



Universidade Estadual de Goiás

Câmpus Anápolis de Ciências Exatas e Tecnológicas - Henrique Santillo

Programa de Pós-Graduação *Stricto Sensu* em Recursos Naturais do Cerrado

THATYANE CAETANO DE ALMEIDA

**VIESES ESPACIAIS EM DADOS DE OCORRÊNCIA DE PEIXES DE ÁGUA DOCE
DO BRASIL**

Anápolis

2019

THATYANE CAETANO DE ALMEIDA

**VIESES ESPACIAIS EM DADOS DE OCORRÊNCIA DE PEIXES DE ÁGUA DOCE
DO BRASIL**

Dissertação apresentada ao Programa de Pósgraduação *Stricto Sensu* em Recursos Naturais do Cerrado, da Universidade Estadual de Goiás para obtenção do título de Mestre em Recursos Naturais do Cerrado. Orientador: Prof. Dr Fabrício Barreto Teresa. Co-orientadora: Prof^a Dr^a Geiziane Tessarolo

Anápolis

2019



**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DE TESES E DISSERTAÇÕES
NA BIBLIOTECA DIGITAL (BDTD)**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Estadual de Goiás a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UEG), regulamentada pela Resolução, CsA n.1087/2019 sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

Dados do autor (a)

Nome Completo: Thatyane Caetano de Almeida

E-mail: athaty_caetano@hotmail.com

Dados do trabalho

Título: Vieses espaciais em dados de ocorrência de peixes de água doce no Brasil.

Tipo

() Tese (X) Dissertação

Curso/Programa Pós graduação em Recursos Naturais do Cerrado

Concorda com a liberação documento [x] SIM

[] NÃO¹

Anápolis, 03/05/2019

Thatyane Caetano de Almeida
Thatyane Caetano de Almeida

Fabício Barreto Teresa
Fabrício Barreto Teresa

Almeida, Thatyane Caetano de

Vieses espaciais em dados de ocorrência de peixes de água doce do Brasil. / Thatyane Caetano de Almeida. – 2019.

31 f.

Orientação: Prof. Dr. Fabrício Barreto Teresa.
Coorientação: Prof. Dra. Geiziane Tessarolo.

Dissertação (Mestrado) – Universidade Estadual de Goiás,
Câmpus de Ciências Exatas e Tecnológicas, 2019.

1. Déficit Wallaceanno. 2. Brasil. 3. Déficit Lineanno. 4.
Vieses espaciais.

THATYANE CAETANO DE ALMEIDA

VIESES ESPACIAIS EM DADOS DE
OCORRÊNCIA DE PEIXES DE ÁGUA DOCE DO
BRASIL

Dissertação defendida no Programa de Pós-Graduação *Stricto Sensu* em Recursos
Naturais do Cerrado da Universidade Estadual de Goiás,
para a obtenção do grau de Mestre, aprovada em 25 de fevereiro de 2019, pela
Banca Examinadora constituída pelos seguintes professores:



Prof. Dr. Fabrício Barreto Teresa
Presidente da banca
Universidade Estadual de Goiás



Prof. Dr. Murilo Sversut Dias
Membro externo
Universidade de Brasília



Prof. Dr. Vitor Hugo Mendonça do Prado
Membro interno
Universidade Estadual de Goiás

AGRADECIMENTOS

A Deus e aos meus familiares que me deram força e apoio para cumprir essa etapa da minha vida.

A esta universidade e a este programa que fizeram parte da minha formação. Ao INCT-EECBio e CNPq pela bolsa concedida. Ao Laboratório de Biogeografia e Ecologia Aquática e todos os seus membros (alunos e professores, em especial a Karine, Pedro e Micael, que estavam comigo todos os dias e compartilhávamos os mesmos desesperos) pelo companheirismo e amizade, vocês foram muito importantes na minha formação como pessoa e pesquisadora.

Ao meu orientador Fabrício Barreto Teresa, que se propôs a me ajudar ao longo dessa jornada, pelo seu suporte, incentivo e pela excelente pessoa que é, e que me motivou continuar nesse caminho e acreditou em mim e na minha capacidade.

A minha co-orientadora Geiziane Tessarolo, por toda paciência que teve, pelo suporte nos momentos de desespero e por toda ajuda que me deu.

E a todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigada.

SUMÁRIO

Resumo.....	04
Abstract.....	05
Introdução.....	06
Métodos.....	07
Coleta de dados.....	07
Fatores Explanatórios dos Vieses Espaciais na Amostragem	08
Análise dos dados.....	10
Resultados.....	11
Discussão.....	15
Referências.....	18
Material Suplementar.....	22

RESUMO

O viés espacial, ou seja, a amostragem predominante em algumas regiões em detrimento de outras, afeta a qualidade e representatividade dos dados de biodiversidade. Esses vieses podem ser gerados por diversos fatores, como busca por localidades com maior proximidade de centros urbanos, presença de vias de acesso, falta de recursos financeiros para financiar amostragens mais especializadas, preferência por amostragens próximas ou em áreas de conservação, atratividade por determinada espécie e/ou local, entre outros. O conhecimento sobre a diversidade de peixes de água doce vem aumentando rapidamente, porém ainda há incompletude das informações de suas ocorrências, sendo que estudos sobre vieses amostrais em ambientes aquáticos são escassos. Sendo assim, avaliamos se as amostragens de peixes de água doce são enviesadas em decorrência de fatores como acessibilidade, infraestrutura de pesquisa e atratividade de áreas protegidas. Para isso, compilamos os registros de ocorrências de 3.358 espécies no Brasil (27.3% da biodiversidade mundial) e modelamos o esforço de amostragem (eventos de coletas) em relação à proximidade com centros de pesquisas, proximidade e densidade de vias de acesso, proximidade à áreas protegidas e a densidade populacional, considerando a extensão do Brasil e de cada região do país separadamente. Constatamos que o esforço de amostragem é enviesado espacialmente, com maior concentração nas regiões Sul e Sudeste do Brasil. Além disso, para o Brasil, os fatores mais importantes para explicar a variação no esforço de amostragem foram, em ordem de importância, a proximidade com centros de pesquisa, densidade populacional, densidade e distância de vias de acesso e proximidade a áreas protegidas. Apesar de algumas variações na importância dos fatores explicando a amostragem entre as regiões, a proximidade dos centros de pesquisa foi importante para todas as regiões. Isso reforça a relevância da interiorização e regionalização da infraestrutura de pesquisa para ampliar o conhecimento sobre a biodiversidade.

Palavras chave: Déficit Wallaceano, déficit Linneano, centro de pesquisa.

ABSTRACT

Spatial sampling bias, i.e., the predominant sampling effort in some regions in detriment of others, affects the quality and representativeness of biodiversity data. These biases can be generated by many factors, such as sampling in areas near to urban centers, presence of roads, lack of financial resources for more spatialized samplings, preference for samplings nearby or within conservation areas, preference for a particular species and/or place, among others. Although the knowledge about freshwater fish diversity has increased rapidly, there is still an knowledge gap of their occurrences, and studies about sampling biases in aquatic environments are scarce. Thus, we evaluated if the sampling of freshwater fishes is biased due to factors such as accessibility, research infrastructure and preference for protected areas. Here, we compiled the occurrence records of 3358 species of Brazil (27.3% of the worldwide biodiversity) and modeled the sampling effort (collection events) in relation to the proximity to research centers, proximity to and density of roads, proximity to protected areas and populational density, considering both the whole Brazilian extension and each country region individually. We found that the sampling effort is spatially biased, more concentrated in the South and Southeast regions of Brazil. In addition, for Brazil, the most important factors to explain the variation in sampling effort were, in a decrescent order of importance, the proximity to research centers, populational density, density and proximity to access routes and proximity to protected areas. Despite variations in the importance of factors explaining sampling effort among regions, the proximity to research centers was significant for all regions. This result reinforces the relevance of the internalization and regionalization of research infrastructure to increase knowledge about biodiversity.

Keywords: Wallacean shortfall, Linnean shortfall, research center.

