UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO

MICAEL ROSA PARREIRA

CONTRIBUIÇÃO RELATIVA DE PREDITORES DE DISPERSÃO SIMÉTRICA E ASSIMÉTRICA NOS MODELOS DE NICHO ECOLÓGICO EM AMBIENTES AQUÁTICOS

GOIÂNIA
2022

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, 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.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [X] Tese [ ] Outro*:_____________

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Micael Rosa Parreira

3. Título do trabalho

Contribuição relativa de preditores de dispersão simétrica e assimétrica nos modelos de nicho ecológico em ambientes aquáticos

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [x] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **MICAEL ROSA PARREIRA**, **Discente**, em 23/08/2022, às

10:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **João Carlos Nabout**, **Usuário Externo**, em 23/08/2022, às 10:19, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3099018** e o código CRC **7B84E452**.

Referência: Processo nº 23070.035084/2022-98 SEI nº 3099018

MICAEL ROSA PARREIRA

CONTRIBUIÇÃO RELATIVA DE PREDITORES DE DISPERSÃO SIMÉTRICA E ASSIMÉTRICA NOS MODELOS DE NICHO ECOLÓGICO EM AMBIENTES AQUÁTICOS

Tese apresentada ao Programa de Pós-Graduação em Ecologia e Evolução, do Instituto de Ciências Biológicas, da Universidade Federal de Goiás (UFG), como requisito para obtenção do título de Doutor em Ecologia e Evolução.
Área de concentração: Ecologia e Evolução
Linha de pesquisa: Macroecologia e Ecologia Evolutiva

Orientador: Professor Doutor João Carlos Nabout

GOIÂNIA
2022

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Parreira, Micael Rosa
Contribuição relativa de preditores de dispersão simétrica e assimétrica nos modelos de nicho ecológico em ambientes aquáticos [manuscrito] / Micael Rosa Parreira. - 2022.
cxix, 119 f.

Orientador: Prof. Dr. João Carlos Nabout.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Ecologia e Evolução, Goiânia, 2022.
Bibliografia. Apêndice.
Inclui mapas, gráfico, tabelas.

1. Mapas de Autovetores Assimétricos. 2. Dispersão direcional. 3. Restrições de movimento. 4. Hidrelétricas. 5. Diversidade Beta. I. Nabout, João Carlos, orient. II. Título.

CDU 574

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

ATA DE DEFESA DE TESE

Ata Nº **117** da sessão de Defesa de Tese de **Micael Rosa Parreira** que confere o título de Doutor em **Ecologia e Evolução**, na área de concentração em **Ecologia e Evolução.**

Aos **vinte e seis dias do mês de julho do ano de dois mil e vinte e dois (26/07/2022),** a partir das **09h00min**, por videoconferência, seguindo Resolução CONSUNI/UFG Nº 141 de 13 de maio de 2022 e orientações do Ofício Circular no. 34/2022/PRPG/UFG (SEI 23070.030951/2022-07), realizou-se a sessão pública de Defesa de Tese intitulada **"Contribuição relativa de preditores de dispersão simétrica e assimétrica nos modelos de nicho ecológico em ambientes aquáticos".** Os trabalhos foram instalados pelo Orientador, **Prof. Dr. João Carlos Nabout (CCET/UEG)**; com a participação dos demais membros da Banca Examinadora: **Prof. Dr. Matheus de Souza Lima Ribeiro (Depto de Biologia/UFJ)**, membro titular interno; **Profa. Dra. Levi Carina Terribile (Depto de Biologia/UFJ)**, membro titular interno, **Prof. Dr. Daniel de Paiva Silva (Depto de Biologia/IFGoiano)**, membro titular externo, **Prof. Dr. Guilherme de Oliveira (CCBBA/UFRB),** membro titular externo. Durante a argüição os membros da banca **não fizeram** sugestão de alteração do título do **trabalho**. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo **Prof. Dr. João Carlos Nabout**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos **vinte e seis dias do mês de julho do ano de dois mil e vinte e dois (26/07/2022).**

TÍTULO SUGERIDO PELA BANCA

Documento assinado eletronicamente por **Daniel de Paiva Silva**, **Usuário Externo**, em 13/09/2022, às 11:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Matheus De Souza Lima Ribeiro**, **Professor do Magistério Superior**, em 13/09/2022, às 20:07, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Guilherme de Oliveira**, **Usuário Externo**, em 14/09/2022, às 09:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Levi Carina Terribile**, **Professor do Magistério Superior**, em 15/09/2022, às 10:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **João Carlos Nabout**, **Usuário Externo**, em 18/09/2022, às 20:36, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3183948** e o código CRC **1F946F87**.

Referência: Processo nº 23070.035084/2022-98 SEI nº 3183948

Dedicatória

Dedico este trabalho primeiramente a minha esposa que sempre me apoiou e me deu forças pra continuar nessa jornada e também a minha família e amigos, que sempre estiveram ao meu lado nos momentos de alegria e tristeza. Também dedico ao meu pai que sempre me apoiou nos estudos de forma direta ou indireta e que neste momento sei que está descansando em um lugar melhor. Saudades eternas.

Agradecimentos

Algo muito importante na carreira acadêmica se chama colaboração, seja de forma direta, como na publicação de um trabalho acadêmico, ou indireta, por conversas durante um lanche no laboratório que te ajudam na resolução de problemas na sua tese. Neste sentido só tenho a agradecer todos que estiveram ao meu lado durante esse período e espero que tenhamos ainda muitas colaborações. Esses agradecimentos são para vocês.

Primeiramente gostaria de agradecer a minha esposa Alinne Dourado que sempre esteve ao meu lado durante meus estudos durante o doutorado, sempre me apoiando nos momentos de angústia, compartilhando momentos de sorrisos e lágrimas. Também agradeço a minha família e meus amigos pelo apoio direto e indireto, especialmente minha mãe, que mesmo as vezes não entendendo o que faço no doutorado, sempre me apoiou nas decisões que tomei e se orgulha do que me tornei.

Agradeço aos meus colegas de laboratório, Karine, Pedro, Filipe, Rafael e todos os demais pelas ajudas nos trabalhos científicos ou mesmo pelas conversas sobre ciência, vida, política e tantos outros assuntos, que sempre me permitiram espairecer e me divertir em meio a preocupações e prazos apertados. Agradeço também aos meus demais colegas de doutorado da UFG pelos momentos divertidos em salas de aula e fora dela.

Agradeço principalmente ao meu orientador João Carlos Nabout que há vários anos vem me orientando ou coorientando na minha carreira de cientista. A orientação dele vai além da parte acadêmica, me dando conselhos e uma palavra amiga em momentos de desânimo e vontade de desistir. Mesmo em meio a correria da vida acadêmica e profissional de um professor, esteve sempre disposto a me ajudar, sendo o alicerce para o meu desenvolvimento acadêmico e profissional.

Agradeço a Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) pela bolsa de doutorado concedida na primeira parte do meu doutorado e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado concedida para a parte final do doutorado. Além disso, agradeço a Universidade Estadual de Goiás por permitir o uso de equipamentos e da estrutura do Centro de Pesquisas ao qual o Laboratório BioEcol está localizado.

Índice

Resumo

Modelos de distribuição de espécies são constituídos principalmente de dados ambientais (em sua maioria climáticos) e de distribuição das espécies para predizer a distribuição potencial das espécies. Neste sentido, a dispersão das espécies (i.e., movimento) é por muitas vezes ignorada nas predições. Em ambientes aquáticos, a dispersão não é restrita somente às barreiras físicas, mas também se dá pelo movimento direcional das espécies através da malha hidrográfica, que por sua vez pode ser incluída nos modelos por meio de preditores espaciais. Neste trabalho, nós buscamos avaliar o efeito da inclusão de preditores de dispersão assimétrica na distribuição geográfica potencial de um peixe de água doce da bacia Tocantins-Araguaia, Brasil. Além disso, nós estimamos se as predições usando preditores climáticos e de dispersão assimétrica variam dependendo da distribuição e amplitude de múltiplas espécies, e dependendo da inclusão de hidrelétricas como barreiras de dispersão nas predições. Para isso, nós geramos modelos usando sete conjunto de variáveis representando modelos climáticos e de dispersão, assim como suas interações. As métricas de avaliação dos modelos foram usadas para comparar a performance dos diferentes conjuntos de modelos (e.g., assimétrico e simétrico, e ambiental), a performance de modelos de múltiplas espécies baseando-se na distribuição das ocorrências entre sub-bacias e amplitude de distribuição, além do efeito da inclusão de barreiras de dispersão nos modelos. Primeiramente, os modelos com maiores performances foram os gerados usando preditores de dispersão assimétrica, tanto sozinhos como em conjunto com variáveis ambientais. Em segundo, espécies mais restritas tiveram modelos com maior performance quando modeladas usando preditores de dispersão assimétrica, especialmente quando distribuídas em diferentes sub-bacias. Por último, predições incluindo hidrelétricas como barreiras de dispersão mostraram uma maior perda de riqueza e composição de espécies, especialmente para as áreas com o maior número de barragens. Portanto, a inclusão de variáveis de dispersão assimétricas, considerando limites de dispersão das espécies, reduziu a extrapolação de predições para áreas climaticamente adequáveis, porém desconectadas pelos rios. Ademais, os modelos usando dispersão assimétrica melhor representaram espécies restritas em ambas sub-bacias e também o efeito de barreiras de dispersão na riqueza e composição de espécies ao longo da bacia. Sendo assim, futuros trabalhos usando modelos de distribuição de espécies, especialmente usando grupos com dispersão assimétrica, deveriam considerar a inclusão de preditores de dispersão assimétrica de forma a melhorar a performance dos modelos e realidade ecológica das predições.

Palavras-chave: Mapas de Autovetores Assimétricos; Dispersão direcional; Restrições de movimento; Hidrelétricas; Diversidade Beta.

Abstract

Species distribution models are based mainly on environmental (mostly climatic) and species distribution data to predict the potential distribution of species. In this sense, the species dispersal (i.e., movement) is often ignored in their predictions. In freshwater habitats, species dispersal is not restricted only by physical barriers but also by the directional movement of the hydrographic network, which can be considered through spatial predictors. Here, we aim to evaluate the effect of including asymmetrical dispersal predictors in the potential geographic distribution of a freshwater fish in the Tocantins-Araguaia River basin, Brazil. Furthermore, we aim to assess if the predictions using climatic and asymmetrical dispersal vary depending on multiple species occurrence distribution and range, and on the inclusion of hydropower plants as dispersal barriers in the predictions. For this, we built models with seven variable sets representing environmental (climatic) and dispersal models, as well as their interactions. The models' accuracy metrics were then used to compare the performance of different model sets (e.g., asymmetrical and symmetrical dispersal and environmental predictors), the performance of multiple species models based on their occurrence distribution among sub-basins and range, and the effect of including dispersal barriers into the models. First, we found that the models with higher performance are those built using asymmetrical dispersal predictors, either solo or combined with environmental variables. Second, species more restricted had models with higher performance when modeled using asymmetrical dispersal predictors, especially when distributed in different sub-basins. Third, predictions including the hydropower plants as dispersal barriers showed a higher loss of species richness and composition, especially for the areas with the highest number of dams. Therefore, the inclusion of asymmetrical dispersal variables, taking into account dispersal limitations of species, decreased the overprediction to climatically suitable but disconnected areas through rivers. Furthermore, those models using asymmetrical dispersal better represented restricted species distributed in both sub-basins of the basin and also the effect of dispersal barriers in the fish species richness and composition along the basin. Therefore, future SDM studies, especially those using species groups with asymmetrical dispersal, should consider the inclusion of asymmetrical dispersal predictors to increase the model's accuracy and ecological reality of predictions.

Keywords: Asymmetric Eigenvector Maps; Directional dispersal; Movement constraints; Hydropower plants; Beta diversity.

Introdução Geral

Atualmente existem diversas abordagens e metodologias que têm sido implementadas nos modelos de distribuição de espécies para melhorar o ajuste e poder de precisão dos mesmos (Araújo et al., 2019; Austin and van Niel, 2011; Howard et al., 2014). Muitos dos pressupostos dos modelos como dispersão livre para áreas adequadas ambientalmente para a espécie não levam em conta a rota de dispersão das espécies ou ainda possível barreiras (naturais ou antropogênicas) nas rotas que possam inviabilizar a migração das espécies para essas áreas (Franklin, 2010; Miller and Holloway, 2015).

Essas barreiras, também conhecidas como restrições de movimento, podem atuar nos modelos de distribuição de espécies como um limitante de dispersão ou taxa de migração, sendo adicionadas em modelos tanto *a priori* (i.e., como preditores) ou *a posteriori* (i.e., sobrepondo áreas acessíveis ou inacessíveis com as áreas preditas) (Mendes et al., 2020). Em geral essas restrições não levam em conta a dispersão direcional ou assimétrica das espécies pela dificuldade de se delinear exatamente as rotas de dispersão das espécies. Contudo, para muitas espécies de peixe de água doce, as rotas são bem definidas visto a estrutura hierárquica e direção de fluxo dos rios ao qual as espécies utilizam para se dispersar para áreas adequadas (Domisch et al., 2015).

Nesse sentido, aqui evidenciamos o uso de filtros espaciais assimétricos (Asymmetric Eigenvector Maps [AEM]) (Blanchet et al., 2008) que podem ser usados como substitutos de dispersão na estruturação de metacomunidades aquáticas (e.g., Rocha et al., 2020). Para isso, nós adaptamos os filtros AEM para serem usados como preditores espaciais de dispersão assimétrica na modelagem da distribuição das espécies. A criação desses filtros se deu pela geração de uma matriz binária de conectividade (0 e 1) entre os rios e tributários da bacia do Tocantins-Araguaia, onde as conexões dos rios (edges) se encontram nas colunas e as células (nodes) nas linhas da matriz. Posteriormente essa matriz foi utilizada como base da criação dos filtros AEM, que posteriormente foram selecionados um número ideal de autovetores para serem utilizados como variáveis de dispersão assimétrica na criação dos modelos.

Portanto, aqui neste trabalho nós evidenciamos a importância do uso de variáveis de dispersão assimétrica nos modelos de distribuição de espécies, comparando os modelos gerados usando essas variáveis com modelos usando variáveis de dispersão simétrica (PCNM) e variáveis climáticas (Bioclim), usando critérios de avaliação de modelos e estrapolação nas predições. Além disso, estimamos os efeitos das variáveis de dispersão assimétricas e ambientais na distribuição potencial de múltiplas espécies de peixes de água doce da bacia, comparando a performance dos modelos dependendo da amplitude de distribuição da espécie e da sub-bacia ao qual a mesma está localizada. Por último, nós utilizamos os preditores de dispersão assimétrica como uma ferramenta de se inserir informações de barreiras de dispersão, neste caso usinas hidroelétricas, *a priori*, ou seja, como preditores no momento da criação dos modelos, avaliando o efeito dessas restrições de movimento na acurácia das predições dos modelos.

Referências

Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. Science Advances 5, 4858–4874. https://doi.org/10.1126/sciadv.aat4858

Austin, M.P., van Niel, K.P., 2011. Improving species distribution models for climate change studies: variable selection and scale. Journal of Biogeography 38, 1–8. https://doi.org/10.1111/j.1365-2699.2010.02416.x

Blanchet, F.G., Legendre, P., Borcard, D., 2008. Modelling directional spatial processes in ecological data. Ecological Modelling 215, 325–336. https://doi.org/10.1016/j.ecolmodel.2008.04.001

Domisch, S., Jähnig, S.C., Simaika, J.P., Kuemmerlen, M., Stoll, S., 2015. Application of species distribution models in stream ecosystems: the challenges of spatial and temporal scale, environmental predictors and species occurrence data. Fundamental and Applied Limnology 186, 45–61. https://doi.org/10.1127/fal/2015/0627

Franklin, J., 2010. Moving beyond static species distribution models in support of conservation biogeography. Diversity and Distributions 16, 321–330. https://doi.org/10.1111/j.1472-4642.2010.00641.x

Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D., Willis, S.G., 2014. Improving species distribution models: the value of data on abundance. Methods in Ecology and Evolution 5, 506–513. https://doi.org/10.1111/2041-210X.12184

Mendes, P., Velazco, S.J.E., Andrade, A.F.A. de, De Marco, P., 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

Miller, J.A., Holloway, P., 2015. Incorporating movement in species distribution models. Progress in Physical Geography: Earth and Environment 39, 837–849. https://doi.org/10.1177/0309133315580890

Rocha, B. da S., Souza, C.A. de, Machado, K.B., Vieira, L.C.G., Nabout, J.C., 2020. The relative influence of the environment, land use, and space on the functional and taxonomic structures of phytoplankton and zooplankton metacommunities in tropical reservoirs. Freshwater Science 39, 321–333. https://doi.org/10.1086/708949

CAPÍTULO 1: Incorporating symmetrical and asymmetrical dispersal into Species Distribution Models in freshwater environments

Micael Rosa Parreira[1*], Geiziane Tessarolo[2], João Carlos Nabout[2]

[1]Universidade Federal de Goiás, Campus Samambaia, 74690-900, Goiânia, GO, Brazil.

[2]Universidade Estadual de Goiás, Campus Central, 75132-903, Anápolis, GO, Brazil.

[*]Corresponding author: Micael Rosa Parreira, Universidade Federal de Goiás, Campus Samambaia, Av. Esperança, s/n - Chácaras de Recreio Samambaia, 74690-900, Goiânia, GO, Brazil. Email: micael_rp@hotmail.com

Abstract

Species Distribution Models are based mainly on environmental (mostly climatic) and species distribution data to predict the potential distribution of species. This prediction is based on the model assumption that species are in equilibrium with the climate, i.e., species occur in all places with suitable environmental conditions. However, the climatic equilibrium relies on the species ability and time to disperse to suitable areas. In freshwater habitats, species dispersal is not restricted only by physical barriers but also by the directional movement of the hydrographic network, which can be considered through spatial predictors. Here, we aim to evaluate the effect of including asymmetrical and symmetrical spatial predictors in the potential geographic distribution of a freshwater fish in the Tocantins-Araguaia River basin, Brazil. For this, we built models with seven variable sets representing the climatic and spatial models, as well as their interactions. We found that the overall best models (higher evaluation and lower variation among modeling methods) are those built using AEM (asymmetrical dispersal [i.e., dispersal along the river flow path]), either solo or in combination with environmental variables (ENV). Moreover, the inclusion of asymmetrical dispersal variables, taking into account dispersal limitations of species, decreased the overprediction to climatically suitable but disconnected areas through rivers. Therefore, future SDM studies, especially those using species groups with directional dispersal, should consider the inclusion of asymmetrical spatial predictors to increase the model's accuracy and ecological reality.

Keywords: Asymmetric Eigenvector Maps; Species Distribution Models; directional dispersal; Principal Coordinates of Neighbour Matrices; spatial modeling.

1. Introduction

Species Distribution Models (SDMs) use environmental (mostly climatic – abiotic) and species distribution (mostly occurrences) data to predict climatically suitable areas for species survival (Araújo and Peterson, 2012; Peterson et al., 2011; Peterson and Soberón, 2012). SDMs have been widely used for predicting the species' potential distribution as a species conservation tool in poorly known areas (Guisan et al., 2013). Moreover, they have been used in different research areas, such as climate change (Anderson, 2013; Nabout et al., 2011), invasive species (Jiménez-Valverde et al., 2011), disease transmission (Peterson et al., 2005), delimitation of conservation areas (de Carvalho et al., 2017), the effect of habitat loss (Peterson et al., 2006), among others (see Peterson, 2006 for more applications).

The factors determining species distribution regarding their niche are represented in the BAM diagram (Biotic, Abiotic, and "Movement") (Soberón, 2007). However, correlative SDMs often use only species occurrence data and environmental factors (abiotic) (Peterson et al., 2011).

Therefore, including biotic and movement factors in SDMs are current knowledge frontiers to improve the predictions of species distributions, and some recent papers have used these new predictors in SDMs (see, for example, Barve et al., 2011; Cardador et al., 2014; da Cunha et al., 2018; Gherghel et al., 2018).

Movement constraints can be inserted as a component of the SDMs by including areas environmentally accessible or inaccessible to species, functioning as a proxy for dispersal limitation or migration rate (Miller and Holloway, 2015). In this sense, identifying relevant movement constraints and successfully incorporating them into SDMs are crucial to understanding the landscape-habitat connectivity and species dispersal (Perrin et al., 2020; Vasudev et al., 2015). Some studies have successfully included dispersal proxies as predictors in SDMs, generally through distance functions (e.g., Euclidean and Kernel distances) and fixed dispersal rate (Barbet-Massin et al., 2012b; Holloway et al., 2016; Monsimet et al., 2020), either *a priori* (i.e., as explanatory variables) or *a posteriori* (i.e., by overlapping accessible and suitable areas) (Mendes et al., 2020). Although such approaches can be suitable to describe the movement of many terrestrial species, they cannot account for the whole complexity of asymmetrical movements.

Asymmetrical movement is the occurrence of a preferential direction of migration, i.e., the probability of moving in one direction is not the same as moving in the opposite direction (Acevedo and Fletcher, 2017; Pringle et al., 2011). Asymmetrical dispersal has been reported in various ecological systems and for a variety of groups. For example, asymmetrical dispersal due to oceanic and stream currents for many species (Riginos et al., 2019), and wind patterns generating directional dispersal in fungal, orchid, and zooplankton species (Acevedo et al., 2015; Horváth et al., 2016; Rieux et al., 2014). Also, asymmetric dispersal can arise or intensify due to landscape fragmentation and climate change (Acevedo et al., 2020; Dalui et al., 2020; Pavlacky Jr et al., 2012). Specifically in freshwater environments, species dispersal is constrained not only by geographical distance but also by the hierarchical structure and flow direction of rivers (Domisch et al., 2015b). Consequently, in these systems, there is a predominant downstream direction of migration (Asymmetrical dispersal: Pringle et al., 2011; Altermatt et al., 2013). Thus, dispersal proxies that account for asymmetrical (directional) movements are more appropriate to characterize species movement in freshwater systems (Altermatt, 2013; Mozzaquattro et al., 2020). Although some recent studies have addressed the impacts of dispersal limitation in SDMs in freshwater environments (Bush and Hoskins, 2017; Perrin et al., 2020), few studies have included flow direction in their dispersal metrics (e.g., Ver Hoef et al., 2006).

