-Arellano et al. (2021) regarding the need for research efforts addressed to unexplored areas and periods, less detectable species, and neglected biomes.

Excluding the possibility of factors not considered in the present study shaping the distributions of the Brazilian Strigidae, several not mutually exclusive possibilities can explain the overall elevated number of omissions for the models. First, misidentifications. For example, Rocha and López-Baucells (2014a) reported a young *Lophostrix cristata* that was indeed a *Strix virgata* Rocha and López-Baucells (2014b). Thus, similar unnoticed mistakes, especially those involving species more alike, can result in false occurrences reported far from the actual range.

Second, vagrancy which, according to the ‘Exodus Hypothesis’ (Flade and Lachmann, 2008), can be triggered by severe habitat loss. Here, most omissions corresponded to taxa from the Atlantic Forest recorded in open areas such as the Cerrado. Currently, the Atlantic Forest covers less than 16% of the original extent, and more than 80% of the fragments are smaller than 50 ha (Ribeiro et al., 2009), a dramatic situation that could have triggered the translocation of owls to neighboring areas the comparatively less impacted Cerrado areas in past decades.

Third, range expansions, as reported for Strigiformes in well-monitored areas of North America, Europe and Oceania, often mediated by human activities. For example, according to Livezey (2009), the forest expansions after the arrival of European settlers (who excluded fires regularly set by Native Americans, planted trees, and overhunted large herbivores) allowed barred owls (*Strix varia*) to overcome the historical lack of trees in the Great Plains of North America granting their expansion westward in the recent years. Similarly, in the case of *A. harrissi*, most of the omissions correspond to records from the last two decades in open biomes such as the Cerrado and the Caatinga (Silva et al., 2021), extensively anthropized during the recent decades according to MAPBIOMAS (<https://plataforma.brasil.mapbiomas.org/>, consulted April 14, 2021), with steady increases in planted tree surfaces (<http://atlasagropecuario.imaflora.org/>, consulted April 16, 2021), perhaps offering newly suitable areas for these tree-nester owls.

Notwithstanding, a more likely option is that the taxonomy of the Brazilian owls is far from being completely understood, with local populations/morphs representing in many cases separate taxa,

as recently demonstrated by Dantas et al. (2021). Thus, it is reasonable to expect SDMs biased toward the best-represented groups, failing to predict regions suitable for the others. Our models for the conspecific subspecies and our results on niche equivalency, even if assigning the occurrences in an approximatively way, reinforced this possibility.

Conservation-related issues

The combination of sampling biases, ecologic and taxonomic knowledge gaps, can lead to misguided conservation decisions. For example, similarly to what had already reported for bats (Delgado-Jaramillo et al., 2020), some of the coldspots in our *sensu stricto* map (Fig. 3.2a) are placed along areas severely deforested during the last decades, particularly along the contact zone between the Amazonia and the Cerrado in the so-called “arch of deforestation” (Becker, 2005). Such coldspots could be misinterpreted as of lower conservation interest, while originated by the absence of information after local extinctions prompted by anthropic impacts (Delgado-Jaramillo et al., 2020).

The greater diversity of Strigidae within the Atlantic Forest corresponds to an extensively degraded biome with small-sized and sparse protected areas, whose efficacy is lower than other regions, such as the Amazonia (Sobral-Souza et al., 2018). Worse yet, the buffer zones around these protected areas in the economically prosperous Brazilian regions tend to be as degraded as the surrounding unprotected ones (Almeida-Rocha and Peres, 2021). In fact, bird species are already extinct or extirpated from large tracts of the Atlantic Forest, especially in the Pernambuco Centre of Endemism in northeast Brazil (Develey and Phalan, 2021). Thus, we agree with Oliveira et al. (2017) who highlighted the deficiencies of the Brazilian network of protected areas given the combination of the scarce knowledge on their biodiversity and an inadequate spatial disposition which offers limited or no protection to most species and even evolutionary lineages. We also agree with Jenkins et al. (2015), regarding the urgent need to substantially improve the network of protected areas in Brazil, with an emphasis in the Atlantic Forest.