INTRODUÇÃO

A obtenção de dados de boa qualidade sobre a ocorrência das espécies é essencial para o conhecimento da biodiversidade (Troia & McManamay, 2017). Isso contribui diretamente para reduzir as limitações de conhecimento sobre a identidade taxonômica das espécies (i.e. lacuna Linneana) e a respeito das suas distribuições geográficas (i.e. lacuna Wallaceana) (Lomolino, 2004; Whittaker et al., 2005; Bini et al., 2006; Hortal et al., 2015). Para isso, o incremento e distribuição dos esforços de amostragem no espaço geográfico é fundamental (Brooks et al., 2004). Entretanto, embora haja uma crescente preocupação com a intensificação e espacialização das coletas (Reddy & Dávalos, 2003; Hortal, et al. 2007; Romo, et al. 2006; Oliveira, et al. 2016; Troia & McManamay, 2017; Girardello, et al. 2018), os vieses nos registros da biodiversidade ainda são grandes (Brooks et al., 2004, Girardello, et al. 2018).

Os vieses de amostragem são lacunas no conhecimento que são decorrentes, em parte, da pequena escala espacial dos estudos, gerando concentração espacial de registros, limitando o conhecimento sobre os padrões de distribuição em escalas maiores (Hecnar & M'Closkey, 1997; Dias et al., 2016). Como consequência dos vieses nos registros das espécies, o conhecimento sobre a biodiversidade é distorcido, prejudicando estudos e inferências de padrões sobre a biodiversidade, que consequentemente afeta ações de conservação (Brooks et al., 2004). A identificação dos fatores que influenciam esses vieses é um passo necessário para minimizar seus efeitos e, assim, obter uma distribuição mais equitativa dos esforços de amostragem (Reddy & Dávalos, 2003; Sánchez-Fernández et al., 2008; Girardello, et al. 2018).

Vários fatores podem estar associados com a distribuição desigual do esforço de amostragem das espécies como, por exemplo, a facilidade de acesso (Romo et al., 2006; Sánchez-Fernández et al., 2008; Oliveira et al, 2016), disponibilidade de recursos financeiros (Anderson, 2012; Yang et al., 2013), atratividade por áreas específicas ou espécies (Reddy & Dávalos, 2003). De fato, pesquisadores tendem a concentrar as suas coletas em locais próximos ao seu local de trabalho, com a presença de rodovias e próximos a grandes cidades, como uma maneira de facilitar a locomoção (Sánchez-Fernández et al., 2008; Sastre & Lobo, 2009) e minimizar a logística e despesas financeiras com as atividades de campo (Botts et al., 2011; Meyer et al., 2015). Além disso, alguns locais também exercem certa atratividade sobre os pesquisadores, como por exemplo, as áreas protegidas, por serem locais mais conservados

onde há maior probabilidade de se encontrar alta riqueza de espécies ou espécies raras (Sánchez-Fernández et al., 2008; Oliveira et al., 2016).

Nos últimos anos, houve uma maior preocupação em relação a identificação dos vieses espaciais nos dados de biodiversidade e suas causas (Reddy & Dávalos, 2003; Hortal et al., 2007; Romo et al., 2006; Sánchez-Fernández et al., 2008; Boakes et al., 2010; Yang et al., 2013; Oliveira et al., 2016; Stropp et al., 2016). Esses estudos revelam que os padrões de amostragem são dirigidos por fatores que facilitam as coletas, resultando em amostragens enviesadas. Entretanto, a grande maioria dos estudos são voltados para espécies de vertebrados terrestres, enquanto que os vieses associados com a biodiversidade aquática são ainda poucos estudados (Sánchez-Fernández et al., 2008, Levêque et al., 2005).

Neste estudo, avaliamos a distribuição espacial no esforço de amostragem de peixes de água doce, buscando compreender fatores causadores dos vieses amostrais. Testamos a hipótese de que as amostragens são enviesadas devido à acessibilidade às áreas para coleta, infraestrutura de pesquisa e atratividade de áreas protegidas. Para isso, compilamos os registros de ocorrência de peixes de água doce no Brasil, país que utilizamos como modelo por apresentar uma megadiversidade, cujos esforços de pesquisas ocorreram mais recentemente e onde as lacunas de conhecimento ainda são grandes (Azevedo et al., 2010; Dias et al., 2016).

MÉTODOS

Coleta de dados

Foram listadas 3.358 espécies de peixes de água doce para o Brasil, com base no banco de dados Fishbase (www.fishbase.org), a partir da utilização dos filtros de busca “Brazil” e “Freshwater”. Espécies marinhas que eventualmente foram encontradas nos dados foram removidas, compondo então a lista apenas espécies de água doce. Para cada espécie, obtivemos os números de registros de ocorrência por meio dos seguintes bancos de dados online: *Global Biodiversity Information Facility* (GBIF; www.gbif.org), *SpeciesLink* (www.splink.org.br), Portal da Biodiversidade (portaldabiodiversidade.icmbio.gov.br/portal/) e Sistema de Informação sobre a Biodiversidade Brasileira (SBBR; www.sibbr.gov.br/) (Figura 1A). Em seguida os dados foram analisados levando em conta: 1) presença de coordenadas geográficas (latitude e longitude); 2) ano da coleta e nome das espécies; e 3) identificação de dados duplicados, sendo considerados como duplicados aqueles dados que se

repetiam nas diferentes bases de dados, apresentando mesma coordenada geográfica, data e nome da espécie. Os dados que não apresentavam essas informações ou eram duplicados foram removidos, juntamente com pontos de ocorrência fora do Brasil. Com o auxílio dos programas QGIS 2.18 e R, pacotes *vegan*, *rgbif* (QGIS Development Team, 2015, R Core Team 2017, Oksanen, et al., 2018) os dados foram agregados a uma grade com resolução de 0,5°, assim cada célula da grade compunha um conjunto de pontos de ocorrência (Figura 1B).

O esforço de amostragem em cada célula (variável resposta) foi quantificado como o somatório dos eventos de amostragem dentro de cada célula. Para isso, os registros das espécies foram avaliados quanto à coordenada do local de amostragem e ao ano. Consideramos como um único evento aqueles registros com as mesmas coordenadas dentro de um mesmo ano. Assim, registros de espécies com as mesmas coordenadas, mas realizados em anos diferentes foram considerados como eventos de amostragem diferentes. Além disso, registros de espécies diferentes com as mesmas coordenadas em um dado ano foram considerados como um único evento de amostragem. Utilizamos o ano ao invés da data exata do registro, pois muitos registros não dispõem da informação sobre a data da coleta. Dessa forma, ao quantificar os eventos de amostragem e não os registros de ocorrência das espécies *per se*, nossa métrica de esforço não é influenciada diretamente pelas diferenças nos padrões geográficos de riqueza de espécies (Vieira et al., 2018).

Fatores Explanatórios dos Vieses Espaciais na Amostragem

Os fatores explanatórios dos vieses espaciais na amostragem dos peixes foram avaliados a partir de variáveis que representam acessibilidade (densidade demográfica, distância à rodovias e densidade de vias de acesso), disponibilidade de infraestrutura de pesquisa (distância de centros de pesquisa), assim como a atratividade de áreas protegidas (distância para áreas protegidas) (Figura 1). Essas variáveis foram obtidas para todas as células da grade com resolução de 0,5°.

Densidade demográfica

Os dados de densidade demográfica dos municípios foram obtidos na plataforma online do Instituto Brasileiro de Geografia e Estatística, censo demográfico de 2010 (IBGE; www.ibge.gov.br). Os valores de densidade demográfica dos municípios foram reamostrados para a resolução 0,5°. Para as células que continham mais de um município foi usado o valor médio para aquele local.

Centros de pesquisa

Consideramos como centro de pesquisa o local de lotação de pesquisador atuante em linha de pesquisa envolvendo peixes de água doce. Adotamos essa estratégia para representar centros de pesquisa que efetivamente empreendem esforços para a amostragem de peixes. Para isso, inicialmente foi realizado o levantamento de todos os programas de pós-graduação brasileiros (PPGs) de quatro grandes áreas, cujas pesquisas têm maior potencial para envolver estudos com peixes (Ciências Ambientais, Biologia Geral, Biodiversidade e Recursos Pesqueiros), totalizando 262 PPGs. Essas áreas são definidas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), órgão governamental que gerencia a pós-graduação no Brasil. Dentro de cada grande área, foram selecionados os PPGs que possuem linhas de pesquisa ligadas ao estudo da biodiversidade de peixes de água doce, para isso os currículos de todos os docentes desses PPGs foram acessados e avaliados quanto a atuação em estudos com peixes de água doce (Tabela S1). O centroide da célula que compreende o local de lotação de docente(s) cujo currículo constava pesquisa com peixes foi utilizado como referência para o cálculo da distância para o centro de pesquisa. Por fim, foi calculada a distância mínima do centróide de cada célula da grade para o centroide da célula mais próxima contendo o centro de pesquisa.