Metacommunity studies have widely used spatial eigenfunction analysis as a surrogate of spatial dispersal (e.g., Griffith & Peres-Neto, 2006; Heino et al., 2015). Among those, we can highlight asymmetrical (AEM - Asymmetric eigenvector maps) (Blanchet et al., 2008) and symmetrical (e.g., PCNM - Principal coordinates of neighbor matrices) spatial predictors (Borcard and Legendre, 2002). These predictors can be used by inserting the directional effect of species dispersal in many macroecological approaches and at many scales (Blanchet et al., 2011). Moreover, recent studies have shown the importance of AEM filters as a surrogate for dispersal in freshwater metacommunity structuring, which, sometimes, contributes more than environmental processes (e.g., Dong et al., 2016; Mozzaquattro et al., 2020; Rocha et al., 2020). Besides the potential of such spatial predictors to account for the dispersal structure of species, to our knowledge, there is no study using them to evaluate dispersal constraints (either asymmetrical or symmetrical) in SDMs.

Therefore, in this study, we aim to evaluate the effect of including asymmetrical and symmetrical dispersal predictors in the potential geographic distribution of *Aspidoras* *eurycephalus*, a Neotropical fish endemic to the studied area. Considering the study objective, we selected one Neotropical basin (Tocantins-Araguaia River basin) and one fish species occurring in this basin. This basin is located in the central portion of Brazil and has a bioclimatic variability mainly along the latitudinal gradient. Moreover, the basin is divided into two major sub-basins (Tocantins and Araguaia) connected in the northernmost part of the basin, where some species have been recorded only in one sub-basin (e.g., our model species). Therefore, the latitudinal gradient of climatic variables and the longitudinal dispersal limitation (except for one connection in the north region) support this basin as a suitable model region to evaluate the influence of directional spatial predictors on SDMs. We hypothesize that models built with variables including the directional dispersal effect (AEM) through the rivers will produce more realistic and accurate models since freshwater fish species have directional dispersal routes through the hydrographic network.

2. Methods

2.1. Study area

The study area of the SDMs is the Tocantins-Araguaia River basin, covering the entire hydrographic network. This basin has two major rivers (Tocantins and Araguaia), forming two sub-basins that merge in the north region, close to the mouth (Figure 1). Therefore, a species occurring exclusively in one sub-basin needs to disperse a long way through the main river course in that sub-basin to occupy the other sub-basin. Thus, this basin is an interesting area for studies on dispersal limitation because it shows the species effort to disperse between the two basins (Tocantins and Araguaia). The entire hydrographic network was rasterized into grid-cells with 0.5° resolution (latitude and longitude), totaling 282 cells for the Tocantins-Araguaia basin (Figure 2A).

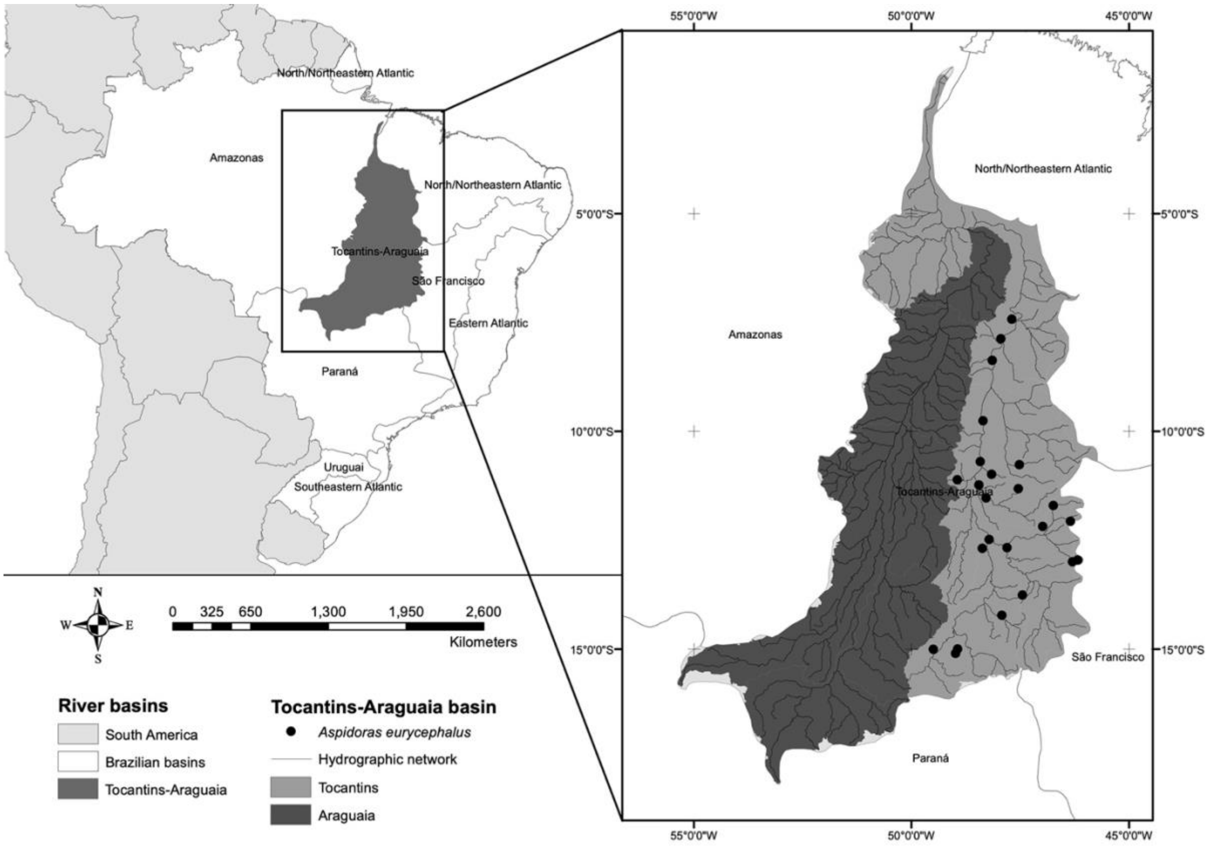


Figure 1 – Study area of the Tocantins-Araguaia River basin and the occurrences of *Aspidoras eurycephalus* (black dots).

2.2. Species occurrence data

We obtained the fish occurrence data from specific online databases: Species Link (specieslink.net/search), Gbif (gbif.org), and FishBase (fishbase.org; Froese & Pauly, 2019). All records are from fish collections and museums, such as the UNT (Fish collection of the Federal University of Tocantins), and the collection time frame varies from 2002 to 2012. We used *Aspidoras eurycephalus* Nijssen & Isbrücker, 1976 (Siluriformes: Callichthyidae) as our model species to evaluate the contribution of spatial predictors in the potential distribution. This small catfish is endemic to the upper Tocantins River sub-basins, with a maximum size of 3 cm and wide distribution in this river (Reis et al., 2003). We removed the duplicated occurrence points (more than 1 point in the same grid-cell), resulting in 23 unique occurrence points of this species (Figure 1). Moreover, because presence-absence methods (e.g., GLM) require absence records, we generated 56 pseudo-absences (~20% of study area cells) randomly distributed in all extent area (1 per grid cell). We avoided grid cells containing occurrence records (Barbet-Massin et al., 2012a).

2.3. Environmental data

The environmental variables (ENV) used were the 19 bioclimatic variables from the online database Worldclim (http://worldclim.com; (Hijmans et al., 2005). We rescaled the variables to a resolution of 0.5º (~ 55 km) using the function *aggregate* from the *raster* package (Hijmans, 2019). We then used a factorial analysis (FA) using varimax rotation to remove collinearity among variables. For this, we used the functions *fa* and *fa.paralell* of *psych* package (Revelle, 2019) of the R software (R Core Team, 2019) to determine the adequate number of non-orthogonal axes. We determined four axes (i.e., the number of factors with eigenvalues higher than simulated eigenvalues) through a scree plot of actual data and simulated data eigenvalues. Then, we selected the variables with the highest loading (- or +) on each axis. The environmental variables selected were: BIO1 = Annual Mean Temperature, BIO2 = Mean Diurnal Temperature Range, BIO13 = Precipitation of Wettest Month, and BIO15 = Precipitation Seasonality (See variables in Supplementary Material). It is important to note that, whenever possible, predictors should be chosen considering the species' biology. However, statistical selection (as used here) has been widely used in the absence of such information. In this sense, the choice of environmental variables based on species biology can increase the importance of environmental variables.

2.4. Dispersal-related predictors

2.4.1. Asymmetrical binary matrix

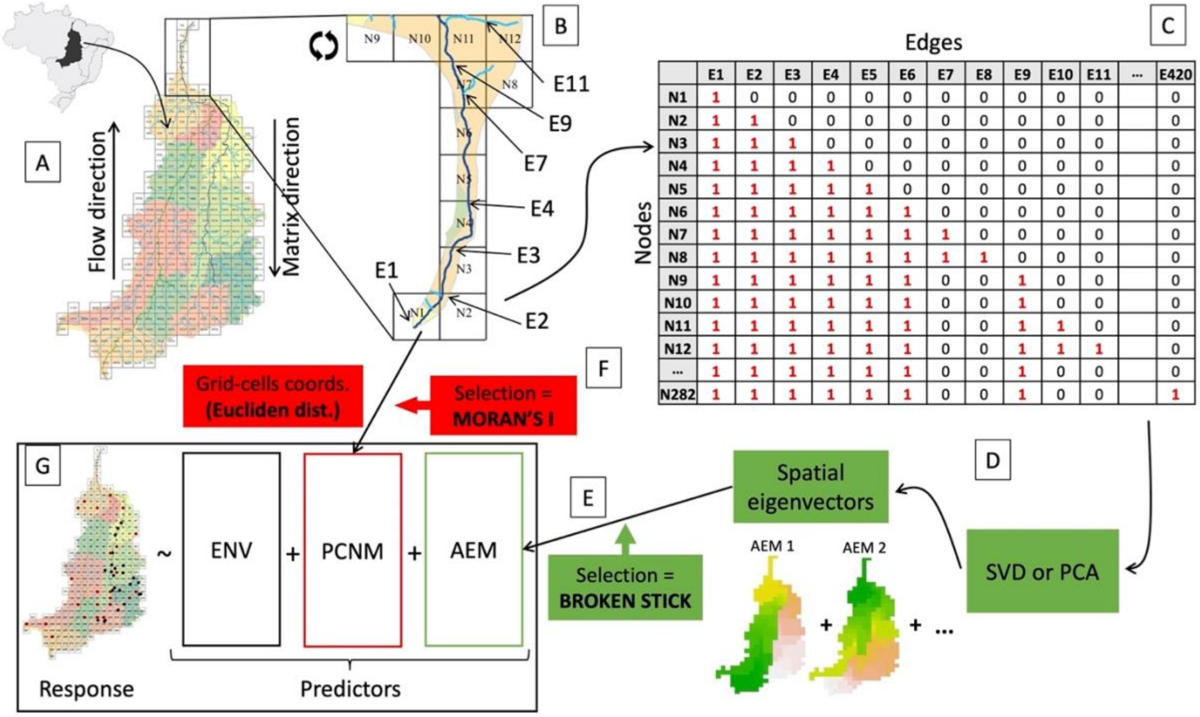
The binary matrix was created through the grid-cells of the Tocantins-Araguaia basin with a resolution of 0.5º (latitude e longitude) (Figure 2A). This matrix consists of river connections (edges) in the columns and numbers of each cell (nodes) in the rows (Figures 2B and C). We started counting the grid-cells containing rivers upstream, with the first grid-cell corresponding to the mouth of the basin. The first connection (E1) received value 1 for all nodes of the basin. The subsequent connections (E2, E3, ...) also received a value of 1 for all nodes, except for the first node with 1 in this connection. We repeated this step until we reached a tributary (lower order) river. In this case, we started a new edge (e.g., E7 in our study), assigning value 1 for all nodes in this connection and subtracting the first node in the following edges of this tributary river. Once we finished the last edge in the tributary, we returned to the main connection (the entire basin), attributing value 1 to the remaining nodes until we reached another tributary. We repeated this process until we reached the last edge, which usually contains the source furthest from the mouth of the basin (Figure 2C). We ended up with an asymmetrical (directional) matrix with 282 grid-cells (nodes) and 420 connections (edges). This methodology is based on Blanchet et al. (2008).

2.4.2. AEM and PCNM calculation and axes selection

We used asymmetrical and symmetrical spatial filters as dispersal predictors in the construction of SDMs. The spatial filters used were PCNM (*Principal coordinates of neighbour matrices*; Borcard and Legendre, 2002) and AEM (*Asymmetric eigenvector maps*; Blanchet et al., 2008). Each filter (orthogonal axes) represents the geographic space, where the first axes represent a large-scale variation, and the last axes represent a fine-scale variation (Borcard and Legendre, 2002). Specifically, in our study, AEM filters represent the flow path of rivers throughout the basin, where the river connections and directionality are translated from a binary connectivity matrix. PCNM filters, on the other hand, represent the non-directional (symmetrical) effect of dispersal into the species distribution, such as a distribution by air. The spatial filters can be incorporated into approaches of multiple regressions (e.g., niche models), inserting spatial autocorrelation into the models (Diniz-Filho and Bini, 2005). We removed the spatial filters (AEM and PCNM) with low importance considering the comparison with null models (broken-stick), their low spatial structure, and their high correlation with bioclimatic variables (redundant spatial filter).

The AEM calculation (function *aem* of *adespatial* package [Dray et al. 2019]) is based on the SVD (Singular Value Decomposition) analysis using the asymmetrical binary matrix as the input matrix. Similar results can be obtained through a PCA (Principal Component Analysis) using the same binary matrix or a PCoA (Principal Coordinates Analysis) using a Euclidean distance matrix from this binary matrix (Figure 2D) (Blanchet et al., 2008). The AEM calculation generated eigenvalues and eigenvectors (axes) that correspond to the dispersal predictors. Since here we are using standardized river connections (edges), which are grid cells of the same size throughout the basin, it was not necessary to include weights in the AEM calculation, which is different from using river segments as edges that may have different lengths. We selected the axes sufficient to explain the total variance (i.e., sum of the eigenvalues) using the broken stick method for AEM based on SVD. We found no correlation between axes with any bioclimatic variable. Therefore, we retained the first seven axes through the broken stick, which were used posteriorly as predictors to construct the niche models (Jackson, 1993) (Figure 2E and Figure A.2).

The construction of PCNMs (function *pcnm* of package *vegan* [Oksanen et al., 2017]) was performed using the centroid coordinates (latitude and longitude) of the 282 grid cells. The analysis resulted in 281 eigenvectors (axes) corresponding to the dispersal predictors. We used the Moran's I test (Cliff and Ord, 1981) to select the axes sufficient to explain the total variation. To calculate the Moran's I, we used the function *moran.randtest* of package *adespatial* (Dray et al., 2019), using the eigenvectors (axes) generated by the PCNM analysis as input. This function computes the Moran's I for each axis generating significant ($p<0.05$) correlation values ranging from 0 to 1 for each axis, where the axes selected had values above 0.7. Finally, we retained the first 20 axes through the Moran's I approach. However, we removed the first axis (PCNM1) due to its collinearity with the climatic variables. Therefore, 19 axes were posteriorly used as symmetrical dispersal predictors (PCNM) in the niche model construction (Figure 2F and Figure A.3).



	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	...	E420
N1	1	0	0	0	0	0	0	0	0	0	0		0
N2	1	1	0	0	0	0	0	0	0	0	0		0
N3	1	1	1	0	0	0	0	0	0	0	0		0
N4	1	1	1	1	0	0	0	0	0	0	0		0
N5	1	1	1	1	1	0	0	0	0	0	0		0
N6	1	1	1	1	1	1	0	0	0	0	0		0
N7	1	1	1	1	1	1	1	0	0	0	0		0
N8	1	1	1	1	1	1	1	1	0	0	0		0
N9	1	1	1	1	1	1	0	0	1	0	0		0
N10	1	1	1	1	1	1	0	0	1	0	0		0
N11	1	1	1	1	1	1	0	0	1	1	0		0
N12	1	1	1	1	1	1	0	0	1	1	1		0
...	1	1	1	1	1	1	0	0	1	0	0		0
N282	1	1	1	1	1	1	0	0	1	0	0		1

Figure 2 – Conceptual diagram of calculation of asymmetrical binary matrix and asymmetrical and symmetrical spatial filters. A) Gridded Tocantins-Araguaia map subdivided into sub-basins; B) Determination of nodes and edges in a stretch of the gridded basin; C) Directionality matrix using the nodes and edges in a stretch of the gridded basin; D) Calculation of AEM using SVD or PCA,

generating the spatial eigenvectors (AEM variables) used for modeling; E) Selection method (Broken stick) to determine the AEM variables to be used for modeling; F) Selection method (Moran's I) to determine the PCNM variables to be used for modeling; G) Conceptual model (Response [Presences and Pseudo-absences] ~ Predictors [ENV+PCNM+AEM]) used to generate the SDMs.

2.5. Species Distribution Models

We generated 50 models for each modeling method using species occurrences and pseudo-absence data, non-collinear environmental variables (ENV), and symmetrical and asymmetrical dispersal predictors (AEM and PCNM) (Figure 2G). Thus, a total of seven sets of SDMs were built: a) only ENV; b) only AEM; c) only PCNM; d) ENV + AEM; e) ENV + PCNM; f) PCNM + AEM; g) All predictors (FULL). Moreover, we used a total of six methods (algorithms) for each set of variables to assess the variability of methods in the predictions. Therefore, we generated in total 2100 models (50 x 7 x 6). The algorithms used were Bioclim (Nix, 1986), Domain (Carpenter et al., 1993), Support Vector Machines (SVM; Schölkopf et al., 2001), Generalized Linear Models (GLM; Nelder & Wedderburn, 1972), Maximum Entropy (MaxEnt; Phillips et al., 2006), and Random Forest (Breiman, 2001). The algorithms were chosen because they consider different statistical methods (climatic envelopes, environmental distances, machine learning, regressions) (Rangel and Loyola, 2012).

Models were built using 75% presence and absence points for training (model's construction) and 25% for testing the models' performance (Guisan and Zimmermann, 2000), randomly chosen. Thus, we have the same set of pseudo-absences for all models, but randomly selecting subsets for each run. In this sense, we can compare the different sets of variables among models with the same initial datasets. All models generated were evaluated using the Area Under the ‘receiver operating characteristic’ Curve (AUC; Swets, 1988), which is an evaluation metric threshold-independent (the limit for determination of presences/absences) that compare predict with observed values. The final map of potential distribution is a consensus map (ensemble) built using the mean suitability values of all models with AUC > 0.7 weighted by the AUC values. The models were generated and evaluated using the dismo package (Hijmans et al., 2016) available in the R software (R Core Team, 2019).

2.6. Data analysis

The eigenvectors (axes) generated by the asymmetrical (AEM) and symmetrical (PCNM) spatial filters, as well as the environmental variables, were used as predictors in the SDMs. These models were evaluated and compared using different statistics for models' performance (AUC, TSS [Sensitivity + Specificity – 1], Sensitivity [True Positive Rate], and Specificity [True Negative Rate]) to assess the effect of including spatial predictors in the niche models. For this, we compared the mean values of the evaluation metrics between the models. For this, we used the mean metric values of the 50 models built using the six modeling methods for each of the seven sets of variables. Therefore, we ended up with a mean value for the 300 (50 models x 6 algorithms) models for each variable set, representing the models with the highest performance among the different sets of variables. Metric values were compared using Venn diagrams, where it is possible to visualize the values in the individual (ENV, AEM, and PCNM) and interacting sets (ENV+AEM, ENV+PCNM, AEM+PCNM, and FULL) through connected circles.

Besides comparing the evaluation metrics for each set of variables, we tested the variation of SDM evaluation metrics (AUC, TSS, Sensitivity, and Specificity) among the variables and methods through an interaction plot between the set of variables and modeling methods used. Since each model built used one set of variables and one modeling method, we arranged the output metric values corresponding to each method and variable. For this, we obtained the mean metric values of all 50 models for each interaction between the set of variables and modeling methods (6 x 7 = 42 points plotted). In this sense, we generated two interaction plots between variable types and modeling methods using the SDM evaluation metrics of these interactions. All analyses in this study were performed in the R software version 3.5.1 (R Core Team, 2019).

3. Results

The consensus maps (ensembles) generated using the sets of variables individually (ENV, AEM, and PCNM), and in addition to other variables (ENV+AEM, ENV+PCNM, AEM+PCNM, and FULL), showed similar predicted distribution outputs. The models predicted suitable areas close to the known species core region. However, it is possible to notice that AEM-based maps limited the species potential distribution for the Tocantins River sub-basin where the species currently occurs (Figure 3).

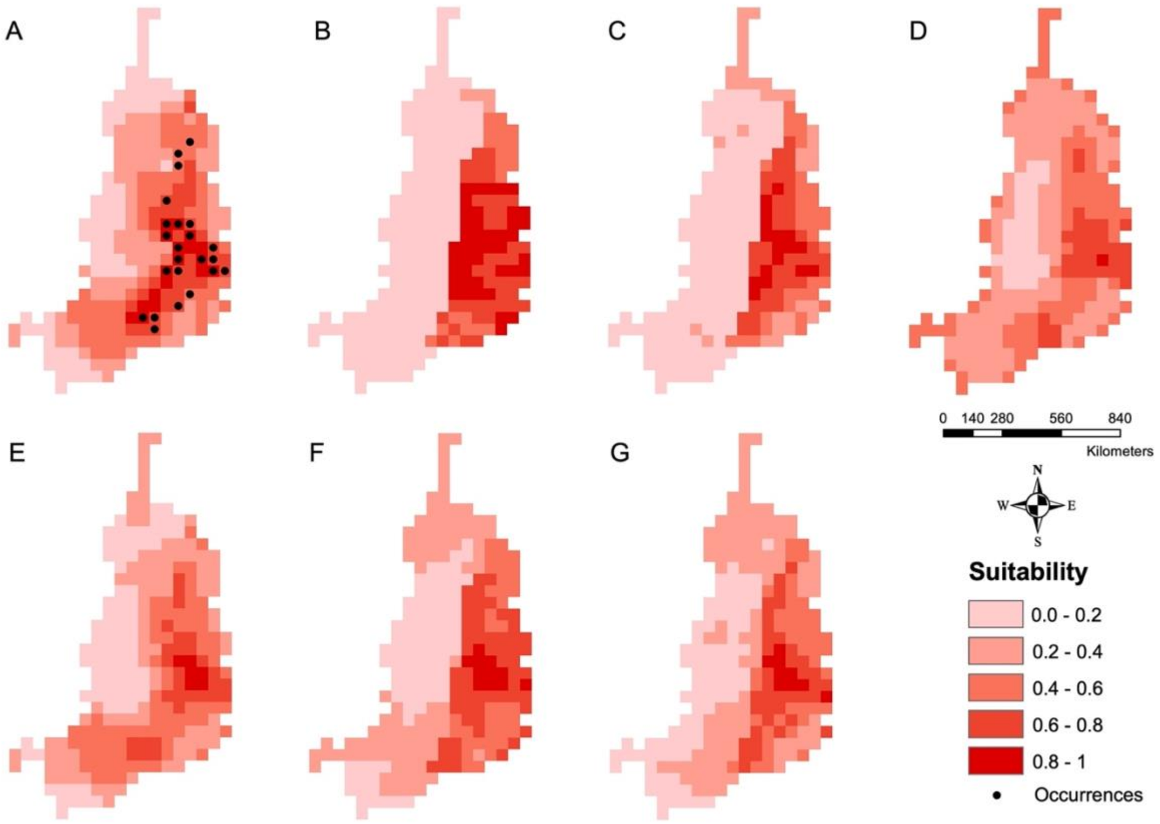


Figure 3 – Consensus maps (ensembles) generated for each set of variables (A=ENV, B=AEM, C=ENV+AEM, D=PCNM, E=ENV+PCNM, F=AEM+PCNM, and G=FULL).