Acknowledgements

We are grateful to Ivo Rohling Ghizoni Jr, Gerald Mayr, Martin Flade and Thadeu Sobral for providing valuable information. P.R.P. also thanks to the Organization of American States, through its Partnerships Program for Education and Training of the Coimbra Group of Brazilian Universities (OAS/PAEC/GCUB).

REFERENCES

- Almeida-Rocha, J.M. de, Peres, C.A., 2021. Nominally protected buffer zones around tropical protected areas are as highly degraded as the wider unprotected countryside. *Biological Conservation* 256, 109068. <https://doi.org/10.1016/j.biocon.2021.109068>
- Alvares, C.A., Stape, J.L., Sentelhas, P.C., Gonçalves, J.L. de M., Sparovek, G., 2013. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>
- Barros, O.G., Cintra, R., 2009. The effects of forest structure on occurrence and abundance of three owl species (Aves: Strigidae) in the Central Amazon forest. *Zoologia* 26, 85–96.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Becker, B.K., 2005. Geopolítica da Amazônia. *Estudos Avançados* 19, 71–86.
- BirdLife International, 2019. IUCN Red List for birds [WWW Document]. URL <http://www.birdlife.org> (accessed 9.18.19).
- Booth, T.H., Nix, H.A., Busby, J.R., Hutchinson, M.F., 2014. Bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions* 20, 1–9. <https://doi.org/10.1111/ddi.12144>
- Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010) (Report No. 2011–1073), Open-File Report. <https://doi.org/10.3133/ofr20111073>
- Dantas, S.M., Weckstein, J.D., Bates, J., Oliveira, J.N., Catanach, T.A., Aleixo, A., 2021. Multi-character taxonomic review, systematics, and biogeography of the Black-capped/Tawny-bellied Screech Owl (*Megascops atricapilla* - *M. watsonii*) complex (Aves: Strigidae). *Zootaxa* 4949, 401–444. <https://doi.org/10.11646/zootaxa.4949.3.1>

- Delgado-Jaramillo, M., Aguiar, L.M.S., Machado, R.B., Bernard, E., 2020. Assessing the distribution of a species-rich group in a continental-sized megadiverse country: Bats in Brazil. *Diversity and Distributions* 26, 632–643. <https://doi.org/10.1111/ddi.13043>
- Develey, P.F., Phalan, B.T., 2021. Bird extinctions in Brazil's Atlantic Forest and how they can be prevented. *Frontiers in Ecology and Evolution* 9, 275. <https://doi.org/10.3389/fevo.2021.624587>
- Donázar, J.A., Cortés-Avizanda, A., Fargallo, J.A., Margalida, A., Moleón, M., Morales-Reyes, Z., Moreno-Opo, R., Pérez-García, J.M., Sánchez-Zapata, J.A., Zuberogoitia, I., Serrano, D., 2016. Roles of raptors in a changing world: From flagships to providers of key ecosystem services. *Ardeola* 63, 181–234. <https://doi.org/10.13157/arla.63.1.2016.rp8>
- Enríquez, P.L., 2017. Neotropical owls: Diversity and conservation, Springer International Publishing. Springer, Cham.
- Fernández-Arellano, G.J., Teixido, A.L., Bernardon, B., Bueno, E.R., Ferreira, T.V., Gonçalves, S.R.A., Jesus, M., Thomas, K.P.C.S., Zucchetto, M., Piacentini, V.Q., Pinho, J.B., 2021. Knowledge gaps and biases in the Pantanal indicate future directions for ornithological research in large wetlands. *Ibis* 163, 784–797. <https://doi.org/10.1111/ibi.12931>
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24, 38–49. <https://doi.org/10.1017/S0376892997000088>
- Flade, M., Lachmann, L., 2008. International species action plan for the Aquatic Warbler *Acrocephalus paludicola*. BirdLife International, Cambridge, UK.
- Fluck, I.E., Cáceres, N., Hedges, C.D., Brum, M. do N., Dambros, C.S., 2020. Climate and geographic distance are more influential than rivers on the beta diversity of passerine birds in Amazonia. *Ecography* 43, 860–868. <https://doi.org/10.1111/ecog.04753>
- Gallo-Cajiao, E., Archibald, C., Friedman, R., Steven, R., Fuller, R.A., Game, E.T., Morrison, T.H., Ritchie, E.G., 2018. Crowdfunding biodiversity conservation. *Conservation Biology* 32, 1426–1435. <https://doi.org/10.1111/cobi.13144>
- Gill, F., Donsker, D., Rasmussen, P., 2021. IOC World bird list 11.1. <https://doi.org/10.14344/IOC.ML.11.1>
- Gómez Tapias, J., Schobbenhaus, C., Montes Ramírez, N., 2019. Geological map of South America 2019.
- Heibl, C., Calenge, C., 2018. phyloclim: Integrating phylogenetics and climatic niche modeling.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangquan, 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., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* 12, 1–40. <https://doi.org/10.1371/journal.pone.0169748>

- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 46, 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- IBGE, 2019. *Biomass e sistema costeiro-marinho do Brasil compatível com a escala 1:250 000, Relatórios Metodológicos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro.
- Jenkins, C.N., Alves, M.A.S., Uezu, A., Vale, M.M., 2015. Patterns of vertebrate diversity and protection in Brazil. *PLOS ONE* 10, 1–13. <https://doi.org/10.1371/journal.pone.0145064>
- Liu, C., White, M., Newell, G., 2016. Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography* 40, 778–789. <https://doi.org/10.1111/jbi.12058>
- Livezey, K.B., 2009. Range expansion of Barred Owls, Part II: Facilitating ecological changes. *The American Midland Naturalist* 161, 323–349. <https://doi.org/10.1674/0003-0031-161.2.323>
- Mainali, K., Hefley, T., Ries, L., Fagan, W., 2020. Matching expert range maps with species distribution model predictions. *Conservation Biology* 34, 1292–1304. <https://doi.org/10.1111/cobi.13492>
- Mendes, P., Velazco, S.J.E., Andrade, A.F.A. de, Marco, P.D., 2020. Dealing with overprediction in species distribution models: How adding distance constraints can improve model accuracy. *Ecological Modelling* 431, 109180. <https://doi.org/10.1016/j.ecolmodel.2020.109180>
- Motta-Junior, J.C., Braga, A.C.R., 2012. Estado del conocimiento sobre la ecología y biología de búhos en Brasil. *Ornitología Neotropical* 23, 227–234.
- Motta-Junior, J.C., Braga, A.C.R., Granzinolli, M.A.M., 2017. The owls of Brazil, in: Enriquez, P.L. (Ed.), *Neotropical owls: Diversity and conservation*. Springer International Publishing, Cham, pp. 97–158. https://doi.org/10.1007/978-3-319-57108-9_6
- Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J., Uriarte, M., Anderson, R.P., 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for ecological niche models. *Methods in Ecology and Evolution* 5, 1198–1205.
- Naimi, B., Hamm, N.A., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., Alibakhshi, S., 2019. ELSA: An entropy-based local indicator of spatial association. *Spatial Statistics* 29, 66–88. <https://doi.org/10.1016/j.spasta.2018.10.001>
- 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., 2017. Biodiversity conservation gaps in the Brazilian protected areas. *Scientific reports* 7, 9141. <https://doi.org/10.1038/s41598-017-08707-2>
- Pereira, G.A., 2010. Avifauna associada a três lagoas temporárias no estado do Rio Grande do Norte, Brasil. *Atualidades Ornitológicas* 156, 53–60.

- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E., 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40, 887–893. <https://doi.org/10.1111/ecog.03049>
- Protected Planet, 2021. The World Database on Protected Areas (WDPA).
- R Core Team, 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramoni-Perazzi, P., Passamani, M., Thielen, D., Padovani, C., Arizapana, M.A., 2021. BrazilClim: dataset of bioclimatic variables for the continental megadiverse Brazil. *International Journal of Climatology*. <https://doi.org/10.1002/joc.7325>
- Ramoni-Perazzi, P., Schuchmann, K.-L., Weller, A., Soto-Werschitz, I.A., Passamani, M., 2020. Niches and radiations: a case study on the Andean sapphire-vented puffleg *Eriocnemis luciani* and coppery-naped puffleg *E. sapphiropygia* (Aves, Trochilidae). *Journal of Avian Biology* 51. <https://doi.org/10.1111/jav.02242>
- Ribeiro, G.V.T., Teixeira, A.L., Barbosa, N.P.U., Silveira, F.A.O., 2016. Assessing bias and knowledge gaps on seed ecology research: implications for conservation agenda and policy. *Ecological Applications* 26, 2033–2043. <https://doi.org/10.1890/15-1852.1>
- Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation* 142, 1141–1153. <https://doi.org/10.1016/j.biocon.2009.02.021>
- Rocha, R., López-Baucells, A., 2014a. Opportunistic predation of the Crested Owl *Lophotrix cristata* upon on Seba's short-tailed bat *Carollia perspicillata*. *Revista Brasileira de Ornitologia* 22, 58. <https://doi.org/10.1007/BF03544230>
- Rocha, R., López-Baucells, A., 2014b. Erratum to: Opportunistic predation by Crested Owl *Lophotrix cristata* upon Seba's Short-tailed Bat *Carollia perspicillata*. *Revista Brasileira de Ornitologia* 22, 244. <https://doi.org/10.1007/bf03544253>
- Sánchez de Dios, R., Cabal Ruano, C., Domínguez Lozano, F., Sainz Ollero, H., Moreno Saiz, J.C., 2017. The role of criteria in selecting important areas for conservation in biodiversity-rich territories. *Diversity and Distributions* 23, 368–380. <https://doi.org/10.1111/ddi.12535>
- Santorelli, S., Magnusson, W.E., Deus, C.P., 2018. Most species are not limited by an Amazonian river postulated to be a border between endemism areas. *Scientific Reports* 8, 2294. <https://doi.org/10.1038/s41598-018-20596-7>
- Silva, A.F.T. da, Melo, H.R.S. de, Ubaid, F., 2021. First records of Buff-fronted Owl, *Aegolius harrisi* (Cassin, 1849) (Aves, Strigiformes), from the state of Maranhão, northeastern Brazil, and the northernmost record for the Cerrado domain. *Check List* 17, 353–358. <https://doi.org/10.15560/17.2.353>

Silva, J.M.C. da, Coelho, G., Gonzaga, L., 2002. Discovered on the brink of extinction: A new species of pygmy-owl (Strigidae: Glaucidium) from Atlantic Forest of northeastern Brazil. *Ararajuba* 10, 123–130.

Silva, H.G. de, Medellín, R.A., 2001. Evaluating completeness of species lists for conservation and macroecology: A case study of Mexican land birds. *Conservation Biology* 15, 1384–1395. <https://doi.org/10.1111/j.1523-1739.2001.00177.x>

Sobral-Souza, T., Vancine, M.H., Ribeiro, M.C., Lima-Ribeiro, M.S., 2018. Efficiency of protected areas in Amazon and Atlantic Forest conservation: A spatio-temporal view. *Acta Oecologica* 87, 1–7. <https://doi.org/10.1016/j.actao.2018.01.001>

Vale, M.M., Jenkins, C.N., 2012. Across-taxa incongruence in patterns of collecting bias. *Journal of Biogeography* 39, 1744–1748. <https://doi.org/10.1111/j.1365-2699.2012.02750.x>

Veloz, S.D., 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography* 36, 2290–2299. <https://doi.org/10.1111/j.1365-2699.2009.02174.x>

Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>

Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21, 335–342. <https://doi.org/10.1890/10-1171.1>

DATA ACCESSIBILITY STATEMENT

Supplementary Material A (<https://doi.org/10.4121/15121137>): additional figures, citations, and tables.

Supplementary Material B (<https://doi.org/10.4121/15121188>): occurrences database.