Áreas protegidas

O mapa com as áreas protegidas foi obtido na plataforma do Ministério do Meio Ambiente (<http://mapas.mma.gov.br/>), tendo sido consideradas todas as unidades federais e estaduais de conservação brasileiras. Posteriormente, nós calculamos o percentual de área protegida em cada célula da grade. Somente as células que continham mais de 50% de sua área coberta por área protegida foram consideradas para essa análise. Calculamos então a distância mínima do centroide de cada célula onde houve esforço de amostragem para o centroide da célula mais próxima contendo área protegida.

Vias de acesso

Para analisar se a distribuição do esforço de amostragem está associada com a proximidade de vias de acesso (estradas de chão, ruas e rodovias), os mapas de rotas de acesso foram obtidos a partir de diferentes plataformas: Ministério do Meio Ambiente (<http://mapas.mma.gov.br/>), Sistema Estadual de Geoinformação (<http://www.sieg.go.gov.br/>) e Instituto Brasileiro de Geografia e Estatística (www.ibge.gov.br). Para quantificar as vias de

acesso utilizamos duas métricas: i) a distância do centroide de cada célula para o centroide da célula mais próxima contendo vias de acesso; ii) a densidade de vias de acesso, estimada por Kernel (no qual cada uma das observações é ponderada pela distância em relação a um valor central) no programa QGIS 2.18, para cada célula onde houve esforço de coleta.

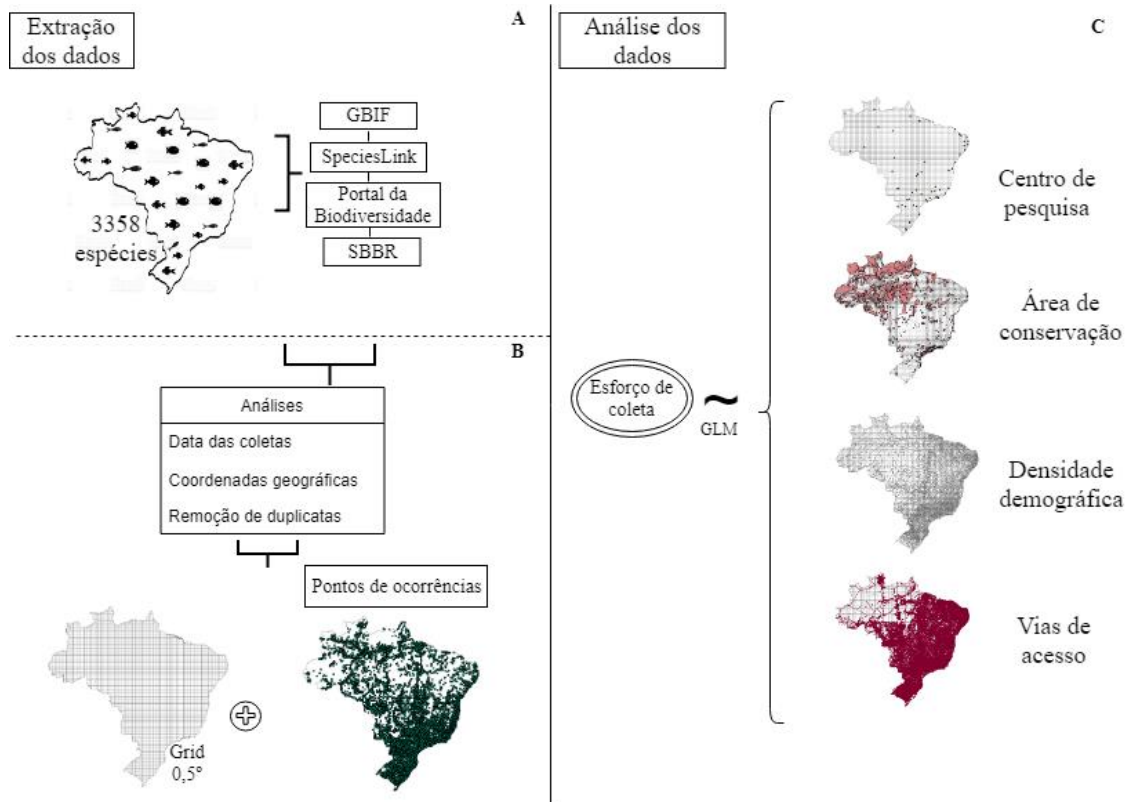


Figura 1: Esquema representativo da metodologia empregada. A) Extração dos dados; B) Filtragem dos dados; C) Análise estatística.

Análise dos dados

Primeiramente, foi realizado um teste de colinearidade (correlação par a par) entre as variáveis predictoras para o Brasil e para suas regiões, utilizando a função *stats* no programa R, foi utilizado como critério para considerar não colinear ter um valor de correção menor que 0,5 (R Core Team 2017). Posteriormente, analisamos a associação da distância para centros de pesquisa, densidade e distância das vias de acesso, distância para áreas protegidas e da densidade populacional com o esforço de coleta, previamente padronizados, por meio da transformação z, onde os valores oscilam de -3 a +3, utilizando quatro tipos de regressões (*Generalized linear model (GLM)*; *Linear regression (LM)*, *Generalized least square (GLS)* e

Zero-Inflated negative binomial regression), para analisar qual modelo de regressão melhor se aplicava aos dados. Os resultados não diferiram qualitativamente entre os modelos e verificamos que o GLM teve melhor ajuste, de acordo com o Critério de Informação de Akaike (AIC) (Tabela S13). Dessa forma, por parcimônia, mostramos somente os resultados das análises usando GLM, família gaussiana. As análises foram realizadas considerando-se todo o território brasileiro e, posteriormente, considerando cada região do Brasil separadamente (Norte, Nordeste, Centro-Oeste, Sul e Sudeste), visando explorar as possíveis diferenças nos fatores associados aos vieses entre as regiões (Figura 1 C).

RESULTADOS

A busca nos bancos de dados online resultou em 339.246 mil registros de ocorrências de espécies, sendo que, após a remoção de dados duplicados e que não continham o ano da coleta, coordenadas geográficas e ocorrências que não pertenciam ao Brasil, restaram 247.169 mil registros. Os bancos de dados online *SpeciesLink* e GBIF foram os mais completos em número de registros, apresentando praticamente a mesma quantidade de registros (Tabela S14), com registros a partir de 1964 e 1900, respectivamente. O agrupamento dos dados de ocorrência das espécies em eventos de coleta em cada célula resultou em 43.538 eventos de amostragem. Os eventos de amostragem estão distribuídos de forma desigual no Brasil, sendo maior nas regiões Sul e Sudeste, embora em todas as regiões tenham sido registrados locais com grande esforço amostral (e.g., Manaus e entorno na região Norte) (Figura 2).