The models generated using the combination of climate and asymmetrical dispersal variables (AEM+ENV) had the best performance for all evaluation metrics, except for sensitivity. Besides, models using only asymmetrical spatial predictors (AEM) also had a high evaluation performance, showing the highest sensitivity of all models. On the other hand, models using only PCNM and together with other variables (ENV+PCNM and AEM+PCNM) generated the models with the lowest performance for both metrics (Figure 4).

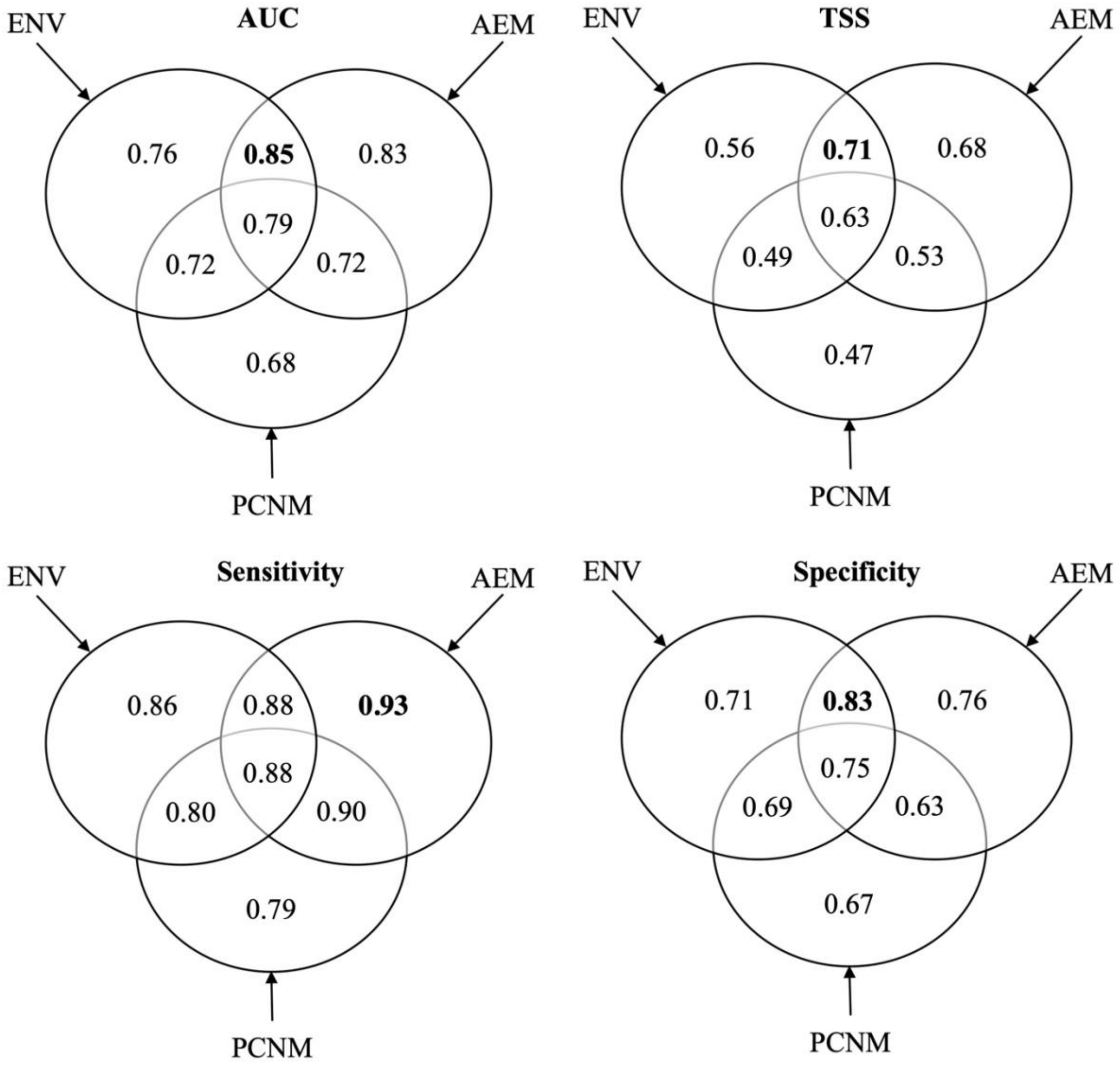


Figure 4 – Venn diagram of evaluation metrics (AUC, TSS, Sensitivity, and Specificity) results for all sets of niche models generated. Higher values (highest in bold) indicate the best adjusted models.

The evaluation metrics of models' performance (AUC, TSS, Sensitivity, and Specificity) showed high variability among modeling methods and variable types (environmental, asymmetrical, and symmetrical spatial variables). In general, AEM-based models had the highest evaluation metrics and the lowest performance variation, especially for models with AEM only and AEM+ENV. On the contrary, models built using ENV and PCNM (solo or together with other variables) generated high variability among methods, especially for PCNM-based models, which had the worst performances. This reduction in performance is accentuated when combined with Bioclim and GLM. The remaining model combinations had consistent performances (Figure 5).

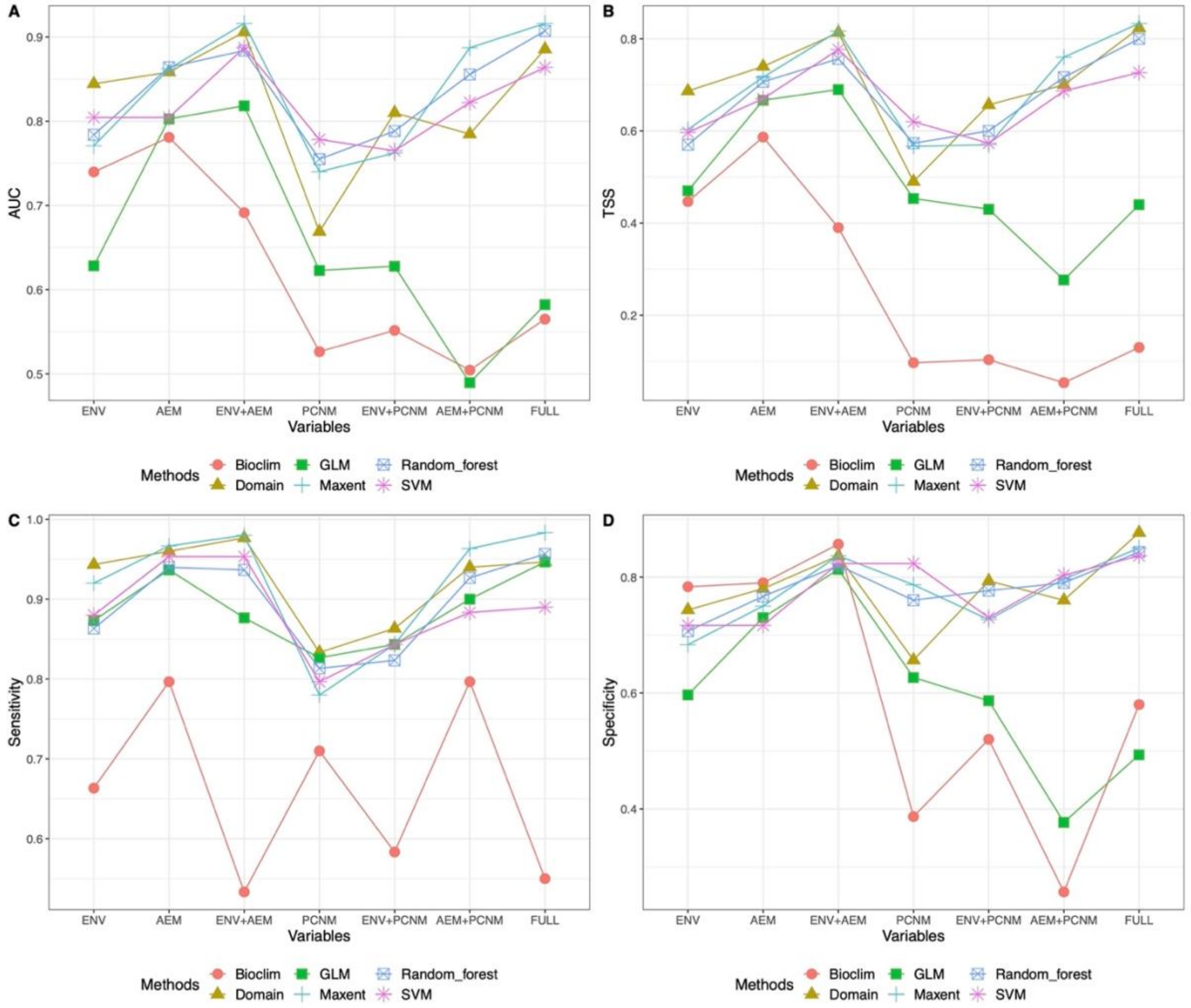


Figure 5 – Interaction plot between methods and variables used in the SDMs, considering the evaluation metrics of model’s performance (AUC, TSS, Sensitivity, and Specificity).

4. Discussion

Our results showed a significant influence of asymmetrical spatial predictors in freshwater SDM performance. Moreover, the inclusion of asymmetrical predictors (AEM) accounted for dispersal limitations, informing the models about areas with difficult access to the species, unlike models based only on climatic aspects. In this sense, AEM-based models limited overprediction in areas climatically similar but disconnected through rivers, such as in the upper Tocantins-Araguaia River basin.

The predictors considered here generated models with similar spatial predictions. However, models based on AEM (solo or together with other predictors) showed the highest restriction to the Tocantins River basin, the known species distribution area. This restriction demonstrates the effect of the dispersal path the species has to travel to reach other regions. Therefore, the representation of directional dispersal along the river course through asymmetrical spatial predictors (AEM) is more suitable for including dispersal routes of rivers in niche models. Besides, different methods have been used to incorporate the dispersal effect in SDMs (Engler et al., 2012). However, dispersal limitation in those studies is often based on symmetrical dispersal or migration rates for areas surrounding the current species distribution (Miller and Holloway, 2015).

The models built using the set of climatic and asymmetrical spatial predictors (ENV+AEM) had the best evaluations, similar to models built with AEM only, which had the highest sensitivity of all models (i.e., more accurately predicted occurrences). The evaluation is directly linked to the model's performance; therefore, the best-evaluated models represent more reliable predictions of the species potential distribution (Allouche et al., 2006). Conversely, models using PCNMs showed the worst evaluations, either solo or together with other predictors, for all tested combinations of variables. Therefore, models built using only the symmetrical space (PCNM) are not reliable to predict the potential distribution of this freshwater fish species. However, hitting or missing presence and absence areas of species may not be enough to assess the quality of the models, as they can be over- or underpredicting by the lack of extrinsic characteristics of the species (e.g., dispersal), generating overpredicted and unrealistic models (Uribe-Rivera et al., 2017). Therefore, besides increasing the model's performance, asymmetrical spatial predictors in SDMs insert dispersal routes (i.e., rivers and streams) that species must travel to be present in other locations, increasing the ecological reality of SDMs.

The models' performance also varied depending on the combination of predictor variables (i.e., spatial and climatic) and modeling methods used to build the SDMs. Some modeling methods showed more variation on their evaluations than others, especially those combined with symmetrical predictors (PCNM). Those variations (uncertainties) in ensemble models among modelling methods are widely known and studied (Diniz-Filho et al., 2009). They can also change depending on the combinations of variables used (Parreira et al., 2019). Among all combinations, models built using only asymmetrical dispersal (AEM) or with other predictors (i.e., AEM-based models), had the lowest variation among methods, showing overall satisfactory performances. Models with spatial restrictions are expected to generate less variation among modeling methods as these "barriers" or dispersal paths may limit the species to occur in other climatically suitable locations, hindering access to inaccessible or less accessible areas (Uribe-Rivera et al., 2017). For example, the upper basin elevation between Araguaia and Tocantins tributaries limits the species dispersal between these two sub-basins, except for their only downstream connection. Therefore, this dispersal restriction in SDMs limits possible overpredictions to climatically similar but disconnected areas (Mendes et al., 2020).

Spatial predictors in freshwater environments can be a valuable tool for building models with lower overprediction of suitable areas for species. In these models, the dispersal routes along the river course, through spatial filters, are included to represent the species movement throughout the basin hydrographic network (Blanchet et al., 2008), improving the accuracy in SDMs predictions for these environments. Moreover, we emphasize that the models built using AEM (solo and together with other predictors) showed higher accuracy than the traditional climatic-only models. Nevertheless, new approaches to automatically generate the asymmetrical binary matrix for the AEM analysis are necessary since manually designing this matrix for large extensions, such as biogeographic regions (e.g., Neotropics) or the entire globe, is yet very demanding. Furthermore, we expect future SDM studies to assess the insertion of this spatial direction in other basins, at different spatial scales, and with a larger species pool, which could allow assessing the effect of different dispersal abilities. For example, asymmetrical predictors may be less relevant for species with higher dispersal abilities or frequent upstream movement, as AEM could be less effective in capturing the dispersal behavior of such species. Besides, niche models using different variable sets, or even using other types of asymmetrical dispersal, such as dispersal through the wind (Horváth et al., 2016), could help the assessment of dispersal limitation in terrestrial SDMs."

In conclusion, our results show an emerging potential for using asymmetrical dispersal filters (AEM) as variables for the construction of aquatic SDMs, solo or with climate-based variables. Besides, other approaches could be developed following this methodology of inserting asymmetrical (directional) dispersal in freshwater SDMs. For example, 1) more accurately predicting the dispersal pathways of freshwater invasive species in new habitats, or 2) evaluating the effect of dispersal interruption, such as current or planned hydropower plants (HPP), on the dispersal of fish species by inserting the disconnection into SDMs built using AEM filters through disconnections in the directional matrix in the HPP areas. Therefore, this approach of using asymmetrical dispersal in freshwater SDMs may contribute to new insights regarding the potential distribution of freshwater species by considering their dispersal routes in the models.

5. Acknowledgments

We thank Fabrício Barreto Teresa for the help in obtaining the species distribution data. MRP and GT received a scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – finance code 001) and PNPD/CAPES, respectively. Our work on Species Distribution Models has been continuously supported by different grants: CNPq, Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG/process 201710267000519), and National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation (MCTI/CNPq/FAPEG/465610/2014-5), and Brazilian Network on Global Climate Change Research (Rede CLIMA). JCN was supported by a CNPq productivity fellowship.

6. Supplementary Material

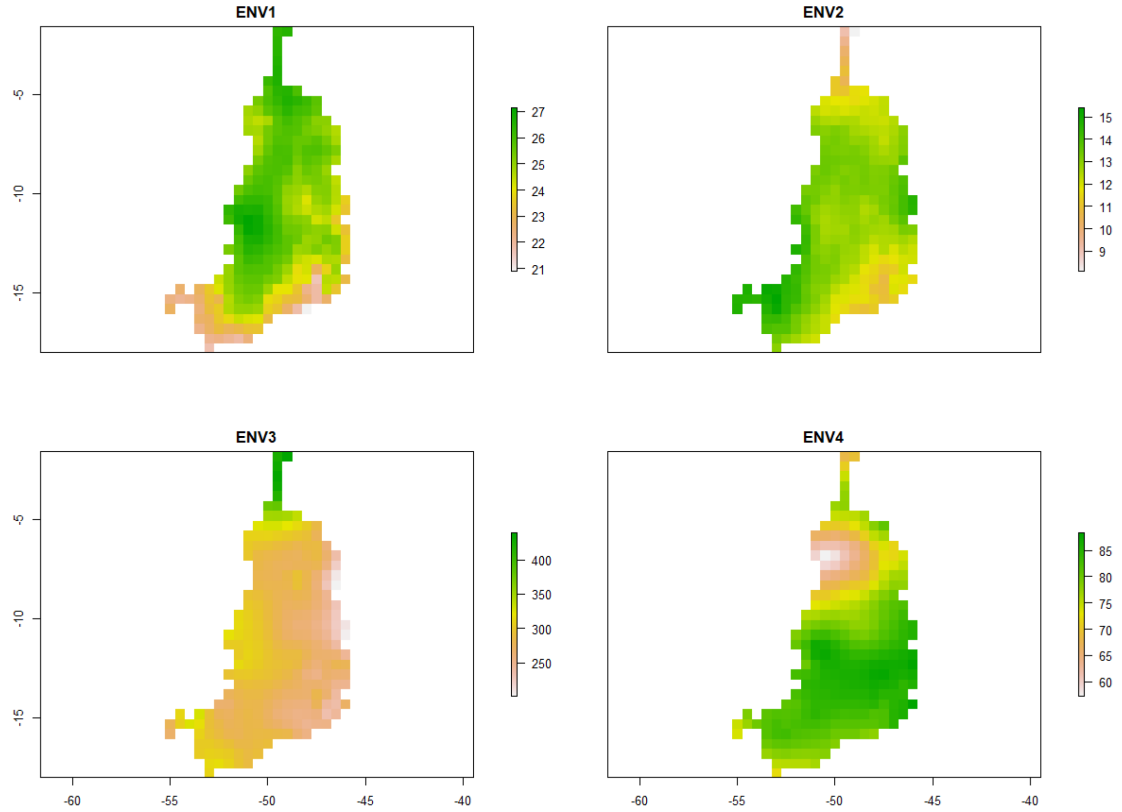


Figure S1 – Environmental (bioclimatic) variables selected for ENM construction.

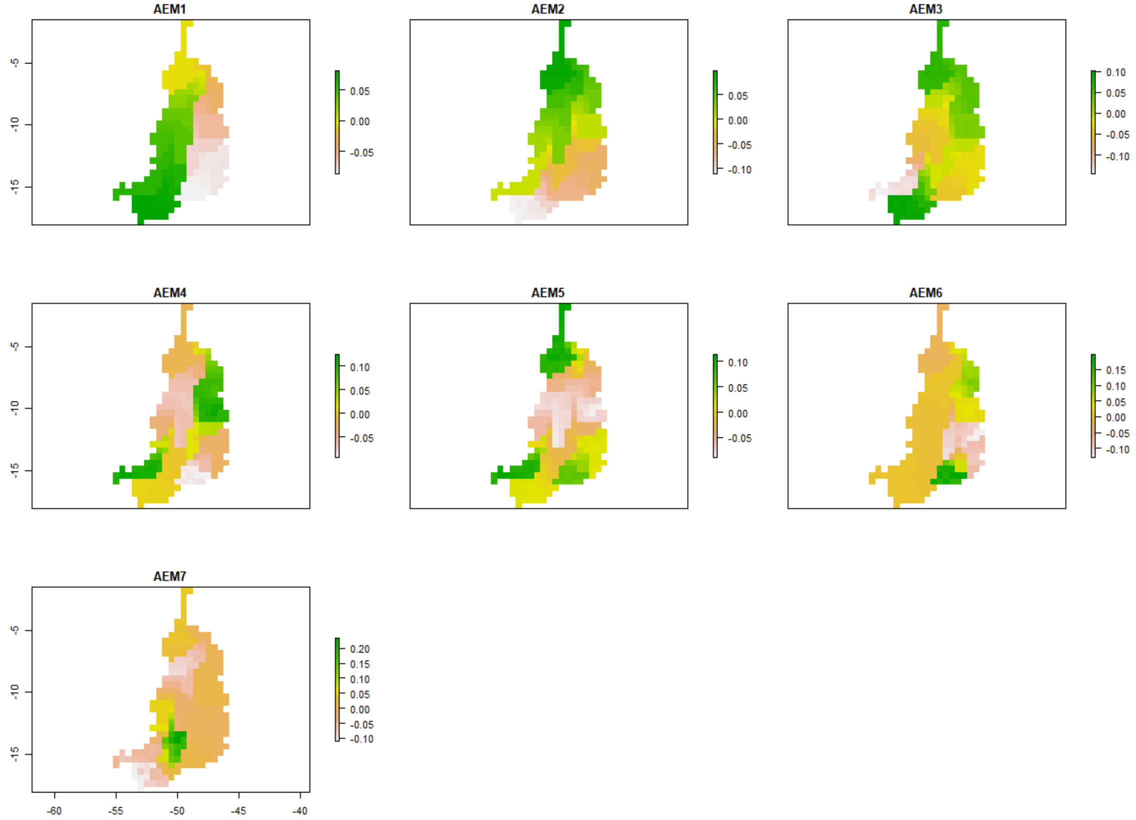


Figure S2 – AEM filters selected for ENM construction.

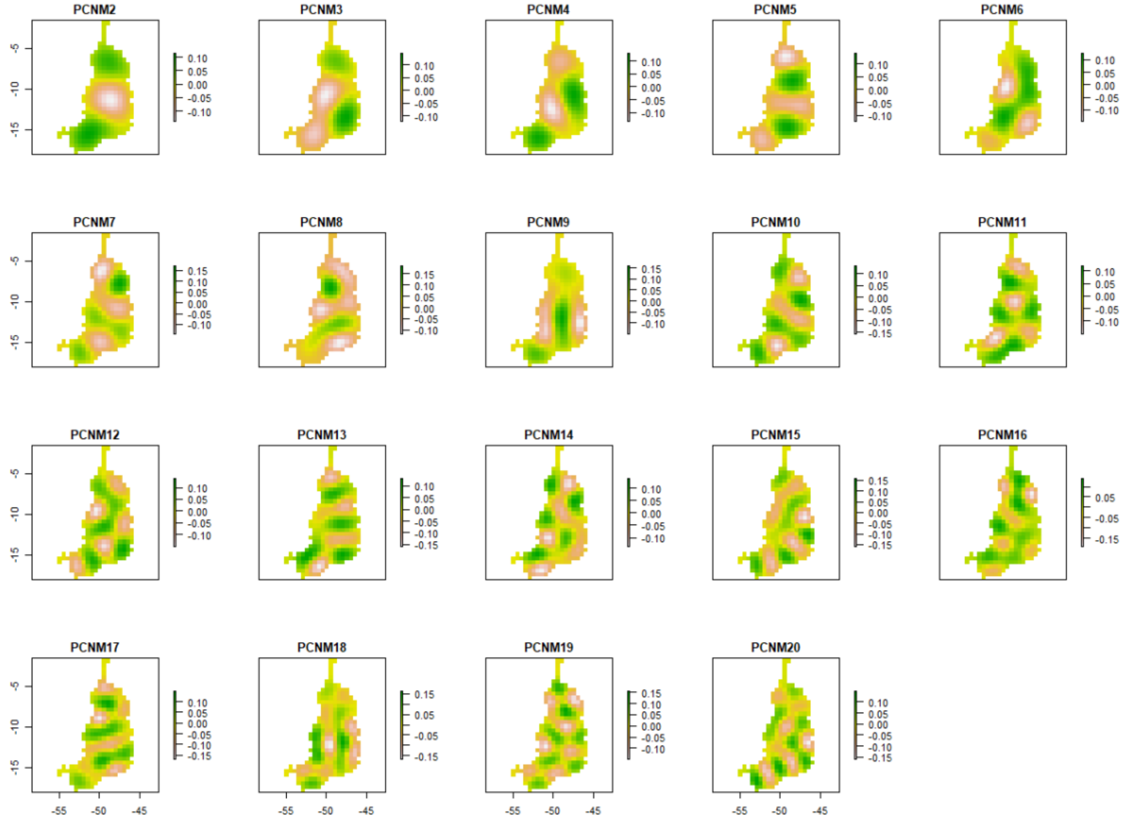


Figure S3 – PCNM filters selected for ENM construction.

7. References

Barbet-Massin, M., Thuiller, W., Jiguet, F., 2012b. The fate of European breeding birds under climate, land-use and dispersal scenarios. Glob. Chang. Biol. 18, 881–890. https://doi.org/10.1111/j.1365-2486.2011.02552.x

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. Ecol. Modell. 222, 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Blanchet, F.G., Legendre, P., Borcard, D., 2008. Modelling directional spatial processes in ecological data. Ecol. Modell. 215, 325–336.

https://doi.org/10.1016/j.ecolmodel.2008.04.001

Blanchet, F.G., Legendre, P., Maranger, R., Monti, D., Pepin, P., 2011. Modelling the effect of directional spatial ecological processes at different scales 357–368. https://doi.org/10.1007/s00442-010-1867-y

Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Modell. 153, 51–68. https://doi.org/10.1016/S0304-3800(01)00501-4

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Bush, A., Hoskins, A.J., 2017. Does dispersal capacity matter for freshwater biodiversity under climate change? Freshw. Biol. 62, 382–396. https://doi.org/10.1111/fwb.12874

Cardador, L., Sardà-Palomera, F., Carrete, M., Mañosa, S., 2014. Incorporating spatial constraints in different periods of the annual cycle improves species distribution model performance for a highly mobile bird species. Divers. Distrib. 20, 515–528. https://doi.org/10.1111/ddi.12156

Acevedo, M.A., Beaudrot, L., Meléndez-Ackerman, E.J., Tremblay, R.L., 2020. Local extinction risk under climate change in a neotropical asymmetrically dispersed epiphyte. J. Ecol. 108, 1553–1564. https://doi.org/https://doi.org/10.1111/1365-2745.13361

Acevedo, M.A., Fletcher, R.J., 2017. The proximate causes of asymmetric movement across heterogeneous landscapes. Landsc. Ecol. 32, 1285–1297. https://doi.org/10.1007/s10980-017-0522-y

Acevedo, M.A., Fletcher, R.J., Tremblay, R.L., Meléndez-Ackerman, E.J., 2015. Spatial asymmetries in connectivity influence colonization−extinction dynamics. Oecologia 179, 415–424. https://doi.org/10.1007/s00442-015-3361-z

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Altermatt, F., 2013. Diversity in riverine metacommunities: a network perspective. Aquat. Ecol. 47, 365–377. https://doi.org/10.1007/s10452-013-9450-3

Anderson, R.P., 2013. A framework for using niche models to estimate impacts of climate change on species distributions. Ann. N. Y. Acad. Sci. 1297, 8–28. https://doi.org/10.1111/nyas.12264

Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93, 1527–1539. https://doi.org/10.1890/11-1930.1

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012a. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3, 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Carpenter, G., Gillison, A.N., Winter, J., 1993. Domain - a Flexible Modeling Procedure for Mapping Potential Distributions of Plants and Animals. Biodivers. Conserv. 2, 667–680. https://doi.org/10.1007/BF00051966

Cliff, A.D., Ord, J. k., 1981. Spatial processes - models and applications. Pion, London, UK.

da Cunha, H.F., Ferreira, É.D., Tessarolo, G., Nabout, J.C., 2018. Host plant distributions and climate interact to affect the predicted geographic distribution of a Neotropical termite. Biotropica 50, 625–632. https://doi.org/10.1111/btp.12555

Dalui, S., Khatri, H., Singh, S.K., Basu, S., Ghosh, A., Mukherjee, T., Sharma, L.K., Singh, R., Chandra, K., Thakur, M., 2020. Fine-scale landscape genetics unveiling contemporary asymmetric movement of red panda (Ailurus fulgens) in Kangchenjunga landscape, India. Sci. Rep. 10, 15446. https://doi.org/10.1038/s41598-020-72427-3

de Carvalho, D.L., Sousa-Neves, T., Cerqueira, P.V., Gonsioroski, G., Silva, S.M., Silva, D.P., Santos, M.P.D., 2017. Delimiting priority areas for the conservation of endemic and threatened Neotropical birds using a niche-based gap analysis. PLoS One 12. https://doi.org/10.1371/journal.pone.0171838

Diniz-Filho, J.A.F., Bini, L.M., 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. Glob. Ecol. Biogeogr. 14, 177–185. https://doi.org/10.1111/j.1466-822X.2005.00147.x

Diniz-Filho, J.A.F., Bini, L.M., Rangel, T.F., Loyola, R.D., Hof, C., Nogués-Bravo, D., Araújo, M.B., 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. Ecography (Cop.). 32, 897–906. https://doi.org/10.1111/j.1600-0587.2009.06196.x

Domisch, S., Amatulli, G., Jetz, W., 2015. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Sci. Data 2, 150073. https://doi.org/10.1038/sdata.2015.73

Dong, X., Li, B., He, F., Gu, Y., Sun, M., Zhang, H., Tan, L., 2016. Flow directionality , mountain barriers and functional traits determine diatom metacommunity structuring of high mountain streams. Nat. Publ. Gr. 1–11. https://doi.org/10.1038/srep24711

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2019. adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-14. https://CRAN.R-project.org/package=adespatial

Engler, R., Hordijk, W., Guisan, A., 2012. The MIGCLIM R package - seamless integration of dispersal constraints into projections of species distribution models. Ecography (Cop.). 35, 872–878. https://doi.org/10.1111/j.1600-0587.2012.07608.x

Froese, R., and Pauly D., 2019. FishBase. World Wide Web electronic publication. https://www.fishbase.org.

GBIF.org, 2019. GBIF Occurrence Download. https://doi.org/10.15468/dl.rzmupe

Gherghel, I., Brischoux, F., Papeş, M., 2018. Using biotic interactions in broad-scale estimates of species' distributions. J. Biogeogr. 1–10. https://doi.org/10.1111/jbi.13361

Griffith, D.A., Peres-Neto, P.R., 2006. Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. Ecology 87, 2603–2613. https://doi.org/10.1890/0012-9658(2006)87[2603:SMIETF]2.0.CO;2

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16, 1424–1435. https://doi.org/10.1111/ele.12189

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Heino, J., Melo, A.S., Siqueira, T., Soininen, J., Valanko, S., Bini, L.M., 2015. Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. Freshw. Biol. 60, 845–869. https://doi.org/https://doi.org/10.1111/fwb.12533

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

https://doi.org/10.1002/joc.1276

Hijmans, R.J., Phillips, S.J., Leathwick, J.R., Hortal, J., 2016. dismo: Species Distribution Modeling. R package version 1.1-1. https://CRAN.R-project.org/package=dismo

Hijmans, R.J., 2019. raster: Geographic Data Analysis and Modeling. R package version 3.4-5. https://CRAN.R-project.org/package=raster

Holloway, P., Miller, J.A., eGillings, S., 2016. Incorporating movement in species distribution models: how do simulations of dispersal affect the accuracy and uncertainty of projections? Int. J. Geogr. Inf. Sci. 30, 1–25. https://doi.org/10.1080/13658816.2016.1158823

Horváth, Z., Vad, C.F., Ptacnik, R., 2016. Wind dispersal results in a gradient of dispersal limitation and environmental match among discrete aquatic habitats. Ecography (Cop.). 39, 726–732. https://doi.org/https://doi.org/10.1111/ecog.01685

Jackson, D.A., 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology 74, 2204–2214. https://doi.org/10.2307/1939574

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P., Lobo, J.M., 2011. Use of niche models in invasive species risk assessments. Biol. Invasions 13, 2785–2797. https://doi.org/10.1007/s10530-011-9963-4

Landeiro, V.L., Magnusson, W.E., Melo, A.S., Espírito-Santo, H.M. V., Bini, L.M., 2011. Spatial eigenfunction analyses in stream networks: do watercourse and overland distances produce different results? Freshw. Biol. 56, 1184–1192. https://doi.org/10.1111/j.1365-2427.2010.02563.x

Mendes, P., Velazco, S.J.E., Andrade, A.F.A. de, De Marco, P., 2020. Dealing with overprediction in species distribution models: How adding distance constraints can improve model accuracy. Ecol. Modell. 431, 109180. https://doi.org/10.1016/j.ecolmodel.2020.109180

Miller, J.A., Holloway, P., 2015. Incorporating movement in species distribution models. Prog. Phys. Geogr. 39, 837–849. https://doi.org/10.1177/0309133315580890

Monsimet, J., Devineau, O., Pétillon, J., Lafage, D., 2020. Explicit integration of dispersal-related metrics improves predictions of SDM in predatory arthropods. Sci. Rep. 10, 16668. https://doi.org/10.1038/s41598-020-73262-2

Mozzaquattro, L.B., Dala-Corte, R.B., Becker, F.G., Melo, A.S., 2020. Effects of spatial distance, physical barriers, and habitat on a stream fish metacommunity. Hydrobiologia 847, 3039–3054. https://doi.org/10.1007/s10750-020-04309-8

Nabout, J.C., Oliveira, G., Magalhães, M.R., Carina, T.L., Almeida, F.A.S. de, 2011. Global Climate Change and the Production of Pequi Fruits (Caryocar brasiliense) in the Brazilian Cerrado. Nat. Conserv. 9, 55–60. https://doi.org/10.4322/natcon.2011.006

Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized Linear Models. J. R. Stat. Soc. Ser. A 135, 370–384. https://doi.org/10.2307/2344614

Nix, H.A., 1986. A biogeographic analysis of Australian elapid snakes, in: Atlas of Elapid Snakes of AustraliaAtlas of Elapid Snakes of Australia. Australian Government Publishing Service, Canberra, pp. 4–15.

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., 2017. Vegan: community ecology package. R package version 2.4-3. https://CRAN.R-project.org/package=vegan

Parreira, M.R., Nabout, J.C., Tessarolo, G., de Souza Lima-Ribeiro, M., Teresa, F.B., 2019. Disentangling uncertainties from niche modeling in freshwater ecosystems. Ecol. Modell. 391, 1–8. https://doi.org/10.1016/j.ecolmodel.2018.10.024

Pavlacky Jr, D.C., Possingham, H.P., Lowe, A.J., Prentis, P.J., Green, D.J., Goldizen, A.W., 2012. Anthropogenic landscape change promotes asymmetric dispersal and limits regional patch occupancy in a spatially structured bird population. J. Anim. Ecol. 81, 940–952. https://doi.org/https://doi.org/10.1111/j.1365-2656.2012.01975.x

Perrin, S.W., Englund, G., Blumentrath, S., Brian O'Hara, R., Amundsen, P., Gravbrøt Finstad, A., 2020. Integrating dispersal along freshwater ecosystems into species distribution models. Divers. Distrib. 26, 1598–1611. https://doi.org/10.1111/ddi.13112

Peterson, A.T., 2006. Uses and Requirements of Ecological Niche Models and Related Distributional Models. Biodivers. Informatics 3, 59–72. https://doi.org/10.17161/bi.v3i0.29

Peterson, A.T., Martínez-Campos, C., Nakazawa, Y., Martínez-Meyer, E., 2005. Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. Trans. R. Soc. Trop. Med. Hyg. 99, 647–655. https://doi.org/10.1016/j.trstmh.2005.02.004

Peterson, A.T., Sánchez-Cordero, V., Martínez-Meyer, E., Navarro-Sigüenza, A.G., 2006. Tracking population extirpations via melding ecological niche modeling with land-cover information. Ecol. Modell. 195, 229–236. https://doi.org/10.1016/j.ecolmodel.2005.11.020

Peterson, A.T., Soberón, J., 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. Nat. Conserv. 10, 102–107. https://doi.org/10.4322/natcon.2012.019

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. Ecological Niches and Geographic Distributions. Princenton University Press, Princenton.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pringle, J.M., Blakeslee, A.M.H., Byers, J.E., Roman, J., 2011. Asymmetric dispersal allows an upstream region to control population structure throughout a species' range. Proc. Natl. Acad. Sci. 108, 15288–15293. https://doi.org/10.1073/pnas.1100473108

R Core Team, 2019. A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Rangel, T.F., Loyola, R.D., 2012. Labeling Ecological Niche Models. Nat. Conserv. 10, 119–126. https://doi.org/10.4322/natcon.2012.030

Reis, R.E., Kullander, S.O., Ferraris, C.J., 2003. Check list of the freshwater fishes of South and Central America, 1st ed. EDIPUCRS, Porto Alegre.

Revelle, W., 2019. psych: Procedures for Personality and Psychological Research. Evanston, Illinois, USA: Northwestern University. https://cran.r-project.org/package=psych

Rieux, A., Soubeyrand, S., Bonnot, F., Klein, E.K., Ngando, J.E., Mehl, A., Ravigne, V., Carlier, J., de Lapeyre de Bellaire, L., 2014. Long-Distance Wind-Dispersal of Spores in a Fungal Plant Pathogen: Estimation of Anisotropic Dispersal Kernels from an Extensive Field Experiment. PLoS One 9, e103225. https://doi.org/10.1371/journal.pone.0103225

Riginos, C., Hock, K., Matias, A.M., Mumby, P.J., van Oppen, M.J.H., Lukoschek, V., 2019. Asymmetric dispersal is a critical element of concordance between biophysical dispersal models and spatial genetic structure in Great Barrier Reef corals. Divers. Distrib. 25, 1684–1696. https://doi.org/10.1111/ddi.12969

Rocha, B. da S., Souza, C.A. de, Machado, K.B., Vieira, L.C.G., Nabout, J.C., 2020. The relative influence of the environment, land use, and space on the functional and taxonomic structures of phytoplankton and zooplankton metacommunities in tropical reservoirs. Freshw. Sci. 39, 321–333. https://doi.org/10.1086/708949

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C., 2001. Estimating the Support of a High-Dimensional Distribution. Neural Comput. 13, 1443–1471. https://doi.org/10.1162/089976601750264965

Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecol. Lett. 10, 1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Swets, J., 1988. Measuring the accuracy of diagnostic systems. Science (80-.). 240, 1285–1293. https://doi.org/10.1126/science.3287615

Uribe-Rivera, D.E., Soto-Azat, C., Valenzuela-Sánchez, A., Bizama, G., Simonetti, J.A., Pliscoff, P., 2017. Dispersal and extrapolation on the accuracy of temporal predictions from distribution models for the Darwin's frog. Ecol. Appl. 27, 1633–1645. https://doi.org/10.1002/eap.1556

Vasudev, D., Fletcher, R.J., Goswami, V.R., Krishnadas, M., 2015. From dispersal constraints to landscape connectivity: lessons from species distribution modeling. Ecography (Cop.). 38, 967–978. https://doi.org/10.1111/ecog.01306

Ver Hoef, J.M., Peterson, E., Theobald, D., 2006. Spatial statistical models that use flow and stream distance. Environ. Ecol. Stat. 13, 449–464. https://doi.org/10.1007/s10651-006-0022-8

CAPÍTULO 2: Asymmetric dispersal in freshwater species distribution models vary depending on the spatial pattern of fish species

Micael Rosa Parreira[1*], João Carlos Nabout[2]

[1]Universidade Federal de Goiás, Campus Samambaia, 74690-900, Goiânia, GO, Brazil.

[2]Universidade Estadual de Goiás, Campus Central, 75132-903, Anápolis, GO, Brazil.

[*]Corresponding author: Micael Rosa Parreira, Universidade Federal de Goiás, Campus Samambaia, Av. Esperança, s/n - Chácaras de Recreio Samambaia, 74690-900, Goiânia, GO, Brazil. Email: micael_rp@hotmail.com

Abstract

The species movement (dispersal) is commonly ignored in species distribution models (SDMs), generating model overprediction and inaccurate predictions. Besides, different dispersal capacities and distribution patterns should be considered since restricted species are expected to generate more overprediction to climatically similar and geographically close but unreachable areas, especially for climate-only SDMs. In this sense, we estimate and compare the accuracy of models built using environmental and asymmetric dispersal predictors, assessing if the difference in predictions varies depending on the species occurrence distribution and range. Generally, the models built using dispersal predictors generated more accurate predictions than environmental variables. However, even though we found no significant difference in the accuracy among models built using different variables, their accuracy metrics were correlated with the species range and distribution across sub-basins. Besides, more restricted species (i.e., lower range and distribution limited to one sub-basin) showed a higher difference in model accuracy between models built using dispersal and environmental predictors, with more accurate models being generated for restricted species when modeled using dispersal-related predictors only. Therefore, the use of asymmetric dispersal predictors in SDM, besides developing accurate models, avoids/reduces the model's overpredictions to geographically close and climatically suitable areas. This is especially true for restricted species by predicting the species distribution based on their dispersal routes through the actual directional gradient of the basin's hydrographic network.

Keywords: Restricted distribution; Range; ANCOVA; Asymmetric Eigenvector Maps; Species dispersal.

1. Introduction

Dispersal plays a fundamental role in the survival and persistence of species, structuring of populations and communities, and geographic distribution of species, among others (Clobert et al., 2004). Besides, dispersal is essential for species facing changes in habitats due to anthropogenic or natural pressures, where, when they move to other environments, they counteract the effects of environmental changes (Chaine and Clobert, 2012). In freshwater environments, the physical organization of the basin's hydrographic network significantly influences the dispersal of organisms (Tonkin et al., 2018). Specifically for freshwater fish, the dendritic structure of rivers, besides being vital for the development of the life cycle of these organisms (e.g., spawning), is essential to determine their geographic distribution, as they use river and stream channels as dispersal routes (Padial et al., 2014). In this sense, when evaluating the geographic distribution of freshwater species, such as for conservation approaches, the methods used for predicting the potentially suitable areas should consider the dispersal routes used by the species.

Species Distribution Models (SDMs) is a modeling tool often used by biogeographers and ecologists to predict potentially suitable areas for the species distribution (Peterson et al., 2011). There are different frameworks to be used when building an SDM depending on the objective (Anderson, 2013; Soliman et al., 2012), which can generate inaccurate and biased predictions when ignoring their conceptual basis (Araújo and Peterson, 2012; Jiménez-Valverde et al., 2008; Peterson and Soberón, 2012). In this sense, there are three main components for the studied distribution area of species: the Abiotic (A) factors the species exists within, the Biotic (B) interactions, and the “Movement” (M) defined by the reachable areas and limitations of species dispersal (Barve et al., 2011; Soberón, 2007; Soberon and Peterson, 2005). However, many studies using SDMs consider only the abiotic component, as environmental variables, in their predictions (Peterson et al., 2011). Therefore, the species dispersal capacity (movement) is often ignored when modeling the species distribution (Miller and Holloway, 2015), which can generate overpredictions to unreachable but climatically suitable areas for the species (Mendes et al., 2020).

The species movement can be inserted into models by including accessible areas through spatial barriers or the historical species dispersal capacity (Barve et al., 2011; Miller and Holloway, 2015). Specifically for freshwater environments, the species movement can be incorporated by taking into account the hierarchical structure and flow directionality of rivers, which are dispersal routes for species to migrate (Perrin et al., 2020; Tonkin et al., 2018). A way of assessing dispersal in the model’s predictions is by incorporating symmetric spatial filters as predictors in SDMs (Allouche et al., 2008; De Marco et al., 2008). At the same time, in freshwater environments, it should consider the directionality and hierarchical structure of the hydrographic network. In this case, asymmetric spatial filters (AEM) are more appropriate since they take into account the dispersal routes and directionality through rivers, which is assessed through an edge-by-sites directional binary matrix simulating the connections of the basin’s hydrographic network

(Blanchet et al., 2008).

However, it is necessary to consider the dispersal capacity of fish species in predicting potential distribution. For terrestrial models, it is possible to evaluate different dispersal ranges via dispersal/migration rates, using, for example, distance buffers around the current distribution or species-specific fixed dispersal rates. The dispersal rates may be included as dispersal-related predictors in SDMs (Holloway et al., 2016; Monsimet et al., 2020). This becomes important as models can be affected by the size of the accessible area used for modeling, where more restricted species (i.e., lower dispersal capacity) generate models with higher performances given their smaller historically accessible areas (Barve et al., 2011). Indeed, there are many factors, other than extension size, that may affect the performance of SDMs (Elith and Graham, 2009), such as spatial autocorrelation (Guélat and Kéry, 2018; Segurado et al., 2006), uncertainty (Buisson et al., 2010), and species-specific factors, such as prevalence (i.e., the presence/absence ratio; (Jiménez-Valverde et al., 2009), environmental tolerance (Hernandez et al., 2006), species rarity (Franklin et al., 2009), and the sample size/distribution range (Liu et al., 2019; Stockwell and Peterson, 2002; Wisz et al., 2008). Therefore, since species dispersal is a fundamental mechanism for the distribution of freshwater fish and varies depending on species-specific characteristics (Tonkin et al., 2018), it becomes crucial to evaluate how models incorporating dispersal-related predictors perform and are affected by species-specific characteristics in comparison to those using traditional climate variables.