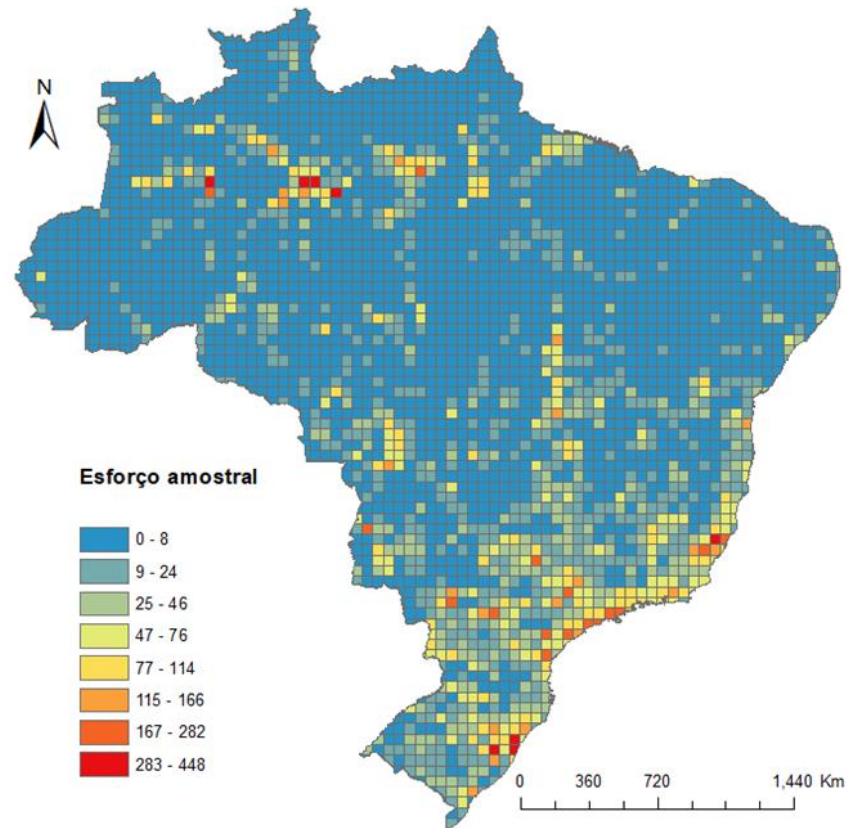


Figura 2: Distribuição do esforço de amostragem (medido a partir da soma do número de coletas em anos e coordenadas geográficas diferentes em cada célula) de peixes de água doce no Brasil.

As variáveis avaliadas não foram colineares (Tabela S2), assim realizamos as análises considerando todas as variáveis preditoras. Tendo o Brasil como extensão, encontramos que a proximidade com centros de pesquisa, proximidade com área protegida, densidade demográfica e densidade de vias de acesso foram significativamente associadas com os esforços de amostragem de peixes de água doce ($p < 0,05$; $R^2 = 0,10$). A maior explicação foi encontrada para o viés relacionado à centros de pesquisa que correspondeu a 6,98% da variação no esforço de amostragem, apresentando uma relação negativa, ou seja, quanto menor à distância para o centro de pesquisa, maior o esforço amostral; seguida pela densidade de vias de acesso e densidade populacional (1,64 % e 1,51%, respectivamente), apresentando relação positiva, ou seja, quanto maior as densidades de vias de acesso e populacional, maior o esforço amostral. As demais variáveis apresentaram valores de explicação próxima de zero. (Tabela 1).

Tabela 1. Resultados do modelo GLM e valor explicativo para cada fator associado ao viés de amostragem.

Variáveis	Coefficiente	Erro padrão	Valor de P	Importância (%)*
Centro de pesquisa	-0,2237	0,01917	<2e-16	6,98
Distância das vias	-0,03899	0,02981	0,08	0,56
Densidade de vias	0,1583	0,02195	<4e-13	1,51
Área protegida	-0,1203	0,02025	<2,78e-9	0,40
Densidade demográfica	0,13	0,01735	<8,80e-14	1,64

*O valor da importância foi calculado através da soma dos quadrados de cada variável dividida pelo seu total e multiplicado por 100.

Para cada região do Brasil, as variáveis preditoras também não apresentaram colinearidade (Tabela S3 a S7). Considerando as regiões separadamente, as regiões Sul e Sudeste foram as que tiveram maior proporção da variação no esforço de amostragem explicada pelas preditoras (20% e 26%, respectivamente). As demais regiões tiveram o esforço de amostragem pobremente explicado pelo modelo (Tabela 2). A distância para o centro de pesquisa foi importante em todas as regiões, sendo o principal determinante dentre os avaliados, para as regiões Norte, Nordeste, Centro-Oeste e Sudeste. A distância para vias de acesso foi importante apenas na região Norte, embora com baixo poder preditivo; a densidade de vias de acesso foi importante nas regiões Centro-Oeste e Sul; distância para área protegida foi importante em todas as regiões, com exceção para a região Norte, sendo o principal preditor na região Sul; a densidade demográfica foi importante para todas as regiões, exceto para a região Nordeste (Tabela 2, Figura 3 e Tabela S8 a S12).

Tabela 2: Proporção da importância (%) atribuída às variáveis preditoras por região do Brasil. Foram consideradas as variáveis que tiveram valores significativos para o GLM.

Regiões	Centro de pesquisa	Distância para via	Densidade de vias	Área protegida	Densidade demográfica	R ^{2*}
Norte	1,83	1,22	--	--	1,76	0,05
Nordeste	5,60	--	--	1,19	--	0,07
Centro Oeste	4,53	--	1,10	2,61	0,81	0,09
Sul	7,13	--	0,76	9,69	7,70	0,26
Sudeste	14,90	--	--	1,67	3,34	0,20

*R²= coeficiente de determinação, que mede o componente da regressão que decorre da variação das variáveis estudadas.

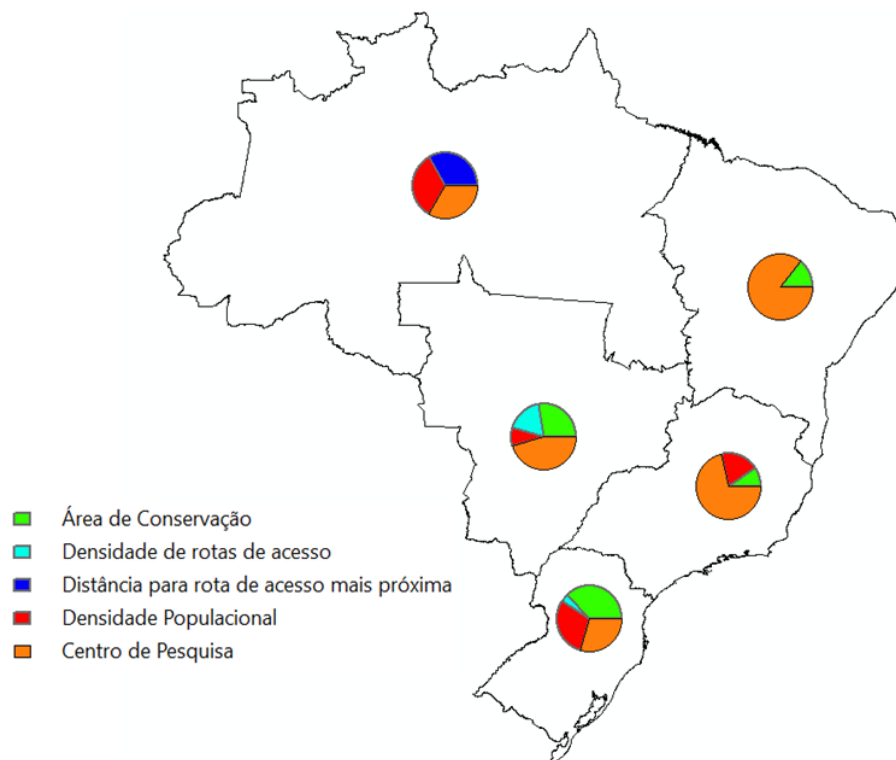


Figura 3: Importância relativa (porcentagem da variação explicada pelas preditoras) das variáveis relacionadas aos vieses espaciais na coleta de dados de peixes de água doce para cada região do Brasil.

DISCUSSÃO

O incremento nos esforços de amostragem ao longo do tempo ocorre para quase todos os grupos, incluindo os peixes de água doce (Dias et al., 2016), resultando em um número expressivo de registros de espécies. Porém nossos resultados evidenciam que essas amostragens de peixes de água doce estão espacialmente enviesadas no Brasil, com uma tendência de maior esforço nas regiões Sul e Sudeste. A proximidade para centros de pesquisa foi o principal fator associado com o esforço de amostragem, destacando a importância de infraestrutura de pesquisa para o conhecimento da ictiofauna.

As diferentes regiões do Brasil variaram quanto a importância dos fatores geradores de vieses. As regiões Sul e Sudeste apresentaram maior previsibilidade na distribuição do esforço de amostragem. A baixa explicação dos modelos nas demais regiões pode ter sido afetada pelo fato de que há muitas áreas onde não há qualquer esforço de amostragem, o que gera grande ruído nas análises e reduz a sua previsibilidade. É notável o pequeno esforço de amostragem em algumas regiões, como o Nordeste. Nesta região, em especial, a maior parte dos centros de pesquisa, infraestrutura e recursos humanos fica concentrada no litoral, contrastando com a malha hidrográfica de água doce que se estende ao interior. Na região Norte, as amostragens são principalmente associadas aos grandes rios (e.g. Amazonas, Tapajós, Madeira, Negro e Solimões) que, além de representarem importantes rotas para deslocamento nesta região, também atraem pesquisadores pela rica biodiversidade e endemismo que apresentam (Siqueira-Souza & Freitas, 2004; Camargo & Giarrizzo, 2007).