In this sense, we estimate here the relative contribution of environmental and dispersal predictors in distribution models of freshwater species with different distribution ranges. Specifically, we tested if the models' performances vary depending on the species range and distribution among sub-basins, and if this relationship with those predictors is higher depending on the type of variables (environmental or asymmetrical dispersal) used for modeling. We expect a significant effect of species range and distribution among sub-basins on models' performance, especially for models built using asymmetrical dispersal (AEM) since these models consider the species directionality throughout the basin. This prediction is based on the hypothesis that inserting asymmetrical dispersal into SDMs will generate more accurate and realistic models. We expect this will be especially true for species with more restricted distribution (lower range and limited to one sub-basin) due to higher dispersal cost. We expect this result because it adds to the prediction of the directional effect of dispersal through rivers, avoiding overprediction to climatically similar but non-accessible areas or with difficult access.

2. Material and Methods

2.1. Study area

The study area used for modeling is the Tocantins-Araguaia River basin, covering the entire hydrographic network. This basin has two major rivers (Tocantins and Araguaia), forming two sub-basins that merge at one point up to the north region, close to the mouth (Figure 1). Therefore, a species occurring exclusively in one sub-basin must disperse a long way through the main river course occupy the other sub-basin. Thus, this basin is an interesting area for studies on dispersal limitation because it shows the species effort to disperse between the two basins (Tocantins and Araguaia), via the downstream connectivity. The entire hydrographic network was rasterized into grid-cells of 0.5° spatial resolution (latitude and longitude), totaling 282 cells for the Tocantins-Araguaia basin.

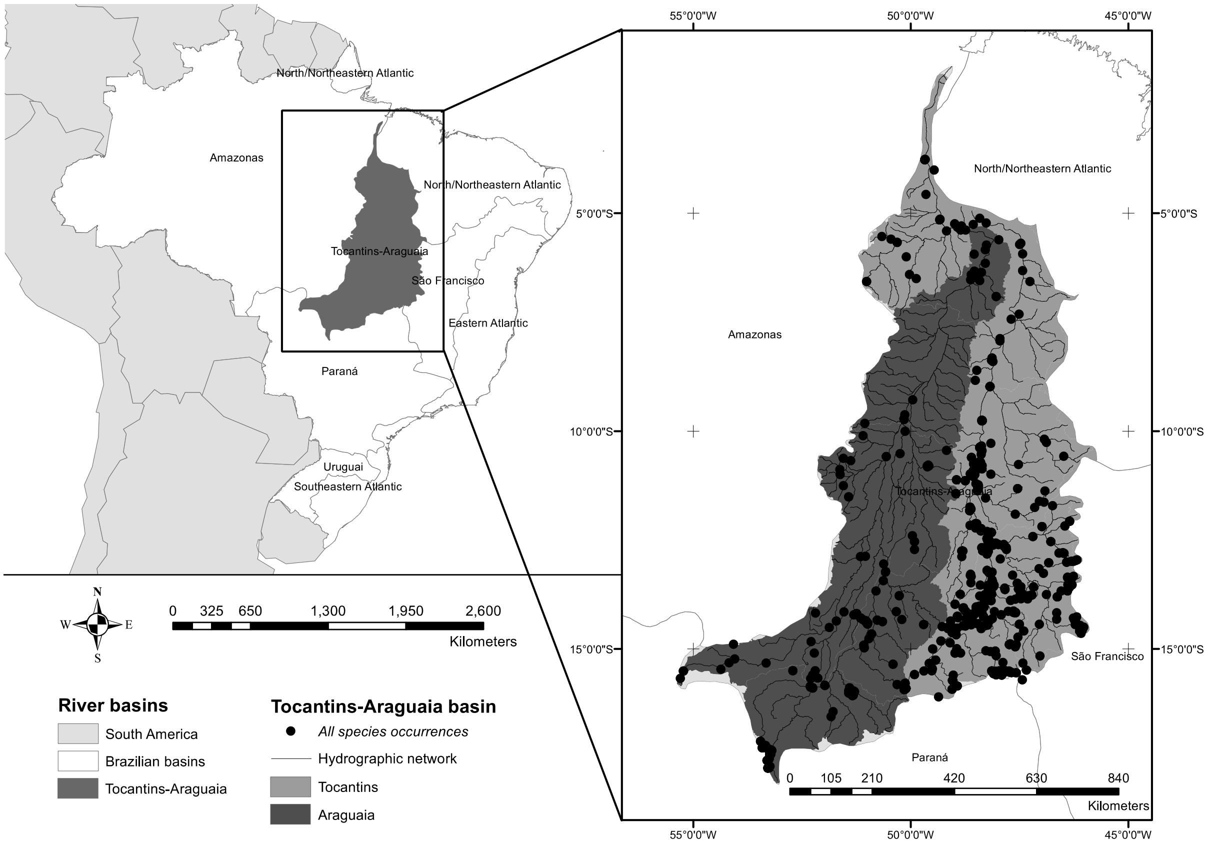


Figure 1 – Study area of the Tocantins-Araguaia River basin together with all species occurrences (black dots). Each species distribution is available in Figure S1.

2.2. Species occurrences

We selected 29 freshwater species native to the Tocantins-Araguaia River basin. The fish occurrence records were obtained in the online databases *Species Link* (specieslink.net)*, GBIF* (gbif.org), and *FishBase* (fishbase.org). The species have different distribution range sizes, where the species with the most restricted range is *Gymnotocinclus anosteos* and with the widest range is *Hypostomus ericae* (Table S1). These species were chosen based on their distributions being restricted to the studied basin and having more than five unique occurrences (non-duplicated within grid-cells). In this sense, for each species, we removed duplicated occurrences (more than one coordinate within cells), by addressing the cell id to each coordinate that falls within, and then, removing the points with duplicated cells ids, i.e., ending up with unique occurrences (only one occurrence per cell). These unique occurrence records (coordinates) for each species were later used to generate the models.

2.3. Environmental variables

The environmental variables used here derive from the 19 bioclimatic variables available in the Worldclim online database for the current scenario (worldclim.com). These variables are derived from the monthly temperature and rainfall values to generate more biologically meaningful variables (Hijmans et al., 2005). These variables were rescaled to a spatial resolution of 0.5º (55.6 km at the equator) using the function *aggregate* of package *raster* (Hijmans, 2019) available in the R software (R Core Team, 2019). The non-collinear environmental variables (ENV) used for modeling were selected using functions of the package *psych* (Revelle, 2019) available in the R software. The number of necessary non-orthogonal axes was determined using the *fa* function that compares the eigenvalues of factors (axes) in a scree plot, which retained four factors. We then selected the variables with the highest positive or negative loadings in each of the four axes. Therefore, the following variables were selected: BIO1 = Annual Mean Temperature, BIO2 = Mean Diurnal Temperature Range, BIO13 = Precipitation of Wettest Month, and BIO15 = Precipitation Seasonality (Figure S2).

2.4. Asymmetric dispersal variables

The asymmetric eigenvector maps (AEM) variables (hereafter called asymmetric dispersal variables) were generated using the function *aem* of the *adespatial* (Dray et al., 2019) available in the R software. The calculation of these variables in this function is based on the singular value decomposition (SVD) analysis using an asymmetric edge-by-sites binary matrix. In our study, this matrix is represented by a binary matrix of directional connectivity (asymmetric) using the grid cells as sites with nodes between grid cells and tributaries, besides the edge as the river path, until it reaches a riverbed. The first edge (E1) represents the longest river path until it reaches the farthest riverbed. The subsequent edge is the same as the previous minus the first node (grid-cell), only changing when it reaches a connection to a lower-order river tributary, which creates one or more edges (depending on tributary size and sub-connections) exclusively to this river branching. After finishing every tributary edge(s), the edges counting return to the main connection (river), receiving 1s for the remaining grid-cells (nodes) in the edge and 0s for the nodes already included in previous edges until it reaches the farthest node (grid-cell with the farthest source) (see Parreira et al. 2021 [submitted] for more explanation on the matrix's construction). The construction of this binary matrix is based on (Blanchet et al., 2008) modeling of directional spatial processes in freshwater environments. Posteriorly, we used the broken stick method (Jackson, 1993) to determine the number of non-collinear asymmetric eigenvectors (AEM). These selected axes (eigenvectors) were then used as asymmetrical dispersal predictors for generating the species distribution models (see Figure S3 for mapped predictors).

2.5. Modeling Procedures

Species distribution models (SDMs) were built using 75% of presence and absence points for training (model's construction) and 25% for testing the models' performance (Guisan and Zimmermann, 2000), randomly selected from each species' unique presences and the 56 randomly generated pseudo-absences. The set of presence and absence points selected for training (constructing) were used together with modeling methods (algorithms) and sets of environmental and spatial variables. We used six algorithms: Bioclim (Nix, 1986), Domain (Carpenter et al., 1993), Support Vector Machines (SVM; Schölkopf et al., 2001), Generalized Linear Models (GLM; Nelder and Wedderburn, 1972), Maximum Entropy (MaxEnt; Phillips et al., 2006), and Random Forest (Breiman, 2001). The algorithms used are available in the R package *dismo* (Hijmans et al., 2016) and were chosen because they consider different statistical methods (climatic envelopes, environmental distances, machine-learning, regressions) (Rangel and Loyola, 2012). In addition to algorithms, the models were built using different types of variables (predictors): environmental variables (ENV), asymmetric dispersal (AEM), and the combination of environmental and dispersal variables (ENV+AEM). The model's training and validation occurred 20 times through cross-validation (using the selected presences and pseudo-absences) for all combinations of algorithms and variables for each species. In this sense, we built 360 models (20 repetitions x 6 algorithms x 3 types of variables) for each of the 29 species, totaling 10,440 models.

All models generated were evaluated using the Area Under the "receiver operating characteristic" Curve (AUC; Swets, 1988), which is an evaluation metric threshold-independent (the limit for determination of presences/absences) that compare predict with observed values. We also used True Skill Statistic (TSS; Allouche et al., 2006), which compares the number of correct predictions minus those attributable to random guessing, which is not affected by the prevalence and size of the validation set. The final potential distribution map is a consensus map (ensemble) built using the mean suitability values of all models with AUC > 0.7 weighted by the AUC values for each species. The models were generated and evaluated using functions from the *dismo* package (Hijmans et al., 2016) available in the R software (R Core Team, 2019).

2.6. Calculation of variables

During model evaluation in each run, we obtained the evaluation metrics (AUC and TSS) for each model. The metrics (accuracy) were summarized for each species. The overall mean AUC and TSS values for each species were calculated using the mean accuracy metric values of models built using the six different algorithms for each of the three sets of variables (AEM, ENV, AEM+ENV). Thus, we ended up with overall mean AUC and TSS values for each species for each set of variables, which are used as response variables for the analysis after that.

The distribution range of species was calculated using the Minimum Convex Polygon (MCP) approach, which is the smallest possible polygon with straight lines containing all occurrence records. We calculated the MCP in Km2 using the *Minimum Bounding Geometry* function available in ArcGIS software version 10.5. We also calculated the number of non-duplicated records for each species to be used as species range. However, since this variable was highly correlated with MCP values (r=0.87), we kept only MCP since it better represents the distribution of poorly surveyed species (e.g., few occurrences widely distributed throughout the basin).

Finally, through the distribution maps of each species, we assessed whether the occurrence records fall within only one sub-basin (Tocantins or Araguaia) or two sub-basins (Tocantins and Araguaia). Besides, for species with distribution only in one sub-basin, we determined which sub-basin it is. These variables are mainly used to control the effect of accuracy on models using species with small range size but distributed across both sub-basins (e.g., *Acestrocephalus maculosus* – Figure S1.1) and large range but within only one sub-basin (e.g., *Moenkhausia tergimacula* – Figure S1.22).

2.7. Data analysis

First, we calculated an Analysis of Variance (ANOVA) to test the difference between the mean AUC and TSS values of each species among the type of variables. We used an Analysis of Covariance (ANCOVA) to test the relationship between species range and the model's evaluation metrics (AUC and TSS) and if this relationship varies according to the variables used (ENV, AEM, ENV+AEM). For these analyses, AUC and TSS are the dependent variables (continuous, varying from 0 to 1), and species range is the continuous independent variable. The type of variables used for modeling is the covariate (nominal categoric) with three levels: ENV (environmental variables), AEM (asymmetric dispersal variables), and ENV+AEM (environment and asymmetric dispersal variables). (Table S1).

Subsequently, we tested whether the accuracy metrics are significantly different between AEM-based and ENV-based models by calculating delta AUC and TSS. The delta accuracy here is the mean accuracy values from AEM models subtracted from mean accuracy values from ENV models. In this sense, AUC and TSS positive values represent models for those specific species, AEM generated models with higher performance than models built using ENV. On the contrary, AUC and TSS negative values show for which species the models built using ENV had higher performance than AEM-based models. Besides, we took those delta accuracy values (response variables) and calculated an ANOVA to test a relationship between delta accuracy and the Number of basins each species occurs within (Delta_AUC ~ Nbasins and Delta_TSS ~ Nbasins). We also calculated a linear regression between Delta accuracy and species distribution range.

3. Results

Among the 29 species modeled, most species (62%) had models with the best performance (both AUC and TSS) when modeled using AEM predictors. At the same time, 21% had better performance using ENV predictors, and 17% using the combination of AEM and ENV predictors. Generaly, the models built for each species had good performances (Figure 2). Among the models built using only asymmetric dispersal (AEM), 24 (83%) species had models considered adequate (AUC >= 0.75; TSS >= 0.5). For models using environmental variables (ENV), 21 (72%) had performances above the acceptance threshold, and for models using both ENV and AEM, 23 (79%) species had models with performances above the threshold. Therefore, AEM-based models, besides having on average the highest accuracy (AUC = 0.82; TSS = 0.71), had more species with models' performance above the acceptance threshold. However, the overall difference among types of variables using the mean accuracy values of each species model was not statistically significant (AUC: p = 0.87; TSS: p = 0.62), therefore indicating that AEM and ENV models presented similar performances (Figure 2).

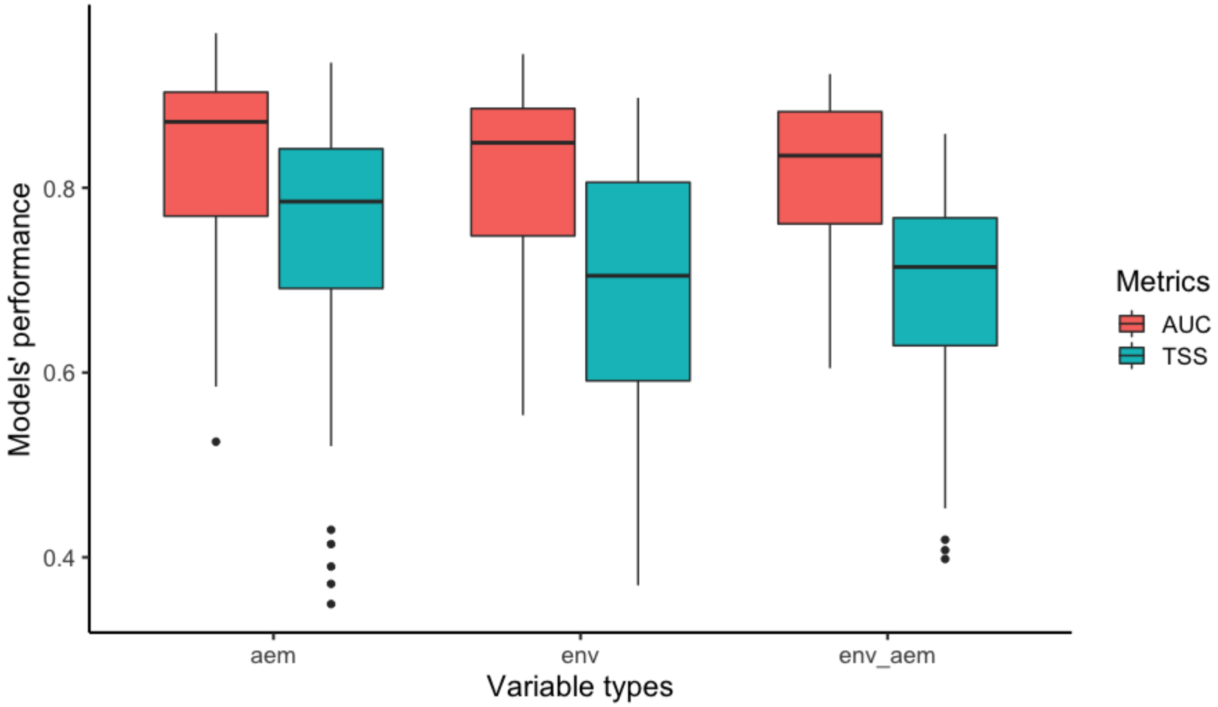


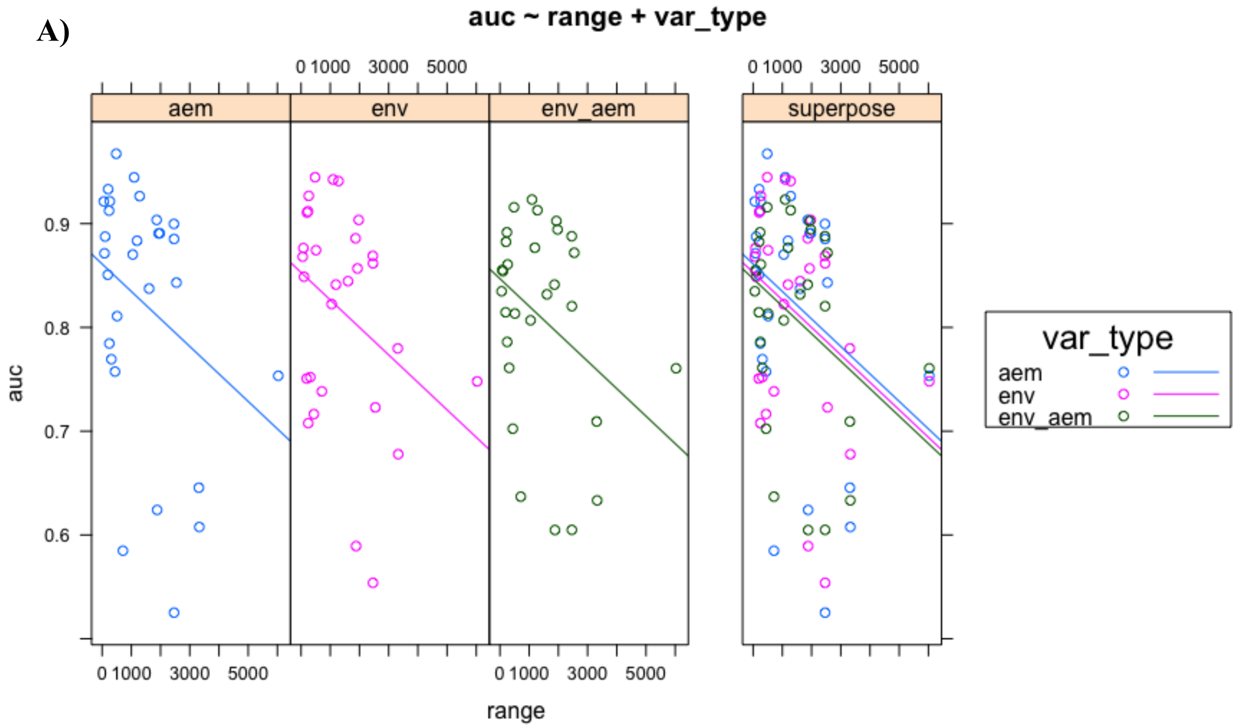
Figure 2 – Difference in model's performance, using AUC and TSS evaluation metrics, among models built using different types of variables for all species.

ANCOVA analysis for both evaluation metrics (AUC and TSS) showed no significant difference between types of variables. However, the species range showed a significant relationship with both accuracy values. Therefore, models built for species with more restricted distribution show better performance, i.e., with more realistic predictions, regardless of the model type (Table 1 and Figure 3). Nevertheless, range effect on AEM models is slightly higher than on ENV models. Therefore, although there is not a significant difference between the performance of models using different types of variables and species range, the effect of the species range is higher for models built using asymmetric dispersal predictors (AEM).

Table 1 – ANCOVA results for both response variables (AUC and TSS). *Significance level at 1% of confidence.

AUC						TSS					
	Df	SS	MS	F	*P*		Df	SS	MS	F	*P*
range	1	0.108	0.108	10.49	0.00*	range	1	0.435	0.435	22.88	0.00*
var_type	2	0.003	0.001	0.14	0.87	var_type	2	0.020	0.010	0.59	0.56
Residuals	83	0.851	0.010			Residuals	83	1.578	0.019		

A)



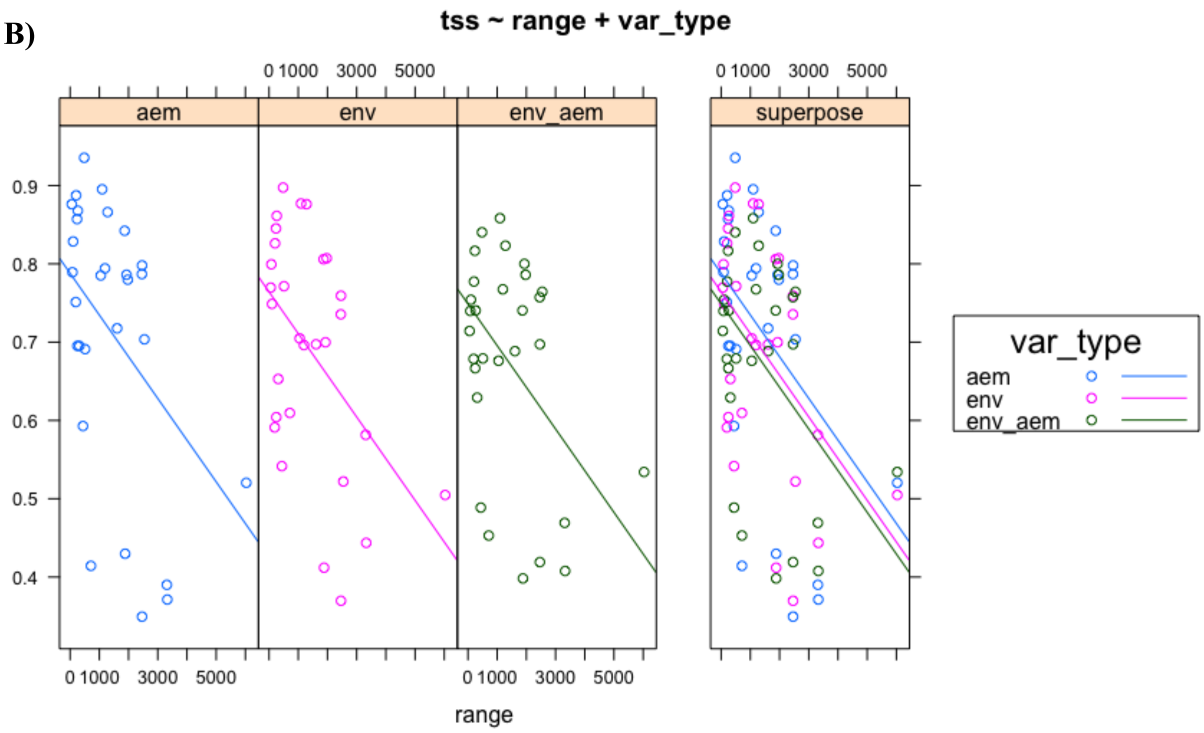


Figure 3 – Relationship between A) AUC and B) TSS with the species range (km2/100) among the models built using different types of variables.