A distância para os centros de pesquisa foi o principal fator explicando o esforço de amostragem, tendo sido importante em todas as regiões do Brasil, com maior importância relativa nas regiões Sul e Sudeste. Essas regiões somam 30 centros de pesquisa, número que corresponde a metade do total encontrado para o Brasil. Resultados semelhantes foram encontrados para besouros aquáticos (Sánchez-Fernández et al., 2008), mostrando que a tendência de concentrar amostragens próximas a centros de pesquisa é um viés importante para o ambiente aquático. A busca por locais de amostragem próximos ao local de trabalho é uma conveniência que torna a pesquisa mais barata e com maior facilidade logística para o pesquisador (Reddy & Dávalos, 2003; Romo et al., 2006; Sánchez-Fernández et al., 2008). Esses resultados também sugerem que a especialização na fixação de pesquisadores tem papel importante na distribuição espacial dos esforços de amostragem.

O desenvolvimento de infraestrutura de vias de acesso aumenta a acessibilidade e o número de coletas próximas a esses locais (Kadmon et al., 2004; Romo et al., 2006;

Hortal et al., 2007; Everill et al., 2014). A concentração dos registros das espécies em locais com presença de vias de acesso, maior densidade de vias ou com pequena distância para a via mais próxima é um viés importante e bem conhecido para registros de organismos terrestres (Romo et al., 2006; Oliveira et al., 2016; Daru et al., 2017). Nossos resultados também destacam a associação das vias de acesso para a amostragem no ambiente aquático, embora com um baixo poder preditivo, sendo importante nas regiões Norte, Centro Oeste e Sul. A região Norte foi a única com resultados significativos para o viés distância para via mais próxima. Nessa região a malha viária é menos desenvolvida que em outras regiões, as quais devem ser mais aproveitadas para acessos pelos pesquisadores. Por outro lado, nas demais regiões, a malha viária é bem desenvolvida, abrangendo grande parte do território. Com isso, em muitos locais acessíveis há pouco ou nenhum esforço amostral, o que reduz a previsibilidade das amostragens a partir desse fator. A realização de coletas preferencialmente próximo às vias de acesso, além de gerar viés espacial, pode ocasionar uma falha na representatividade de ambientes distintos (Kadmon et al., 2004; Oliveira et al., 2016). Adicionalmente, a presença de estradas pode alterar as condições ambientais locais e o padrão natural de distribuição das espécies, de forma que, o viés por acessibilidade pode levar a vieses taxonômicos nas amostragens, ou seja, maior probabilidade de se coletar espécies ou grupos menos sensíveis aos efeitos negativos das rodovias (Laurance, 2009; Griffith et al., 2010; Li et al., 2014).

Em locais mais povoados há a tendência de maior amostragem da biodiversidade (Luck, 2007). Isso ocorre porque áreas com maior densidade populacional são associadas a grandes centros urbanos, onde há infraestrutura logística e de recursos humanos, o que poderia favorecer os esforços de amostragem. A única região não influenciada por esse viés foi o Nordeste, onde há um grande número de habitantes, mas distribuída de forma desigual, com uma grande concentração na região litorânea, onde os estudos são voltados para o ambiente aquático marinho. Além disso, a urbanização tem despertado interesse dos pesquisadores que concentram suas pesquisas nesses locais, uma vez que esses ecossistemas urbanizados podem apresentar uma heterogeneidade de habitats, causados pelo distúrbio da existência humana (Araújo, 2003; Pautasso & McKinney, 2007; Luck, 2007).

A atratividade de determinadas áreas, como por exemplos as áreas de protegidas, pode ser explicada pela tendência dos pesquisadores em buscarem locais menos perturbados, onde supostamente podem ter maior probabilidade de encontrar alta riqueza de espécies (Reddy & Dávalos, 2003; Sánchez-Fernández, 2008). Na região Norte esse viés não foi explicativo, o que pode ser atribuído a grande quantidade de áreas protegidas e suas grandes

extensões que contrasta com o padrão espacialmente concentrado dos esforços de amostragem. De fato, mais da metade do território da região são terras indígenas (54%, dados da Funai, 2018), o que pode inclusive limitar as coletas nessa região. Em regiões mais antropizadas, como o Sul e Sudeste, há poucas áreas com vegetação nativa, a maioria delas localizadas em áreas protegidas. Essas áreas representam um último refúgio para muitas espécies, inclusive espécies ameaçadas de extinção. A busca por essas espécies pode dirigir os esforços de amostragens para essas áreas. Além disso, o planejamento de criação/expansão e manejo de unidades de conservação exige que essas áreas sejam minimamente conhecidas o que é, em alguns casos, estimulado por programas específicos de incentivo à pesquisa dessas áreas (p.ex., Fundação Grupo Boticário).

Bancos de dados online são bons recursos para disponibilização e acesso de dados, porém a sua atualização depende da iniciativa dos pesquisadores (Amano et al., 2016; Girardello et al., 2018). Dessa forma, é possível que a escassez de eventos de amostragens em determinadas regiões (e.g. Nordeste) não reflita unicamente as tendências no esforço de amostragem, mas a falta de disponibilização dos dados por parte dos pesquisadores. Isso pode ser atribuído a vários motivos, entre eles a falta de financiamento, tempo e pessoas para digitalização de dados presentes em coleções de museus de história natural, ou a falta dessas coleções (Vollmar, 2010; Beaman & Cellinese, 2012). Atualmente, a digitalização ocorre de forma oportunista e desigual, sendo realizada de acordo com o interesse dos museus, aumentando o conhecimento sobre as distribuições das espécies de forma enviesada (Vollmar, 2010; Scoble, 2010). Essa prática facilita a obtenção de dados para estudos de identificação, classificação e mapeamento de espécies, além de representar um conjunto extenso de dados sobre a biodiversidade do mundo (Soberón, 2010; Scoble, 2010). Em uma busca pelo número de coleções de peixes por região do Brasil através do site *SpeciesLink* podemos notar que as regiões Sul e Sudeste apresentam mais coleções em relação as demais regiões, observando esse déficit podemos destacar a necessidade de investimento nas coleções e digitalização de dados de todo o país.

Para peixes de água doce, Langeani et al. (2007) destacaram que 50 espécies ainda estavam por serem descritas na ecorregião mais bem amostrada do Brasil (Alto Paraná, no Sudeste). Podemos antecipar então, que muito mais espécies ainda estão por serem descobertas em outras regiões, especialmente naquelas pouco amostradas. Para muitas das descritas o conhecimento sobre a distribuição é igualmente incipiente. Nossos resultados evidenciam que os esforços de amostragem possuem algum grau de previsibilidade e a presença de centros de pesquisa foi o principal fator associado com os esforços de

amostragem. Dessa forma, as superações das lacunas Linneana e Wallaceana dependem de investimento em esforços de amostragens em locais mais distantes de centros de pesquisas e menos acessíveis. Devemos considerar também que outros fatores podem ser responsáveis pelo viés de amostragem no ambiente aquático, uma vez que a explicação, mesmo dos centros de pesquisa, foi baixa. Por exemplo, talvez a distância não linear, tempo de navegação ou tamanho dos rios/lagos sejam importantes e devem ser testados em estudos futuros. Para reduzir/minimizar esses vieses algumas estratégias podem ser empregadas, como por exemplo, planejar as coletas de forma a aumentar a cobertura espacial, investir na criação e manutenção de centros de pesquisa nas diversas regiões do país, incentivar a criação de coleções zoológicas para depósito e digitalização de dados nas diferentes regiões.