The comparison of delta accuracy values (AUC and TSS) between AEM- and ENV-models for each species had different but complementary results among the number of sub-basins and distribution ranges for each species. We found no relationship between range and accuracy metrics between AEM and ENV models, for both AUC ($R^2 = 0.04$, $p < 0.29$) and TSS ($R^2 = 0.04$, $p < 0.26$) metrics. However, the delta accuracy was significantly different between models for species distributed only in one sub-basin and distributed in both sub-basins, for both AUC ($F_{(1, 27)} = 9.18$, $p < 0.01$) and TSS ($F_{(1, 27)} = 8.50$, $p < 0.01$) metrics (Figure 4). It means that species distributed in only one basin modeled using asymmetric dispersal (AEM) generally had higher performance than using environmental variables (ENV). Besides, this positive delta was significantly different from models from species distributed in both basins. That difference is especially higher for more restricted species distributed within the same sub-basin (Figure 4). Except for species widely distributed latitudinally but still within the same sub-basin, such as *Moenkhausia tergimacula* (Figure S1.21), which also had quite higher accuracy when modeled using AEM than ENV. Nevertheless, ENV models for each species had a generally higher performance for highly dispersed species in both basins (regardless of distribution range). For example, *Moenkhausia pyrophthalma* and *Ammoglanis diaphanus* have quite different distribution range sizes but both had higher performance when modeled using ENV than using AEM (Figure S1.1 and Figure S1.21).

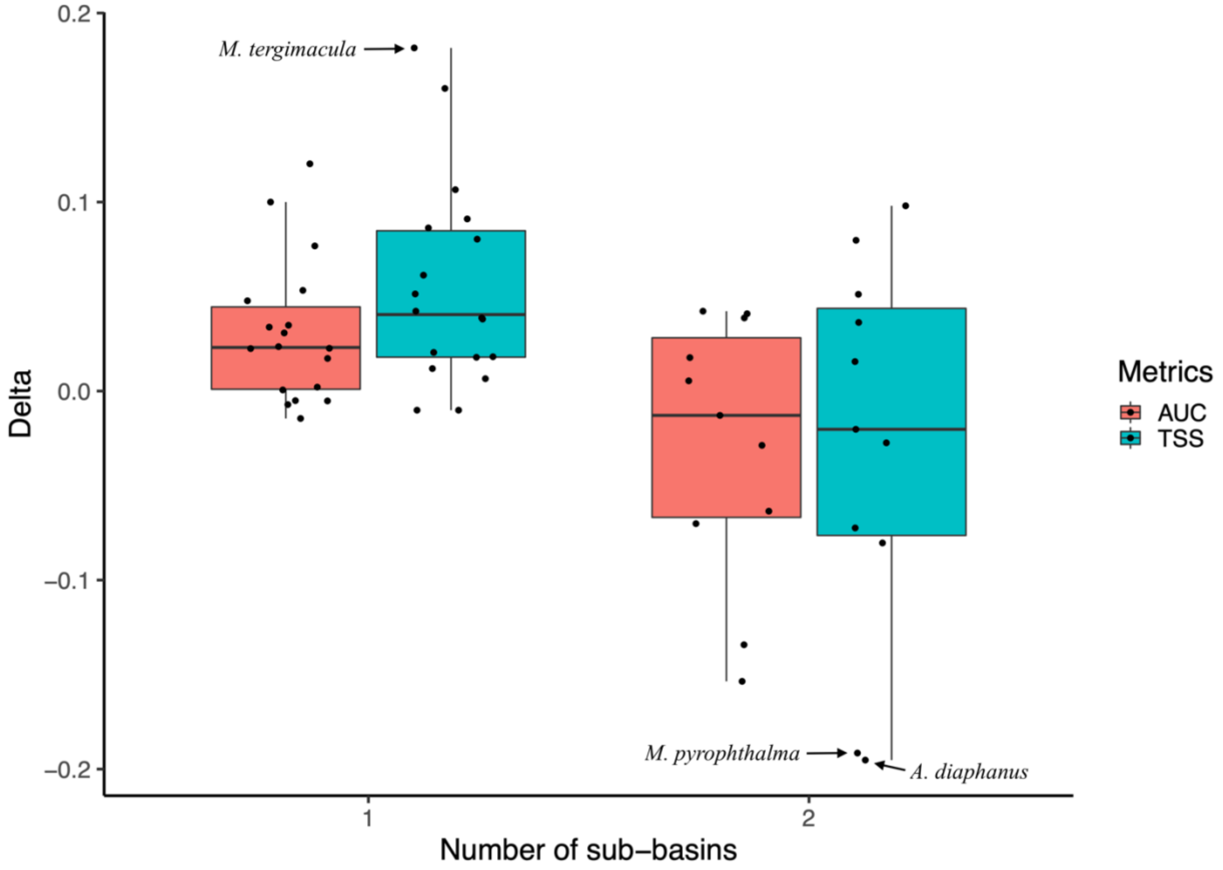


Figure 4 – Delta values of accuracy metrics (AUC and TSS) of all species modeled using AEM and ENV distributed within one or two sub-basins.

4. Discussion

Even though not significantly different, in our study, the models' accuracy is generally higher for AEM models than ENV models. Moreover, this accuracy difference among the built SDMs varies depending on the species range size combined with the species distribution within the main sub-basins. When modeled using AEM, restricted species had models with higher accuracy and less overprediction to areas climatically suitable but longitudinally disconnected. Therefore, the insertion of asymmetric dispersal (AEM) predictors into freshwater SDMs, besides generating models with overall high accuracy for the fish species evaluated (higher for most species), avoids or reduces the overprediction to climatically close but, in many cases, disconnected areas that occur when using only traditional environmental variables.

In general, models built using asymmetric dispersal predictors (AEM) had higher performance than those built using only environmental predictors (ENV) and the combination of AEM and ENV predictors. This trend continues specifically for most species, with models with higher performance and above the acceptable threshold using AEM predictors. Even though this effect could be due to environmental or physical variables not measured here (e.g., water flow velocity), the inclusion of spatial predictors in species distribution models can generate accurate models by including the spatial structure (variation) of the studied area (e.g., coordinates [x and y gradients] and large- and fine-spatial filters) (Bahn and McGill, 2007; Stockwell and Peterson, 2002), spatial autocorrelation (Guélat and Kéry, 2018), and species dispersal, either symmetric through dispersal rates (Holloway et al., 2016; Monsimet et al., 2020) or asymmetric (Parreira et al. 2022 [in submission]). In this sense, the inclusion of species dispersal (movement) through asymmetric dispersal (AEM) predictors can generate more reliable and accurate models by reducing overprediction to close and climatically similar but disconnected areas (e.g., rivers separated by physical barriers [natural: elevation; artificial: dams]). This occurs by including a directional gradient into the predictors that matches the hierarchy and flow direction of river basins (Blanchet et al., 2008).

Species more restricted (i.e., small range) tend to have models with higher accuracy (Stockwell and Peterson, 2002), with no overall differences in this relationship depending on the predictors used for modeling. However, the difference among predictors arises when considering the basin limitation (limited accessible areas) for the species distribution. In our study, species with more restricted distribution due to their small range and distribution limited to a sub-basin had models with higher performances (accuracy). These restricted species can generate models with higher performances (accuracy) due to their limited historically accessible areas (Barve et al., 2011), with prediction accuracy decreasing as the range size increases (Stockwell and Peterson, 2002). Furthermore, by correlating these predictors with the model's delta accuracy for each species, we found a higher and significant difference in performance between models built using only AEM and only ENV, with models built using AEM having higher accuracy, especially for restricted species. AEM-based models increase accuracy since it includes the longitudinal disconnection between the main sub-basins (Araguaia and Tocantins), which act as a dispersal barrier between sub-basins; thus, including into models the directionality for the species dispersal. Hence, the insertion of asymmetric dispersal into freshwater SDMs, besides generating more accurate predictions, more realistic represents the actual distribution through rivers of continental fish, increasing the ecological reality of predictions.

One of the issues of using only environmental predictors when modeling is to not consider the dispersal movement (Miller and Holloway, 2015), either *a priori* during the modeling process either (e.g., as explanatory variables) or a posteriori (e.g., by overlapping accessible and suitable areas), generating overpredictions (Mendes et al., 2020). In this study, the species most affected by overprediction when modeled using only environmental variables are species with restricted or wide distributions throughout one sub-basin but with few occurrence points in the other sub-basin, geographically close to the species core distribution. For example, both *Moenkhausia pyrophthalma* (wide distribution) and *Ammoglanis diaphanus* (restricted distribution) had quite higher performance when modeled using ENV than AEM predictors. This happened due to not considering the species dispersal route throughout the rivers to reach the geographically close and climatically suitable areas but disconnected in the upper portion of the basin. Thus, it generated overpredicted distribution when modeling these species with only environmental variables.

Therefore, the models built using environmental and dispersal-related variables show different accuracies (positive or negative deltas) depending on the species range size and distribution across sub-basins. However, those differences in accuracy for models built using those variables for each species are small and not statistically significant for the overall difference among models. Besides, some species have quite small sample sizes due to their narrow distribution in large grid cells (i.e., small unique presences and low prevalence). This, in turn, generates models with poor accuracy, expected results for accuracy in models built with small samples (Liu et al., 2019), and low prevalence (Jiménez-Valverde et al., 2009). Despite that, AEM-based models were able to address this problem, avoiding overprediction to areas outside the restricted species core distribution, generating overall higher performance for those species with limited distribution and sample size.

We conclude that the predicted suitability distributions and accuracy between AEM and ENV models did not vary much or were slightly higher for restricted species using asymmetric dispersal (AEM) predictors. Dispersal predictors can be used as a surrogate of climate-based environmental variables or as complementary variables to insert dispersal restrictions into SDMs, aiming to reduce model overprediction. Furthermore, asymmetric dispersal in SDMs should be explored more in larger-scale studies (e.g., assessing dispersal restrictions among basins) and in freshwater studies modeling species distribution in areas with anthropogenic dispersal restrictions. For example, building SDM with asymmetric dispersal predictors enables the insertion of river disconnections due to anthropic actions, such as hydroelectric power plants (HPP) or water dams. These variables can be inserted as disconnections (0s in the asymmetric binary matrix) in grid cells that have built or planned HPPs, assessing this effect on the species' potential distribution. Therefore, dispersal-related predictors into SDMs allow a better representation of the species' habitat and movement, avoiding/reducing overprediction to unreachable, climatically suitable areas.

5. Acknowledgments

We thank Fabrício Barreto Teresa for the help in obtaining the species distribution data. MRP thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – finance code 001) and FAPEG for the PhD scholarship received. Our work on SDMs has been continuously supported by different grants: CNPq, Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG/process 201710267000519), and National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation (MCTI/CNPq/FAPEG/465610/2014-5), and Brazilian Network on Global Climate Change Research (Rede CLIMA). JCN was supported by a CNPq productivity fellowship.

6. Supplementary Material

Table S1 – Output table with variables used in the ANCOVA. Species = The 29 species selected for modeling; AUC and TSS = models' evaluation metrics; Var_type = sets of variables used for modeling; Range = species' distribution area (MCP); N_basin = Number of sub-basins the species is currently present; Sub_basin = Sub-basins the species are present (T = Tocantins, A = Araguaia, TA = Tocantins and Araguaia sub-basins).

ID	Species	AUC	TSS	Var_type	Range (km2)	N_basin	Sub_basin
1	*Acestrocephalus maculosus*	0.714	0.566	AEM	44,050.60	2	TA
2	*Ammoglanis diaphanus*	0.510	0.335	AEM	71,209.73	2	TA
3	*Ancistrus aguaboensis*	0.773	0.622	AEM	197,380.80	2	TA
4	*Ancistrus minutus*	0.727	0.584	AEM	160,910.87	1	T
5	*Apareiodon cavalcante*	0.777	0.646	AEM	104,877.81	1	T
6	*Apteronotus camposdapazi*	0.849	0.761	AEM	109,633.95	1	T
7	*Aspidoras eurycephalus*	0.824	0.681	AEM	193,315.62	1	T
8	*Astyanax courensis*	0.824	0.734	AEM	20,591.77	1	T
9	*Astyanax unitaeniatus*	0.907	0.832	AEM	26,361.83	1	T
10	*Corumbataia tocantinensis*	0.836	0.723	AEM	128,408.65	1	T
11	*Corydoras maculifer*	0.617	0.452	AEM	188,335.24	1	A
12	*Creagrutus atrisignum*	0.825	0.679	AEM	245,773.82	1	T
13	*Creagrutus saxatilis*	0.832	0.747	AEM	186,715.05	2	TA
14	*Gymnotocinclus anosteos*	0.768	0.700	AEM	5,660.79	1	T
15	*Hemiancistrus cerrado*	0.852	0.786	AEM	10,217.47	2	TA
16	*Hemiancistrus micrommatos*	0.726	0.603	AEM	51,218.94	2	TA
17	*Hemiancistrus spinosissimus*	0.781	0.646	AEM	246,277.88	1	T
18	*Hypostomus asperatus*	0.910	0.843	AEM	48,196.39	1	T

19	*Hypostomus ericae*	0.651	0.380	AEM	603,029.38	2	TA
20	*Loricaria lata*	0.450	0.262	AEM	246,254.50	2	TA
21	*Moenkhausia pyrophthalma*	0.622	0.366	AEM	331,170.63	2	TA
22	*Moenkhausia tergimacula*	0.756	0.583	AEM	254,380.43	1	T
23	*Pimelodus joannis*	0.743	0.649	AEM	24,770.60	1	T
24	*Pimelodus stewarti*	0.828	0.732	AEM	19,377.87	1	T
25	*Rineloricaria osvaldoi*	0.755	0.637	AEM	119,420.86	2	TA
26	*Steindachnerina gracilis*	0.568	0.318	AEM	332,734.87	2	TA
27	*Steindachnerina notograptos*	0.876	0.796	AEM	8,047.82	1	T
28	*Sternarchorhynchus mesensis*	0.819	0.767	AEM	23,789.12	1	T
29	*Xenurobrycon coracoralinae*	0.729	0.632	AEM	31,655.08	1	A
1	*Acestrocephalus maculosus*	0.634	0.459	ENV	44,050.60	2	TA
2	*Ammoglanis diaphanus*	0.690	0.539	ENV	71,209.73	2	TA
3	*Ancistrus aguaboensis*	0.838	0.704	ENV	197,380.80	2	TA
4	*Ancistrus minutus*	0.746	0.588	ENV	160,910.87	1	T
5	*Apareiodon cavalcante*	0.771	0.664	ENV	104,877.81	1	T
6	*Apteronotus camposdapazi*	0.901	0.831	ENV	109,633.95	1	T
7	*Aspidoras eurycephalus*	0.819	0.641	ENV	193,315.62	1	T
8	*Astyanax courensis*	0.869	0.767	ENV	20,591.77	1	T
9	*Astyanax unitaeniatus*	0.878	0.790	ENV	26,361.83	1	T
10	*Corumbataia tocantinensis*	0.870	0.776	ENV	128,408.65	1	T
11	*Corydoras maculifer*	0.572	0.398	ENV	188,335.24	1	A
12	*Creagrutus atrisignum*	0.791	0.629	ENV	245,773.82	1	T
13	*Creagrutus saxatilis*	0.766	0.659	ENV	186,715.05	2	TA
14	*Gymnotocinclus anosteos*	0.770	0.644	ENV	5,660.79	1	T

15	*Hemiancistrus cerrado*	0.800	0.703	ENV	10,217.47	2	TA
16	*Hemiancistrus micrommatos*	0.762	0.614	ENV	51,218.94	2	TA
17	*Hemiancistrus spinosissimus*	0.755	0.605	ENV	246,277.88	1	T
18	*Hypostomus asperatus*	0.880	0.807	ENV	48,196.39	1	T
19	*Hypostomus ericae*	0.642	0.382	ENV	603,029.38	2	TA
20	*Loricaria lata*	0.500	0.285	ENV	246,254.50	2	TA
21	*Moenkhausia pyrophthalma*	0.696	0.460	ENV	331,170.63	2	TA
22	*Moenkhausia tergimacula*	0.671	0.462	ENV	254,380.43	2	TA
23	*Pimelodus joannis*	0.612	0.512	ENV	24,770.60	1	T
24	*Pimelodus stewarti*	0.700	0.515	ENV	19,377.87	1	T
25	*Rineloricaria osvaldoi*	0.809	0.665	ENV	119,420.86	2	TA
26	*Steindachnerina gracilis*	0.651	0.407	ENV	332,734.87	2	TA
27	*Steindachnerina notograptos*	0.909	0.831	ENV	8,047.82	1	T
28	*Sternarchorhynchus mesensis*	0.813	0.739	ENV	23,789.12	1	T
29	*Xenurobrycon coracoralinae*	0.713	0.603	ENV	31,655.08	1	A
1	*Acestrocephalus maculosus*	0.639	0.446	ENV_AEM	44,050.60	2	TA
2	*Ammoglanis diaphanus*	0.584	0.376	ENV_AEM	71,209.73	2	TA
3	*Ancistrus aguaboensis*	0.791	0.620	ENV_AEM	197,380.80	2	TA
4	*Ancistrus minutus*	0.727	0.568	ENV_AEM	160,910.87	1	T
5	*Apareiodon cavalcante*	0.762	0.619	ENV_AEM	104,877.81	1	T
6	*Apteronotus camposdapazi*	0.873	0.785	ENV_AEM	109,633.95	1	T
7	*Aspidoras eurycephalus*	0.838	0.684	ENV_AEM	193,315.62	1	T
8	*Astyanax courensis*	0.836	0.719	ENV_AEM	20,591.77	1	T
9	*Astyanax unitaeniatus*	0.833	0.709	ENV_AEM	26,361.83	1	T
10	*Corumbataia tocantinensis*	0.850	0.731	ENV_AEM	128,408.65	1	T

11	*Corydoras maculifer*	0.568	0.369	ENV_AEM	188,335.24	1	A
12	*Creagrutus atrisignum*	0.812	0.659	ENV_AEM	245,773.82	1	T
13	*Creagrutus saxatilis*	0.721	0.584	ENV_AEM	186,715.05	2	TA
14	*Gymnotocinclus anosteos*	0.762	0.649	ENV_AEM	5,660.79	1	T
15	*Hemiancistrus cerrado*	0.778	0.643	ENV_AEM	10,217.47	2	TA
16	*Hemiancistrus micrommatos*	0.715	0.532	ENV_AEM	51,218.94	2	TA
17	*Hemiancistrus spinosissimus*	0.780	0.628	ENV_AEM	246,277.88	1	T
18	*Hypostomus asperatus*	0.897	0.824	ENV_AEM	48,196.39	1	T
19	*Hypostomus ericae*	0.652	0.379	ENV_AEM	603,029.38	2	TA
20	*Loricaria lata*	0.481	0.275	ENV_AEM	246,254.50	2	TA
21	*Moenkhausia pyrophthalma*	0.646	0.393	ENV_AEM	331,170.63	2	TA
22	*Moenkhausia tergimacula*	0.778	0.605	ENV_AEM	254,380.43	2	TA
23	*Pimelodus joannis*	0.729	0.606	ENV_AEM	24,770.60	1	T
24	*Pimelodus stewarti*	0.806	0.661	ENV_AEM	19,377.87	1	T
25	*Rineloricaria osvaldoi*	0.787	0.641	ENV_AEM	119,420.86	2	TA
26	*Steindachnerina gracilis*	0.598	0.341	ENV_AEM	332,734.87	2	TA
27	*Steindachnerina notograptos*	0.835	0.708	ENV_AEM	8,047.82	1	T
28	*Sternarchorhynchus mesensis*	0.816	0.714	ENV_AEM	23,789.12	1	T
29	*Xenurobrycon coracoralinae*	0.672	0.538	ENV_AEM	31,655.08	1	A

Table S2 – Delta accuracy values and predictors used for Linear Regression Models. Delta_AUC and Delta_TSS = AUC and TSS values of AEM-based models subtracted from AUC and TSS values of ENV-based models.