REFERÊNCIAS

- AMANO, T., LAMMING, J. D. L., SUTHERLAND, W. 2016. Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. **BioScience**. 66, 5: 393- 400.
- ANDERSON, R. P. 2012. Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. **Annals of the New York Academy of Sciences**. 1-15.
- ARAÚJO, M. 2003. The coincidence of people and biodiversity in Europe. **Global Ecology & Biogeography**. 12: 5- 12.
- AZEVEDO, P.G., MESQUITA, F.O., YOUNG, R.J. 2010. Fishing for gaps in science: a bibliographic analysis of Brazilian freshwater ichthyology from 1986 to 2005. **J. Fish Biol.** 76:2177–2193.
- BEAMAN, R. & CELLINESE, N. 2012. Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. **ZooKeys** 209: 7–17.
- BINI, L. M.; DINIZ-FILHO, J. A. F.; RANGEL, T. F. L. V. B.; BASTOS, R. P. E PINTO, M. P. 2006. Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. **Diversity and Distributions, (Diversity Distrib.)**. 12:475–482.
- BOAKES, E. H. et al. 2010. Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. **PLoS Biology**. 8: 1-11.
- BOTTS, E. A., BAREND F. N. ERASMUS, B. F. N. & ALEXANDER, G. J. 2011. Geographic sampling bias in the South African Frog Atlas Project: implications for conservation planning. **Biodivers Conserv.** 20:119–139.

- BROOKS, T. H., FONSECA, G. A. B., E RODRIGUES, A. S. L. 2004. Protected areas and species. **Conservation Biology**. 18, 616–618.
- CAMARGO, M. & GIARRIZZO, T. 2007. Fish, Marmelos Conservation Area (BX044), Madeira River basin, states of Amazonas and Rondônia, Brazil. **Check list**. 3:291–296.
- DARU, B. H. et al. 2017. Widespread sampling biases in herbaria revealed from large-scale digitization. **New Phytologist**. 217: 939–955.
- DIAS, M. S. et al. 2016. Trends in studies of Brazilian stream fish assemblages. **Natureza & Conservação**. 14, 2:106-111.
- EVERILL PH, PRIMACK RB, ELLWOOD EE, MELLAAS EK. 2014. Determining past leafout times of New England’s deciduous forests from herbarium specimens. **American Journal of Botany**. 101: 1–8.
- Fundação Nacional do Índio- FUNAI. Terras indígenas: o que é? Disponível em: <<http://www.funai.gov.br/index.php/nossas-acoess/demarcacao-de-terras-indigenas?start=2#>>. Acessado em: jan. 2019.
- GIRARDELLO, M., MARTELOS, S., PARDO, A. & BERTOLINO, S. 2018. Gaps in biodiversity occurrence information may hamper the achievement of international biodiversity targets: insights from a cross-taxon analysis. **Environmental Conservation**. 1-8.
- GRIFFITH E. H., SAUER J. R. & ROYLE J. A. 2010. Traffic effects on bird counts on North American breeding bird survey routes. **Auk**. 127: 387–393.
- HECNAR, S. J., & R. T. M’CLOSKEY. 1997. Spatial scale and determination of species status of the green frog. **Conservation Biology**. 11:670–682.
- HORTAL, J., LOBO, J.M. & JIMENEZ-VALVERDE, A. 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. **Conservation Biology**. 21:853–863.
- HORTAL, J., BELLO, F., DINIZ-FILHO, J. A., 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.
- KADMON, R., FARBER, O. & DANIN, A. 2004. Effects of roadside bias on the accuracy of predictive maps produced by bioclimatic models. **Ecological Applications**. 14, 401–413.
- LANGANI, F. et al. 2007. Diversidade da ictiofauna do Alto Rio Paraná: composição atual e perspectivas futuras. **Biota Neotropica**. 7, 3: 181-197.
- LAURANCE, W.F. 2009. Roads to ruin: expanding transportation networks imperil global biodiversity. **The multiple faces of globalization (ed. by C. Gandarias)**. 198–211.
- LÉVÊQUE, C., BALIAN, E.V. & MARTENS, K. 2005. An assessment of animal species diversity in continental waters. **Hydrobiologia**. 542, 39–67.

- LI Y, YU J, NING K, DU S, HAN G, QU F, WANG G, FU Y, ZHAN C. 2014. Ecological effects of roads on the plant diversity of coastal wetland in the Yellow River Delta. **Scientific World Journal**. 2014: 952051.
- LOMOLINO, M.V. 2004. Conservation biogeography. *Frontiers of Biogeography: new directions in the geography of nature* (ed. by M.V. Lomolino and L.R. Heaney), **Sinauer Associates**, Sunderland, Massachusetts, 293–296.
- LUCK, G. 2007. A review of the relationships between human population density and biodiversity. **Biol. Rev.** 82, 607–645.
- MEYER, C., KREFT, H., GURALNICK, R. & JETZ, W. 2015. Global priorities for an effective information basis of biodiversity distributions. **Nature**. 6:8221.
- OKSANEN, J. F. BLANCHET, G., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P. R., O'HARA, R. B., SIMPSON, G. L., SOLYMOS, P., STEVENS, M. H. H., SZOECS, E. & WAGNER, H. 2018. vegan: Community Ecology Package. R package version 2.5-1. <https://CRAN.R-project.org/package=vegan>.
- OLIVEIRA, U. et al. 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. **Diversity and Distributions**. 1–13.
- PAUTASSO, M. & MCKINNEY, M. L. 2007. The botanist effect revisited: Plant species richness, county area and human population size in the United States. **Conserv Biol**. 21: 1333–1340.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- REDDY, S. & DÁVALOS, L. M. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. **Journal of Biogeography**. 30:1719–1727.
- ROMO, H.; BARROS, E. G. & LOBO, J. M. 2006. Identifying recorder-induced geographic bias in an Iberian butterfly database. **Ecography**. 29: 873-885.
- SÁNCHEZ-FERNÁNDEZ, D., et al. 2008. Bias in freshwater biodiversity sampling: the case of Iberian water beetles. **Diversity and Distributions**. 14: 754–762.
- SASTRE, P., LOBO, J.M. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. **Biol. Conserv**. 142: 462–467.
- SCOBLE, M. 2010. Natural history collections digitization: rationale and value. **Biodiversity Informatics**. 7: 77 – 80.
- SIQUEIRA-SOUZA, F. K. & FREITAS, C. E. C. 2004. Fish diversity of floodplain lakes on the lower stretch of the Solimões River. **Braz. J. Biol**. 64: 501- 510.
- SOBERÓN J. M. 2010. Niche and area of distribution modeling: a population ecology perspective. **Ecography**. 33: 159–167.

STROPP, J. et al. 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. **Global Ecology and Biogeography**. 25: 1085–1096.

TROIA, M. & MCMANAMAY, R. A. 2017. Completeness and coverage of open-access freshwater fish distribution data in the United States. **Biodiversity Research**. 1-17.

VIEIRA, T. B., et al. 2018. A multiple hypothesis approach to explain species richness patterns in neotropical stream-dweller fish communities. **Plos one**. 13, 9.

VOLLMAR, A., MACKLIN, J. A., AND FORD, L.S. 2010. Natural history specimen digitization: challenges and concerns, **J. Biodiversity Informatics**. 7: 27-46.

WHITTAKER, R. J., et al. 2005. Conservation Biogeography: assessment and prospect. **Diversity and Distributions**. 11:3–23.

YANG, W., KEPING MA, K. & KREFT, H. 2013. Geographical sampling bias in a large distributional database and its effects on species richness–environment models. **Journal of Biogeography (J. Biogeogr.)**.

MATERIAL SUPLEMENTAR

Tabela S1: Centros de pesquisas que possuem pesquisadores em biodiversidade de peixes de água doce

Centros de pesquisa

Centro Universitário de Anápolis (UNIEVANGELICA)
 Fundação Oswaldo Cruz
 Fundação Universidade Federal de Mato Grosso do Sul (UFMS)
 Fundação Universidade Federal de Sergipe (FUFSE)
 Instituto Nacional de Pesquisas da Amazônia (INPA)
 Museu Paraense Emílio Goeldi (MPEG)
 Pontifícia Universidade Católica de Minas Gerais (PUC/MG)
 Pontifícia Universidade Católica do Rio Grande do Sul (PUC/RS)
 Universidade Anhanguera (UNIDERP)
 Universidade de Brasília (UNB)
 Universidade de São Paulo (USP)
 Universidade de Taubaté (UNITAU)
 Universidade do Estado de Mato Grosso (UNEMAT)
 Universidade do Estado do Rio de Janeiro (UERJ)
 Universidade do Estado do Rio Grande do Norte (UERN)
 Universidade do Vale do Rio dos Sinos (UNISINOS)
 Universidade Estadual da Paraíba (UEPB)
 Universidade Estadual de Goiás (UEG)
 Universidade Estadual de Londrina (UEL)
 Universidade Estadual de Maringá (UEM)
 Universidade Estadual de Ponta Grossa (UEPG)
 Universidade Estadual de Roraima (UERR)
 Universidade Estadual de Santa Cruz (UESC)
 Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF)
 Universidade Estadual do Oeste do Paraná (UNIOESTE)
 Universidade Estadual do Sudoeste da Bahia (UESB)
 Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)
 Universidade Federal da Bahia (UFBA)

Universidade Federal da Fronteira Sul (UFFS)
Universidade Federal da Integração Latino-Americana (UNILA)
Universidade Federal da Paraíba (UFPB)
Universidade Federal de Alagoas (UFAL)
Universidade Federal de Goiás (UFG)
Universidade Federal de Lavras (UFLA)
Universidade Federal de Mato Grosso (UFMT)
Universidade Federal de Minas Gerais (UFMG)
Universidade Federal de Pelotas (UFPEL)
Universidade Federal de Pernambuco (UFPE)
Universidade Federal de Rondônia (UNIR)
Universidade Federal de Santa Catarina (UFSC)
Universidade Federal de São Carlos (UFSCAR)
Universidade Federal de São João Del-Rei (UFSJ)
Universidade Federal de Viçosa (UFV)
Universidade Federal do Acre (UFAC)
Universidade Federal do Amapá (UNIFAP)
Universidade Federal do Amazonas (UFAM)
Universidade Federal do Ceará (UFC)
Universidade Federal do Espírito Santo (UFES)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Universidade Federal do Oeste do Pará (UFOPA)
Universidade Federal do Pará (UFPA)
Universidade Federal do Paraná (UFPR)
Universidade Federal do Rio de Janeiro (UFRJ)
Universidade Federal do Rio Grande (FURG)
Universidade Federal do Rio Grande do Norte (UFRN)
Universidade Federal do Rio Grande do Sul (UFRGS)
Universidade Federal do Sul da Bahia (UFESBA)
Universidade Federal do Tocantins (UFT)
Universidade Federal Rural de Pernambuco (UFRPE)
Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Universidade Federal Rural do Semi-Árido (UFERSA)

 Universidade Santa Cecília (UNISANTA)

 Universidade Vila Velha (UVV)

Tabela S2: Teste de colinearidade entre as variáveis preditoras para o Brasil.

	Esforço amostral	Centro de pesquisa	Densidade demográfica	Distância vias	Área protegida	Densidade de vias
Esforço amostral	1	-0.26	0.16	-0.14	0.06	0.20
Centro de pesquisa	-0.26	1	-0.16	0.28	-0.36	-0.33
Densidade demográfica	0.16	-0.16	1	-0.07	0.10	0.06
Distância vias	-0.14	0.28	-0.07	1	-0.38	-0.53
Área protegida	0.06	-0.36	0.10	-0.38	1	0.46
Densidade de vias	0.20	-0.33	0.06	-0.53	0.46	1

Tabela S3: Teste de colinearidade entre as variáveis preditoras da região Nordeste.

	Esforço amostral	Centro de pesquisa	Densidade demográfica	Distância vias	Área protegida	Densidade de vias
Esforço amostral	1	-0.24	0.03	-0.07	0.04	0.09
Centro de pesquisa	-0.24	1	-0.17	0.11	-0.52	-0.05
Densidade demográfica	0.03	-0.17	1	-0.08	0.18	-0.17
Distância vias	-0.07	0.11	-0.08	1	-0.19	-0.29
Área protegida	0.04	-0.52	0.18	-0.19	1	0.10
Densidade de vias	0.09	-0.05	-0.17	-0.29	0.10	1

Tabela S4: Teste de colinearidade entre as variáveis preditoras da região Norte.

	Esforço amostral	Centro de pesquisa	Densidade demográfica	Distância vias	Área protegida	Densidade de vias
Esforço amostral	1	-0.14	0.18	-0.13	0.11	0.15
Centro de pesquisa	-0.14	1	-0.17	0.18	-0.08	-0.18
Densidade demográfica	0.18	-0.17	1	-0.13	0.13	0.27
Distância vias	-0.13	0.18	-0.13	1	-0.28	-0.52
Área protegida	0.11	-0.08	0.13	-0.28	1	0.54
Densidade de vias	0.15	-0.18	0.27	-0.52	0.54	1

Tabela S5: Teste de colinearidade entre as variáveis preditoras da região Centro Oeste.

	Esforço amostral	Centro de pesquisa	Densidade demográfica	Distância vias	Área protegida	Densidade de vias
Esforço amostral	1	-0.21	0.19	-0.09	-0.08	0.20
Centro de pesquisa	-0.21	1	-0.28	0.29	-0.31	-0.49
Densidade demográfica	0.19	-0.28	1	-0.08	-0.03	0.30
Distância vias	-0.09	0.29	-0.08	1	-0.27	-0.44
Área protegida	-0.08	-0.31	-0.03	-0.27	1	0.22
Densidade de vias	0.20	-0.49	0.30	-0.44	0.22	1

Tabela S6: Teste de colinearidade entre as variáveis preditoras da região Sudeste.

	Esforço amostral	Centro de pesquisa	Densidade demográfica	Distância vias	Área protegida	Densidade de vias
Esforço amostral	1	-0.39	0.30	-0.08	-0.17	0.08
Centro de pesquisa	-0.39	1	-0.26	0.20	0.07	-0.23
Densidade demográfica	0.30	-0.26	1	-0.13	-0.17	0.08
Distância vias	-0.08	0.20	-0.13	1	0.00	-0.31
Área protegida	-0.17	0.07	-0.17	0.00	1	-0.05
Densidade de vias	0.08	-0.23	0.08	-0.31	-0.05	1

Tabela S7: Teste de colinearidade entre as variáveis preditoras da região Sul.

	Esforço amostral	Centro de pesquisa	Densidade demográfica	Distância vias	Área protegida	Densidade de vias
Esforço amostral	1	-0.27	0.38	-0.11	-0.30	0.20
Centro de pesquisa	-0.27	1	-0.29	0.01	-0.03	-0.24
Densidade demográfica	0.38	-0.29	1	-0.13	-0.13	-0.04
Distância vias	-0.11	0.01	-0.13	1	0.00	-0.18
Área protegida	-0.30	-0.03	-0.13	0.00	1	-0.11
Densidade de vias	0.20	-0.24	-0.04	-0.18	-0.11	1

Tabela S8: Resultados do modelo GLM para cada variável explanatória do viés de amostragem da região Nordeste.

	Estimativa	Erro quadrado	Valor de P
Centro de pesquisa	-0.30	0.05	<0.05
Distância via de acesso	-0.03	0.04	0.42

Área protegida	-0.14	0.05	<0.05
Densidade de vias	0.08	0.04	0.052
Densidade demográfica	0.01	0.04	0.79

Tabela S9: Resultados do modelo GLM para cada variável explanatória do viés de amostragem da região Norte.

	Estimativa	Erro quadrado	Valor de P
Centro de pesquisa	-0.09	0.03	<0.05
Distância via	-0.07	0.03	<0.05
Área protegida	0.05	0.03	0.13
Densidade de vias	0.04	0.04	0.27
Densidade demográfica	0.14	0.03	<0.05

Tabela S10: Resultados do modelo GLM para cada variável explanatória do viés de amostragem da região Centro Oeste.

	Estimativa	Erro quadrado	Valor de P
Centro de pesquisa	-0.17	0.05	<0.05
Distância via de acesso	-0.03	0.04	0.450609
Área protegida	-0.16	0.04	<0.05
Densidade de vias	0.11	0.05	<0.05
Densidade demográfica	0.10	0.04	<0.05

Tabela S11: Resultados do modelo GLM para cada variável explanatória do viés de amostragem da região Sudeste.

	Estimativa	Erro quadrado	Valor de P
Centro de pesquisa	-0.33	0.05	<0.05
Distância via de acesso	0.01	0.05	0.88812
Área protegida	-0.11	0.05	<0.05
Densidade de vias	-0.01	0.05	0.77246
Densidade demográfica	0.19	0.05	<0.05

Tabela S12: Resultados do modelo GLM para cada variável explanatória do viés de amostragem da região Sul.