ID	Species	Delta_AUC	Delta_TSS	Range (km2)	N_basin	Sub_basin

1	*Acestrocephalus maculosus*	0.081	0.107	44050.60	2	TA
2	*Ammoglanis diaphanus*	-0.180	-0.204	71209.73	2	TA
3	*Ancistrus aguaboensis*	-0.065	-0.082	197380.80	2	TA
4	*Ancistrus minutus*	-0.019	-0.003	160910.87	1	T
5	*Apareiodon cavalcante*	0.006	-0.018	104877.81	1	T
6	*Apteronotus camposdapazi*	-0.052	-0.070	109633.95	1	T
7	*Aspidoras eurycephalus*	0.005	0.039	193315.62	1	T
8	*Astyanax courensis*	-0.044	-0.033	20591.77	1	T
9	*Astyanax unitaeniatus*	0.028	0.042	26361.83	1	T
10	*Corumbataia tocantinensis*	-0.035	-0.053	128408.65	1	T
11	*Corydoras maculifer*	0.045	0.054	188335.24	1	A
12	*Creagrutus atrisignum*	0.034	0.050	245773.82	1	T
13	*Creagrutus saxatilis*	0.066	0.089	186715.05	2	TA
14	*Gymnotocinclus anosteos*	-0.002	0.056	5660.79	1	T
15	*Hemiancistrus cerrado*	0.052	0.083	10217.47	2	TA
16	*Hemiancistrus micrommatos*	-0.036	-0.011	51218.94	2	TA
17	*Hemiancistrus spinosissimus*	0.025	0.040	246277.88	1	T
18	*Hypostomus asperatus*	0.029	0.036	48196.39	1	T
19	*Hypostomus ericae*	0.009	-0.003	603029.38	2	TA
20	*Loricaria lata*	-0.050	-0.023	246254.50	2	TA
21	*Moenkhausia pyrophthalma*	-0.074	-0.094	331170.63	2	TA
22	*Moenkhausia tergimacula*	0.085	0.121	254380.43	2	TA
23	*Pimelodus joannis*	0.131	0.137	24770.60	1	T
24	*Pimelodus stewarti*	0.127	0.217	19377.87	1	T
25	*Rineloricaria osvaldoi*	-0.054	-0.028	119420.86	2	TA

26	*Steindachnerina gracilis*	-0.083	-0.089	332734.87	2	TA
27	*Steindachnerina notograptos*	-0.033	-0.036	8047.82	1	T
28	*Sternarchorhynchus mesensis*	0.006	0.029	23789.12	1	T
29	*Xenurobrycon coracoralinae*	0.016	0.029	31655.08	1	A

Figure S1 – All species occurrences (black dots) plotted along with the randomly generated pseudo-absences (red Xs) within the gridded Tocantins-Araguaia River basin divided by the two main sub-basins (Dark grey area = Araguaia portion; Light grey area = Tocantins portion). Figure numbers correspond to the species ID available in Tables S1 and S2.

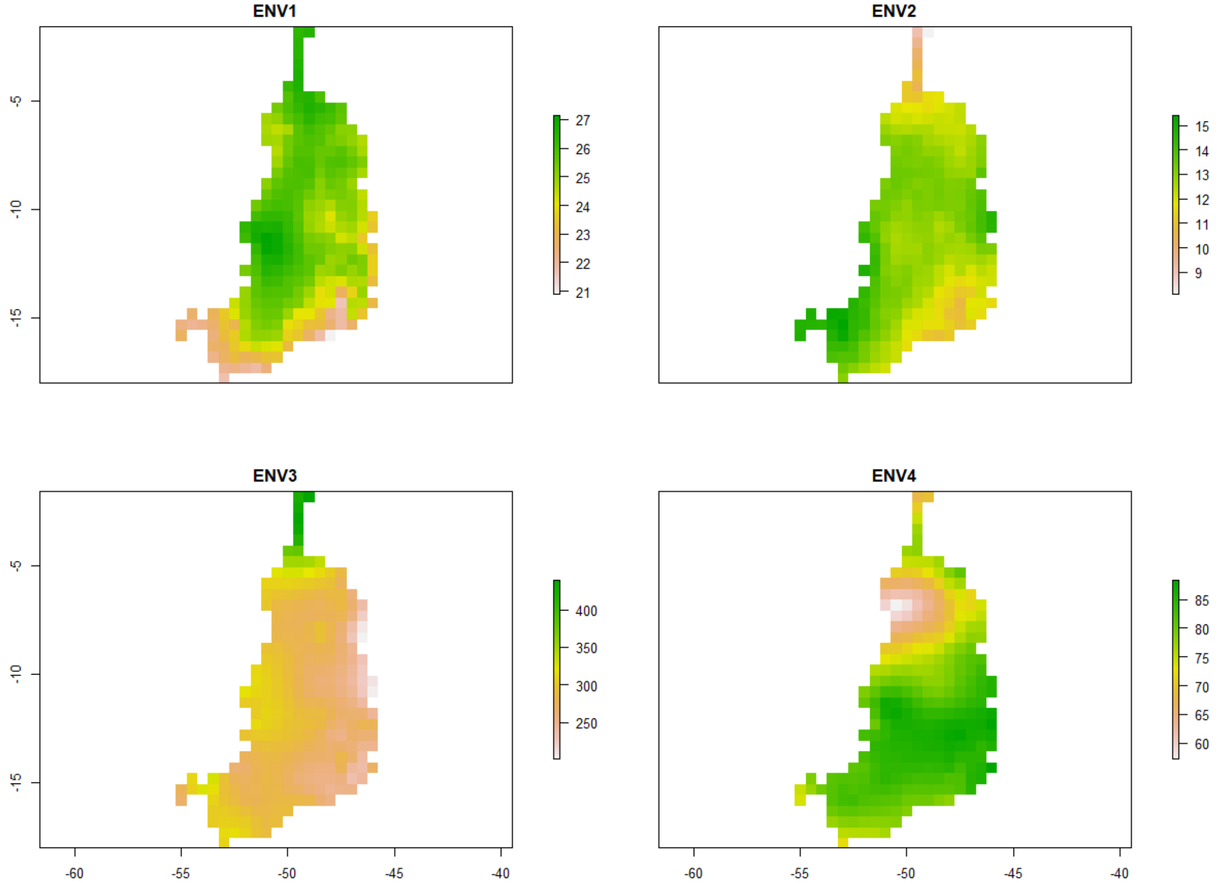


Figure S2 – Environmental (bioclimatic) variables selected for SDM construction.

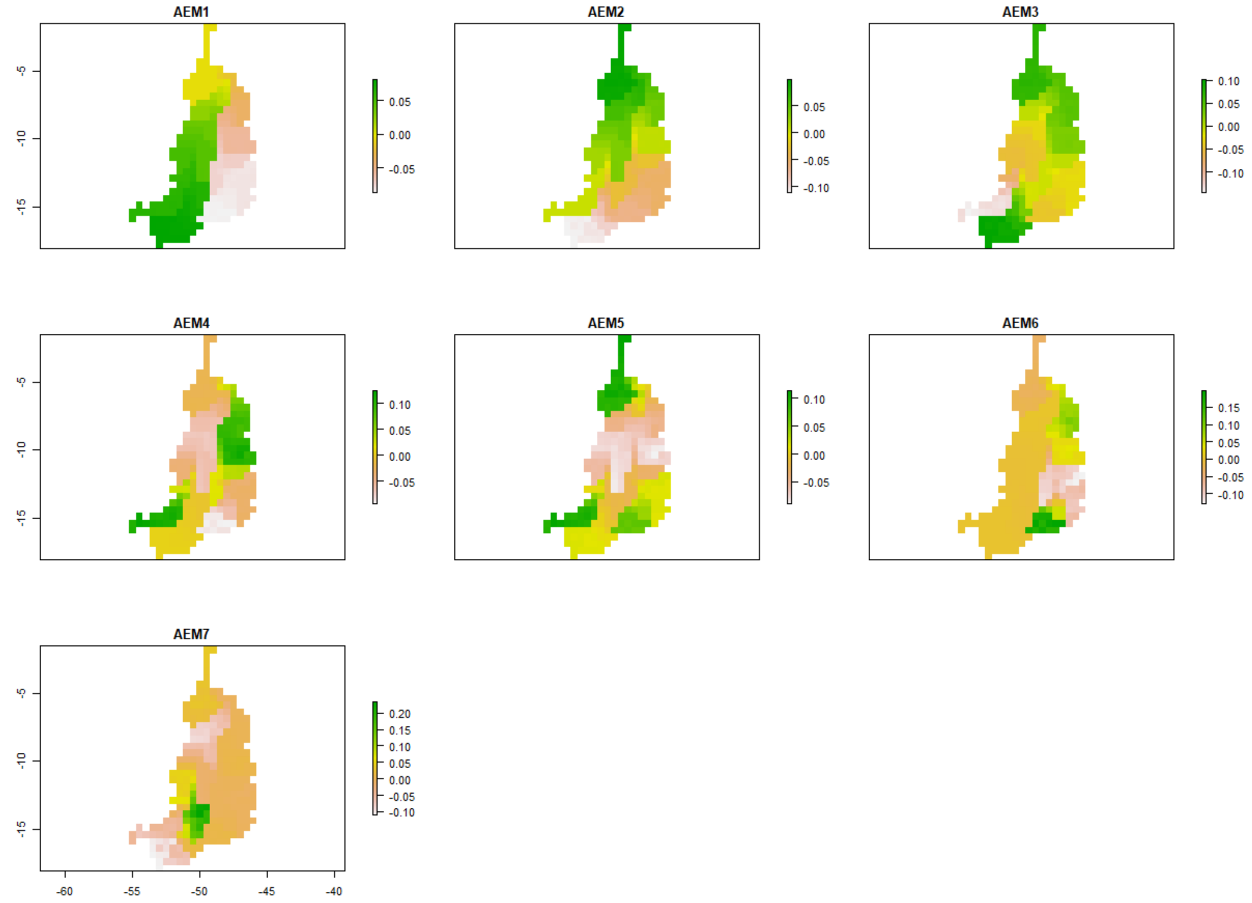


Figure S3 – AEM filters selected for SDM construction.

7. References

Allouche, O., Steinitz, O., Rotem, D., Rosenfeld, A., Kadmon, R., 2008. Incorporating distance constraints into species distribution models. J. Appl. Ecol. 45, 599–609. https://doi.org/10.1111/j.1365-2664.2007.01445.x

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Anderson, R.P., 2013. A framework for using niche models to estimate impacts of climate change on species distributions. Ann. N. Y. Acad. Sci. 1297, 8–28. https://doi.org/10.1111/nyas.12264

Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93, 1527–1539. https://doi.org/10.1890/11-1930.1

Bahn, V., McGill, B.J., 2007. Can niche-based distribution models outperform spatial interpolation? Glob. Ecol. Biogeogr. 16, 733–742. https://doi.org/10.1111/j.1466-8238.2007.00331.x

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. Ecol. Modell. 222, 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Blanchet, F.G., Legendre, P., Borcard, D., 2008. Modelling directional spatial processes in ecological data. Ecol. Modell. 215, 325–336. https://doi.org/10.1016/j.ecolmodel.2008.04.001

Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Chang. Biol. 16, 1145–1157. https://doi.org/10.1111/j.1365-2486.2009.02000.x

Chaine, A.S., Clobert, J., 2012. Dispersal, in: Behavioural Responses to a Changing World. Oxford University Press, pp. 63–79. https://doi.org/10.1093/acprof:osobl/9780199602568.003.0005

Clobert, J., Ims, R.A., Rousset, F., 2004. Causes, mechanisms and consequences of dispersal, in: Ecology, Genetics and Evolution of Metapopulations. Elsevier, pp. 307–335.

De Marco, P., Diniz-Filho, J.A.F., Bini, L.M., 2008. Spatial analysis improves species distribution modelling during range expansion. Biol. Lett. 4, 577–580. https://doi.org/10.1098/rsbl.2008.0210

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2019. adespatial: Multivariate Multiscale Spatial Analysis.

Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? on finding reasons for differing performances of species distribution models. Ecography (Cop.). 32, 66–77. https://doi.org/10.1111/j.1600-0587.2008.05505.x

Franklin, J., Wejnert, K.E., Hathaway, S.A., Rochester, C.J., Fisher, R.N., 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. Divers. Distrib. 15, 167–177. https://doi.org/10.1111/j.1472-4642.2008.00536.x

Guélat, J., Kéry, M., 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. Methods Ecol. Evol. 9, 1614–1625. https://doi.org/10.1111/2041-210X.12983

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography (Cop.). 29, 773–785. https://doi.org/10.1111/j.0906-7590.2006.04700.x

Hijmans, R.J., 2019. raster: Geographic Data Analysis and Modeling.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R.J., Phillips, S.J., Leathwick, J.R., Hortal, J., 2016. dismo: Species Distribution

Modeling. R package version 1.1-1.

Holloway, P., Miller, J.A., Gillings, S., 2016. Incorporating movement in species distribution models: how do simulations of dispersal affect the accuracy and uncertainty of projections? Int. J. Geogr. Inf. Sci. 30, 1–25. https://doi.org/10.1080/13658816.2016.1158823

Jackson, D.A., 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology 74, 2204–2214. https://doi.org/10.2307/1939574

Jiménez-Valverde, A., Lobo, J., Hortal, J., 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Community Ecol. 10, 196–205. https://doi.org/10.1556/ComEc.10.2009.2.9

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

Liu, C., Newell, G., White, M., 2019. The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. Ecography (Cop.). 42, 535–548. https://doi.org/10.1111/ecog.03188

Mendes, P., Velazco, S.J.E., Andrade, A.F.A. de, De Marco, P., 2020. Dealing with overprediction in species distribution models: How adding distance constraints can improve model accuracy. Ecol. Modell. 431, 109180. https://doi.org/10.1016/j.ecolmodel.2020.109180

Miller, J.A., Holloway, P., 2015. Incorporating movement in species distribution models. Prog. Phys. Geogr. 39, 837–849. https://doi.org/10.1177/0309133315580890

Monsimet, J., Devineau, O., Pétillon, J., Lafage, D., 2020. Explicit integration of dispersal-related metrics improves predictions of SDM in predatory arthropods. Sci. Rep. 10, 16668. https://doi.org/10.1038/s41598-020-73262-2

Padial, A.A., Ceschin, F., Declerck, S.A.J., De Meester, L., Bonecker, C.C., Lansac-Tôha, F.A., Rodrigues, L., Rodrigues, L.C., Train, S., Velho, L.F.M., Bini, L.M., 2014. Dispersal Ability Determines the Role of Environmental, Spatial and Temporal Drivers of Metacommunity Structure. PLoS One 9, e111227. https://doi.org/10.1371/journal.pone.0111227

Perrin, S.W., Englund, G., Blumentrath, S., Brian O'Hara, R., Amundsen, P., Gravbrøt Finstad, A., 2020. Integrating dispersal along freshwater ecosystems into species distribution models. Divers. Distrib. 26, 1598–1611. https://doi.org/10.1111/ddi.13112

Peterson, A.T., Soberón, J., 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. Nat. Conserv. 10, 102–107. https://doi.org/10.4322/natcon.2012.019

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. Ecological Niches and Geographic Distributions. Princenton University Press, Princenton.

Revelle, W., 2019. psych: Procedures for Personality and Psychological Research.

Segurado, P., Araújo, M.B., Kunin, W.E., 2006. Consequences of spatial autocorrelation for niche-based models. J. Appl. Ecol. 43, 433–444. https://doi.org/10.1111/j.1365-2664.2006.01162.x

Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecol. Lett. 10, 1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Soberon, J., Peterson, A.T., 2005. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. Biodivers. Informatics 2, 1–10. https://doi.org/10.17161/bi.v2i0.4

Soliman, T., Mourits, M.C.M., van der Werf, W., Hengeveld, G.M., Robinet, C., Lansink, A.G.J.M.O., 2012. Framework for Modelling Economic Impacts of Invasive Species, Applied to Pine Wood Nematode in Europe. PLoS One 7, e45505. https://doi.org/10.1371/journal.pone.0045505

Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Modell. 148, 1–13. https://doi.org/10.1016/S0304-3800(01)00388-X

Tonkin, J.D., Altermatt, F., Finn, D.S., Heino, J., Olden, J.D., Pauls, S.U., Lytle, D.A., 2018. The role of dispersal in river network metacommunities: Patterns, processes, and pathways. Freshw. Biol. 63, 141–163. https://doi.org/10.1111/fwb.13037

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Zimmermann, N.E., 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14, 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

CAPÍTULO 3: Hydropower plants as dispersal barriers in freshwater species distribution models: using restrictions through asymmetrical dispersal predictors

Micael Rosa Parreira[1*], João Carlos Nabout[2]

[1]Universidade Federal de Goiás, Campus Samambaia, 74690-900, Goiânia, GO, Brazil.

[2]Universidade Estadual de Goiás, Campus Central, 75132-903, Anápolis, GO, Brazil.

[*]Corresponding author: Micael Rosa Parreira, Universidade Federal de Goiás, Campus Samambaia, Av. Esperança, s/n - Chácaras de Recreio Samambaia, 74690-900, Goiânia, GO, Brazil. Email: micael_rp@hotmail.com

Abstract

Hydropower plants represent one of the greatest threats to freshwater fish by fragmenting the habitat and avoiding species dispersal. This type of dispersal barrier is often disregarded when predicting freshwater species distribution due to the complexity of inserting the species dispersal routes, and thus the barriers, into the models. Here, we evaluate the impact of hydroelectric dams in species distribution models through asymmetrical dispersal predictors on the predicted geographic distribution of freshwater fish species. For this, we used asymmetrical dispersal (i.e., AEM) as predictors for modeling the distribution of 29 native fish species in the Tocantins-Araguaia basin. After that, we included the hydropower plant (HPP) location into the asymmetrical binary matrix for the AEM construction by removing the connections where the HPP is located, representing the downstream disconnection a dam causes in the fish species dispersal route. Besides having higher predicted accuracy, the models using the HPP information generated more realistic predictions, avoiding overpredictions to areas suitable but limited to species dispersal due to an anthropic barrier. Furthermore, the predictions including HPPs showed higher loss of species richness and nestedness (i.e., loss of species instead of replacement), especially for the southeastern area, which concentrates most planned and built HPPs. Therefore, using dispersal constraints in species distribution models increases the reliability of the predictions by avoiding overpredictions based on the premise of complete access by the species to any area that is climatically suitable, regardless of dispersal barriers or capacity. In conclusion, in this study, we use a novel method of including dispersal constraints into distribution models through *a priori* insertion of their location within the asymmetrical dispersal predictors, avoiding *a posteriori* adjustment of the predicted distribution.

Keywords: Asymmetric Eigenvectors Maps; Freshwater habitats; Hydropower plants; Movement constraints; Species Distribution Models.

1. Introduction

Freshwater environments represent only 0.1% of the Earth's surface, yet they are home to about 10% of all animal species (Balian et al., 2008). Among these species, it is estimated that the fish represent about 32,500 species, having a global distribution, whether in freshwater or saltwater environments (Nelson et al., 2016). Freshwater ecosystems are highly subject to human interference, which has led to habitat degradation, pollution, change in river courses, water extraction, overexploitation in fishing, and the introduction of invasive species (Pfauserová et al., 2022; Strayer and Dudgeon, 2010). Among the anthropic impacts, the construction of dams stands out (Anderson et al., 2006; Liermann et al., 2012), representing one of the greatest threats to the diversity of aquatic species (Vörösmarty et al., 2010). While providing direct economic benefits (e.g., water security and renewable energy) (Frey and Linke, 2002; Kaygusuz, 2016), dams obstruct fish migration, fragment habitat, and reduce nutrient transport between rivers (Hall et al., 2011; Kuriqi et al., 2021), thus, affecting the species distribution (Nilsson, 1997; Pringle et al., 2000).

In Brazil, the use of dams for electricity production stands out as the most used power source nationwide. Among the types of dams used in Brazil (e.g., PCH, UHE, CGH; ANEEL 2020), hydroelectric power plants (UHEs) alone supply more than 56% of all electricity used (ANEEL 2020), which are larger and have more potency, demanding larger fast-moving rivers, and consequently, having a more significant ecological impact (Anderson et al., 2020; Vörösmarty et al., 2010). The demand for energy consumption in Brazil grows by about 4.5% per year, which is why there is an expectation for implementing new hydroelectric plants in the short and medium term. According to ANEEL (National Electric Energy Agency), Brazil has a total of 4,511 enterprises in operation, 213 power enterprises currently under construction, and 627 planned for the coming years (ANEEL, 2020). In this sense, the Brazilian energy production is highly dependable on HPPs, which despite being considered a renewable and sustainable type of energy (Frey and Linke, 2002; Kaygusuz, 2016). It has many ecological impacts, such as altering the natural flow regime and reducing the habitat connectivity, which affect, among other, fish migration (Kuriqi et al., 2021).

Knowing that there are still many uncatalogued fish species in the world (Nelson et al., 2016) and many still have their distribution poorly known (Carneiro et al., 2016), it is difficult to show which species can be affected by dams, mainly due to the lack of knowledge about their distribution (Bini et al., 2006; Whittaker et al., 2005). In this sense, species distribution models (SDMs) can represent an essential tool for assessing these effects (Elith and Leathwick, 2009). SDMs are based on the use of computational algorithms to predict the distribution of species according to their realized niches, using covariates (e.g., climatic variables) and data on the presence/absence of species to search for areas with suitable conditions for their occurrence (Peterson et al., 2011). However, the construction of SDMs for freshwater aquatic species is limited by the lack of specific freshwater environmental variables (Domisch et al., 2015a), variables that allow the insertion of directional effects of fish dispersion (Blanchet et al., 2008), and dispersal constraints (e.g., HPPs) *a priori* during modeling, through asymmetric dispersal variables.

In this sense, we highlight the use of AEMs (Asymmetric Eigenvector Maps; Blanchet et al. 2008) as predictors of asymmetric dispersion to be used in SDMs as dispersal variables that, unlike traditional environmental variables (e.g., bioclimatic variables), allow the insertion of directional dispersal constraints as flow disconnections at hydroelectric installation sites. These disconnections are inserted into a directional binary matrix based on the existing hydrographic network used to generate the AEM eigenvectors, thus representing exactly the disconnections due to HPPs. Therefore, in this study, we aim to evaluate the impact of the insertion of hydroelectric dams on the potential geographic distribution of fish species, verifying this effect from restrictions inserted in the binary hydrographic network of the asymmetric dispersal predictors (AEM).

2. Materials and Methods

2.1. Hydroelectric Power Plants

The hydroelectric power plant (HPP) data was downloaded on the ANEEL’s (Agência Nacional de Energia Elétrica) SIGEL (Sistema de Informações Geográficas do Setor Elétrico) website (sigel.aneel.gov.br). This GIS (Geographic Information System) database has geographic information regarding the Brazilian Department of Energy of the Ministry of Mines and Energy

(MME). Specifically, this database contains maps of types of hydropower plants used in Brazil, wind power, and transmission lines, among others. Besides, the shapefiles with updated information on Brazilian power plants and their location are available to download. Furthermore, more information on all ongoing Brazilian energy enterprises allocated into different categories (e.g., enterprises per state and sub-basin, types of enterprises) can be found on the ANNEL's SIGA (Sistema de Informações de Geração da ANEEL) website (aneel.gov.br/siga).