	Estimativa	Erro quadrado	Valor de P
Centro de pesquisa	-0.15	0.06	<0.05
Distância via de acesso	-0.04	0.05	0.42758
Área protegida	-0.25	0.05	<0.05
Densidade de vias	0.14	0.06	<0.05
Densidade demográfica	0.30	0.06	<0.05

Tabela S13: Teste de Akaike entre os modelos de regressões.

	Teste de Akaike
Modelo linear generalizado- GLM	663.2
Regressão múltipla- LM	663.4
Mínimos quadrados generalizados- GLS	686.9
Zero inflado	72762.3

Tabela S14: Dados coletados por bancos de dados *online*.

Base de dados	Número de registros inicial	Número de registros após análise*
GBIF	215.108	106.161
<i>SpeciesLink</i>	216.816	187.348
Portal da biodiversidade	64.324	64.256
SBBR	55.925	55.903

*esses valores não incluem a remoção de dados duplicados, apenas dados sem data de coleta e coordenadas geográficas.

Mapas das variáveis preditoras



Figura S1: Centros de pesquisas do Brasil com pesquisadores que empreendem esforços com estudos sobre a biodiversidade de peixes de água doce.

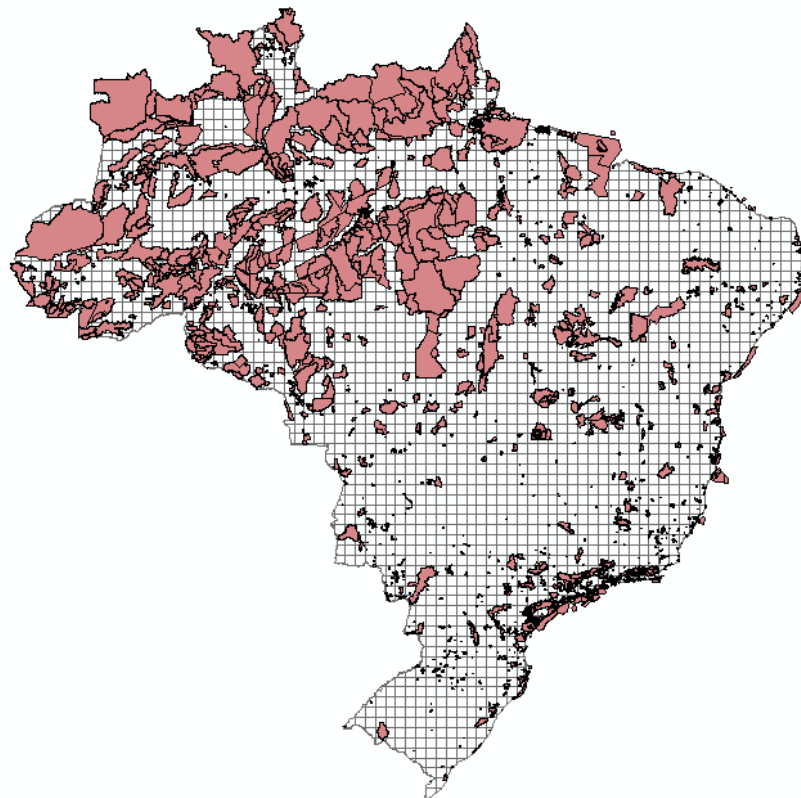


Figura S2: Áreas protegidas do Brasil, sendo consideradas as unidades municipais, estaduais, federais e terras indígenas.

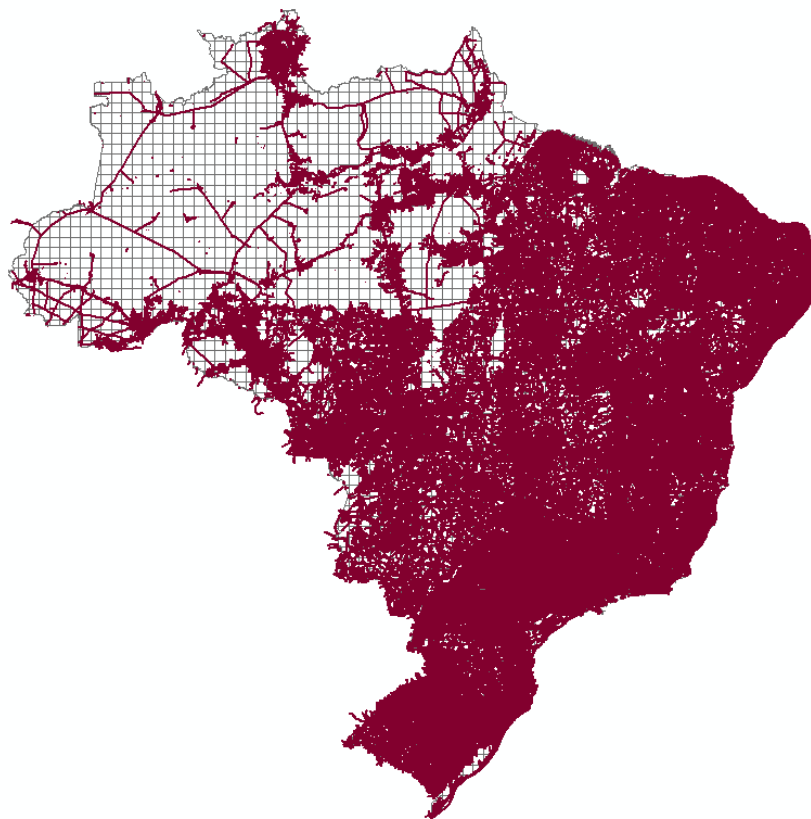


Figura S3: Vias de acesso do Brasil, sendo consideradas estradas de chão, ruas e rodovias.

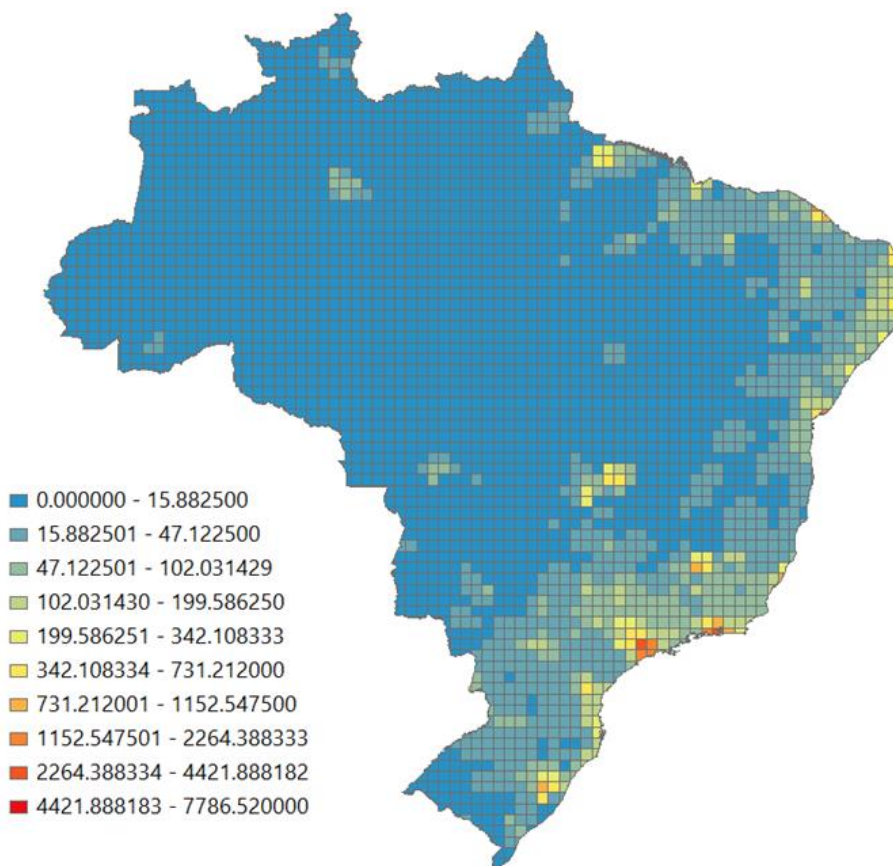


Figura S4: Densidade demográfica por grid 0,5°.