For our study, we obtained all information on UHEs (Usinas Hidrelétricas), hereafter called HPPs, available in both ANNEL's databases. We further categorized those HPPs into built and planned HPPs. The built HPPs correspond to the energy enterprises with finished construction and operating, while planned HPPs are those already approved for construction or with ongoing construction, as well as those still pending for approval but planned for this decade. The latter HPP information is available in the PDEE (Plano Decenal de Expansão de Energia) documents, which are the MME's decennial energy expansion plans updated every year, available now from 2022-2031 on the EPE (Empresa de Pesquisa Energética) website (epe.gov.br). Each PDEE document contains a list of planned HPP for next years that may or may not be already available in the SIGA and SIGEL databases with the enterprise information and location. For those HPPs without information in those databases, we obtained the concession agreements for the companies responsible for building and managing the enterprises, which contain, among other information, the dam's geographic coordinates, predicted energy production, and river and city locations. Information about the built and planned hydropower plants (HPP) is available in Table S2.

2.2. Study area

The study area used for modeling is the Tocantins-Araguaia River basin, covering the entire hydrographic network. This basin has two major rivers (Tocantins and Araguaia), forming two sub-basins that merge at one point up to the north region, close to the mouth (Figure 1). Therefore, a species occurring exclusively in one sub-basin must disperse a long way through the main river course to occupy the other sub-basin. Thus, this basin is an interesting area for studies on dispersal limitation because it shows the species' effort to disperse between the two basins (Tocantins and Araguaia) via downstream connectivity. The entire hydrographic network was rasterized into grid cells of 0.5º spatial resolution (latitude and longitude), totaling 282 cells for the Tocantins-Araguaia basin.

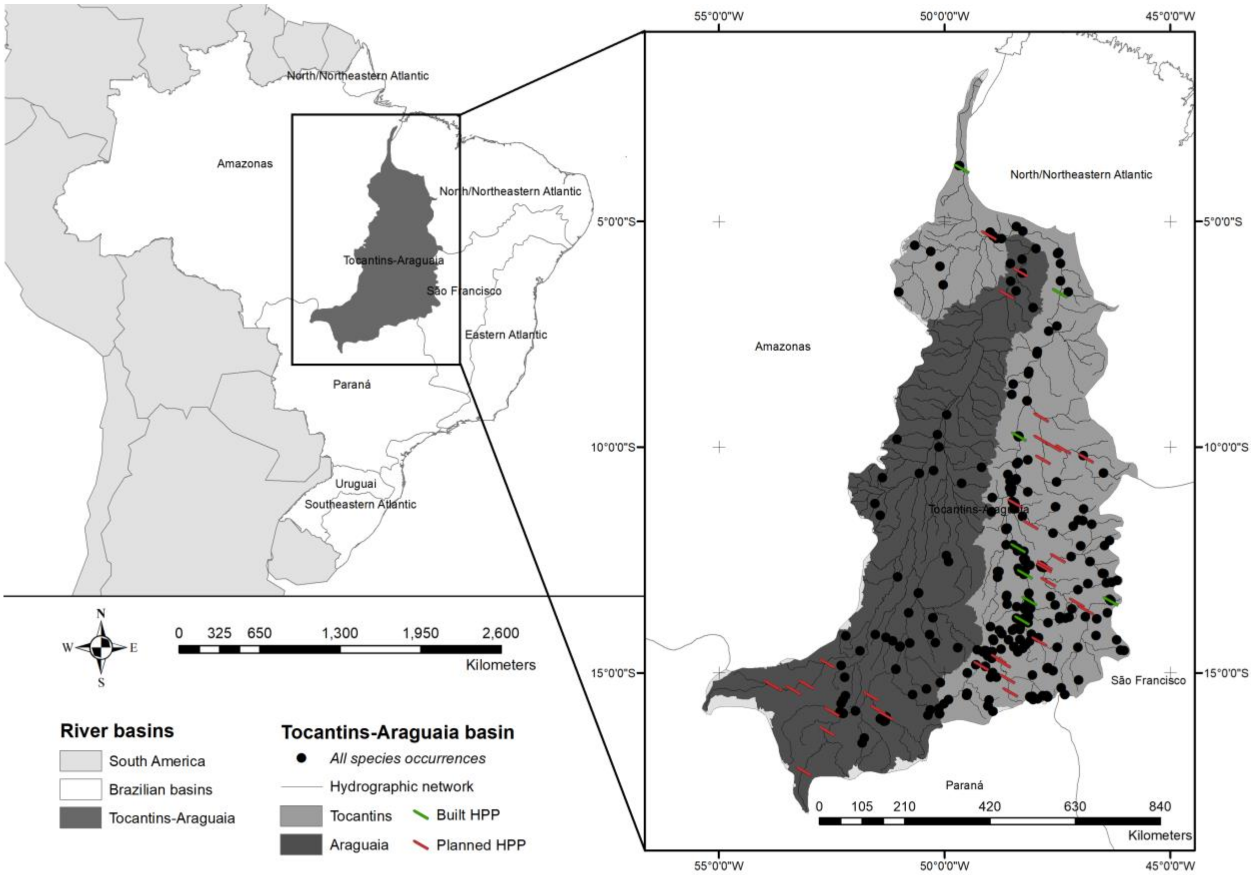


Figure 1 –Study area of the Tocantins-Araguaia River basin together with all species occurrences (black dots) and built and planned HPP locations (green and red dashes, respectively). Each species distribution is available in Figure S1.

2.3. Species occurrences

We selected 29 freshwater species native to the Tocantins-Araguaia River basin. The fish occurrence records were obtained in the online databases *Species Link* (specieslink.net), *GBIF* (gbif.org), and *FishBase* (fishbase.org). These species were chosen based on their distributions being restricted to the studied basin and having more than five unique occurrences (non-duplicated within grid cells). In this sense, for each species, we removed duplicated occurrences (more than one coordinate within cells) by addressing the cell id to each coordinate that falls within and then removing the points with duplicated cells ids, i.e., ending up with unique occurrences (only one occurrence per cell). These unique occurrence records (coordinates) for each species were later used to generate the models.

2.4. Asymmetric dispersal variables

The asymmetric eigenvector maps (AEM) variables (hereafter called asymmetric dispersal variables) were generated using the function *aem* of the *adespatial* (Dray et al., 2019) available in the R software. The calculation of these variables in this function is based on the Singular Value Decomposition (SVD) analysis using an asymmetric edge-by-sites binary matrix. In our study, this matrix is represented by a binary matrix of directional connectivity (asymmetric) using the grid cells as sites with nodes between grid cells and tributaries, besides the edge as the river path, until it reaches a riverbed. The first edge (E1) represents the longest river path until it reaches the farthest riverbed. The subsequent edge is the same as the previous minus the first node (grid cell), only changing when it connects to a lower-order river tributary, which creates one or more edges (depending on tributary size and sub-connections) exclusively to this river branching. After finishing every tributary edge(s), the edges counting return to the primary connection (river), receiving 1s for the remaining grid-cells (nodes) in the edge and 0s for the nodes already included in previous edges until it reaches the farthest node (grid cell with the farthest source) (see Parreira et al. 2021 [submitted] for more explanation on the matrix's construction). The construction of this binary matrix is based on (Blanchet et al., 2008) modeling directional spatial processes in freshwater environments. Posteriorly, we used the broken stick method (Jackson, 1993) to determine the number of non-collinear asymmetric eigenvectors (AEM). These selected axes (eigenvectors) were then used as asymmetrical dispersal predictors for generating the species distribution models.

2.5. Asymmetric dispersal variables with HPP restrictions

The construction of AEM variables accounting for hydropower plants (HPP) restrictions in the hydrographic network followed the same steps of building the directional binary matrix for the AEM variables without restrictions. The difference arises after the binary matrix is built, where disconnections are inserted in the grid cells (nodes) where the built and planned HPP are located. For this, we replace all connections (1s) for disconnections (0s) in the edges of the node (matrix line), where there is an HPP for all connections until it reaches the specific edge (matrix column) to which the HPP node continues the main connection (main rivers) or branches into new connections (tributary rivers). Therefore, this disconnection method represents the downstream disconnection a dam causes in the fish species dispersal route.

2.6. Modeling Procedures

Species distribution models (SDMs) were built using 75% of presence and absence points for training (model's construction) and 25% for testing the models' performance (Guisan and Zimmermann, 2000), randomly selected from each species' unique presences and the 56 randomly generated pseudo-absences. The set of presence and absence points selected for training (constructing) were used together with modeling methods (algorithms) and sets of asymmetrical dispersal variables. We used six algorithms: Bioclim (Nix, 1986), Domain (Carpenter et al., 1993), Support Vector Machines (SVM; Schölkopf et al., 2001), Generalized Linear Models (GLM; Nelder and Wedderburn, 1972), Maximum Entropy (MaxEnt; Phillips et al., 2006), and Random Forest (Breiman, 2001). The algorithms used are available in the R package *dismo* (Hijmans et al., 2016) and were chosen because they consider different statistical methods (climatic envelopes, environmental distances, machine-learning, regressions) (Rangel and Loyola, 2012). In addition to algorithms, the models were built using only asymmetric dispersal (AEM) predictors and asymmetric dispersal predictors with HPP restrictions (AEM+HPP) built in the directional binary matrix. The model's training and validation occurred 20 times through cross-validation (using the selected presences and pseudo-absences) for all combinations of algorithms and types of variables for each species. In this sense, we built 240 models (20 repetitions x 6 algorithms x 2 sets of variables) for each of the 29 species, totaling 6,960 models.

All models generated were evaluated using the Area Under the "receiver operating characteristic" Curve (AUC; Swets, 1988), which is an evaluation metric threshold-independent (the limit for determination of presences/absences) that compare predict with observed values. We also used the True Skill Statistic (TSS; Allouche et al., 2006), which compares the number of correct predictions minus those attributable to random guessing, not affected by the prevalence and size of the validation set. The final potential distribution map is a consensus map (ensemble) built using the mean suitability values of all models with AUC > 0.75 weighted by the AUC values for each species. The models were generated and evaluated using functions from the *dismo* package (Hijmans et al., 2016) available in the R software (R Core Team, 2019).

2.7. Delta species richness

We predicted the potential geographical distribution of fish species using AEM predictors without considering HPP and AEM considering HPP restrictions directional binary matrix. Therefore, we measured the species richness before and after including the HPP restrictions to evaluate the loss, stability, or gain of species richness when the HPP effect as dispersal barriers is included in the SDMs. For this, we transformed the suitability distribution of each species ensemble forecast into a binary distribution of species (presence=1, absence=0), using the mean suitability values of presence points of each species as a threshold (Liu et al., 2005). We calculated the species richness of each grid cell by summing all species presences (1s) in each cell. Thus, we ended up with two raster maps of species richness, one for AEM only and another for AEM with HPP restrictions. Finally, we calculated the delta richness by subtracting the species richness of each grid cell of the AEM with the HPP raster map by the correspondent grid cell of the AEM-only raster map. Therefore, grid cells with negative values represent a loss of richness. In contrast, positive values represent a gain of richness, and grid cells with 0 represent stability (i.e., no loss or gain of richness). In this sense, we calculated the percentage of loss, gain, and stability for the basin area by dividing the number of grid cells in each category by the total number of grid cells of the basin area, multiplying by 100.

Although the delta analysis of species richness helped us understand the losses and gains in the number of species, it does not consider the possible exchanges of species when including HPP restrictions in the stability grid cells. To overcome this problem, we performed analyses regarding the composition of species richness, i.e., we quantified the beta diversity and its components (turnover and nestedness).

2.8. Beta (turnover and nestedness)

We evaluated the possible replacements (turnover) and losses (nestedness) of species between the modeled distribution without HPP restrictions and with HPP restrictions for the species dispersal routes. For this, we estimated the beta dissimilarity (*βsor*) between AEM and AEM+HPP species richness using Eq. 1, adapted for SDM components from (Baselga, 2010)

$$\beta sor = \frac{b+c}{2a+b+c} \tag{1}$$

Where a is the number of species common to both distributions, b is the number of species that occur in the AEM map but not in the AEM+HPP map, and c is the number of species that occur in the AEM+HPP map but not in the AEM-based distribution map.

Posteriorly, we partitioned *βsor* into *βsim* (Turnover) and *βnes* (Nestedness) (Eqs. 2 and 3).

$$\beta sim = \frac{\min(b,c)}{a + \min(b,c)} \tag{2}$$

$$\begin{aligned} \beta nes = \beta sor + \beta sim &= \frac{b+c}{2a+b+c} - \frac{\min(b,c)}{a + \min(b,c)} \\ &= \frac{\max(b,c) - \min(b,c)}{2a + \min(b,c) + \max(b,c)} \, x \, \frac{a}{a + \min(b,c)} \end{aligned} \tag{3}$$

Where, max (*b,c*) and min (*b,c*) represent, respectively, the highest and lowest values between coefficients b and c.

Because *βsor* represents the dissimilarity of beta diversity, i.e., occurring variations, *βsor* = 0 represents the similarity between species richness of AEM and AEM+HPP distributions, with no losses, replacements, or gains. The highest possible βsor value is 1, meaning the species found in a location will not be found in the same location if considering HPP, whether due to the replacement of all species by others or the loss of all species. The values of *βnes* > 0 represent the loss of species in the same area when considering HPPs. *βsim* > 0, on the other hand, corresponds to the species replacement between AEM and AEM+HPP distributions. *βnes* equal to *βsor* and different from 0 means that a species will be added in a given location without losing any species already available there without the interference of HPP restrictions.

3. Results

Generally, most species modeled using only AEM had acceptable models with mean AUC = 0.75 and mean TSS = 0.62. On the contrary, when modeled using AEM with HPP restrictions, we noticed a significant decrease in the model's accuracy, with mean AUC = 0.69 and TSS = 0.54, especially for AUC, where most species performed below the acceptance threshold of 0.75 (n=15) (Table S1). Besides, the predicted distribution (ensemble maps) of the species showed that adding the HPP disconnection into the AEM predictors causes many disconnection patches in the predicted suitability, especially for the specific locations of HPPs (Figures 2 and 3).

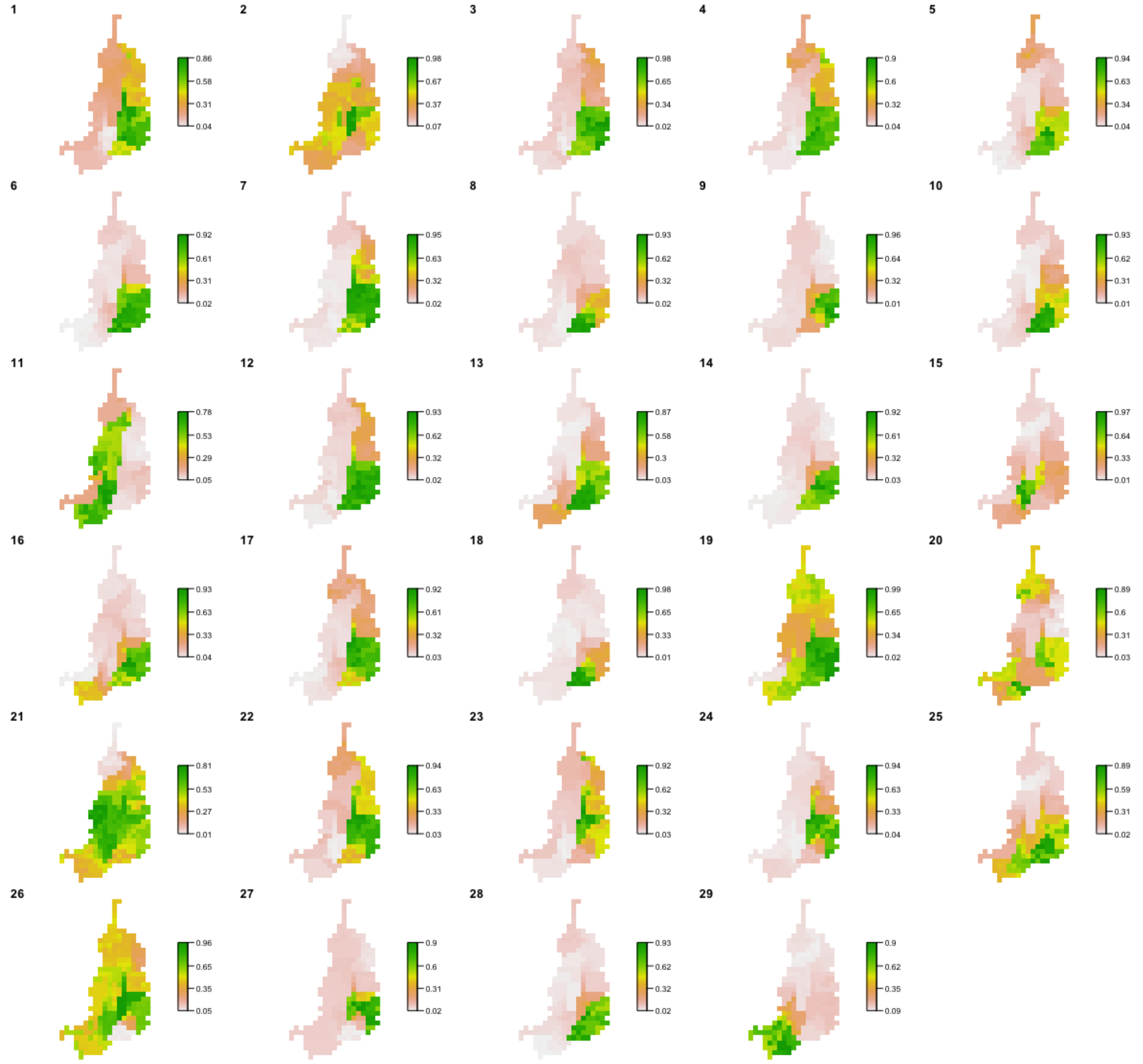


Figure 2 – Ensemble maps of each species modeled using only AEM predictors. Figure numbers correspond to the species ID available in Table S1.

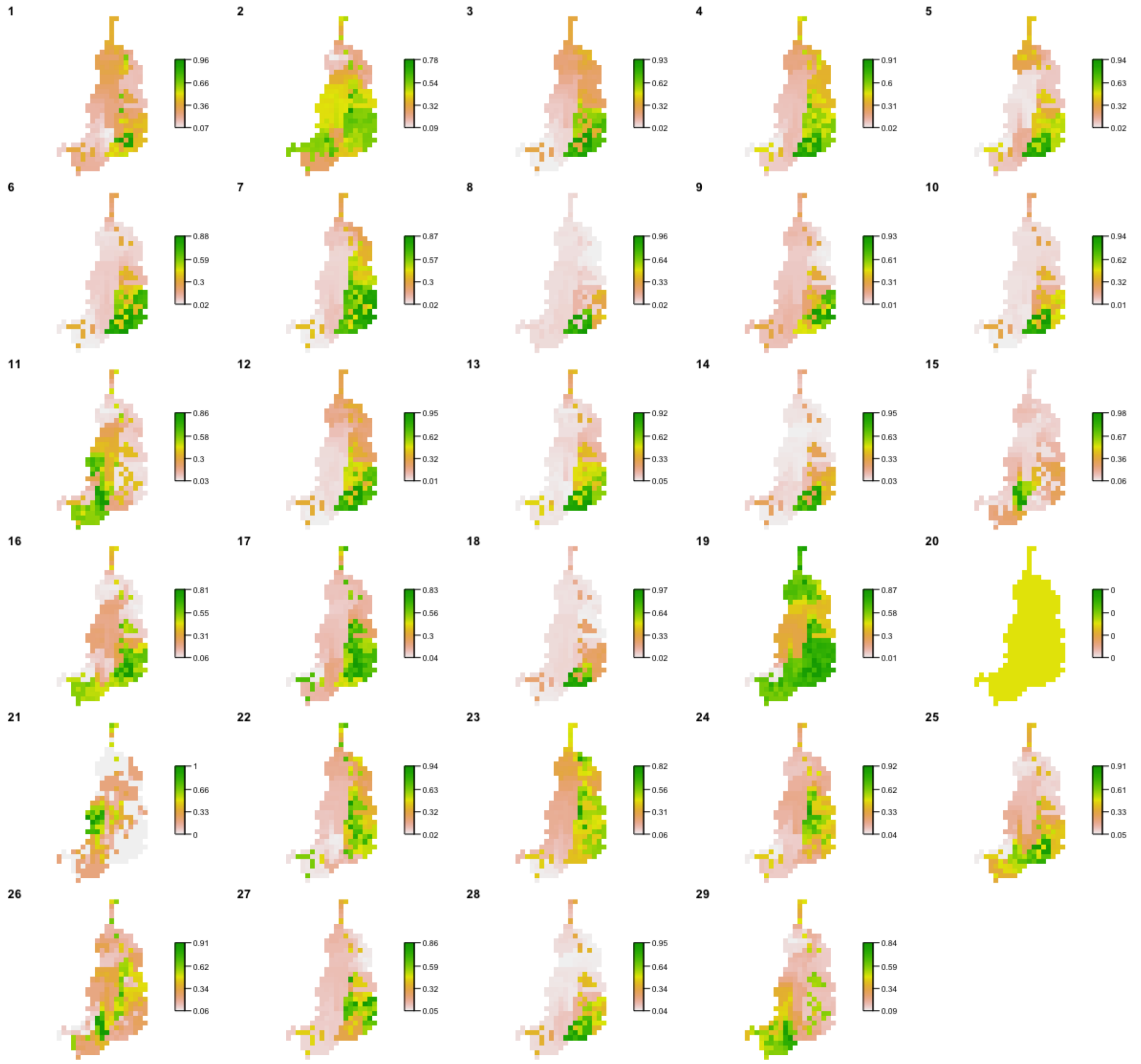


Figure 3 – Ensemble maps of each species modeled using AEM predictors with HPP restrictions built in. *Loricata lata* (20) did not have any model with AUC = 0.75 (Ensemble selection threshold). Figure numbers correspond to the species ID available in Table S1.

The species richness was highly affected by the insertion of HPP restrictions in the SDMs. It is possible to visually notice many areas in the Tocantins-Araguaia River basin with a high decrease of richness, where some areas, such as the middle portion of the Araguaia sub-basin, where the species were predicted extinct. Besides, the delta richness map showed the most loss of richness predicted for the upper Tocantins sub-basin, where there is a high species richness and a significant number of built and planned hydropower plants (Figures 1 and 4). Furthermore, the delta analysis showed that most areas (54%) would not have changed species richness (i.e., Stability). However, most areas with changes in species richness will lose species (30%), with only 15% of the areas gaining species richness after considering the HPP in the SDM analysis.

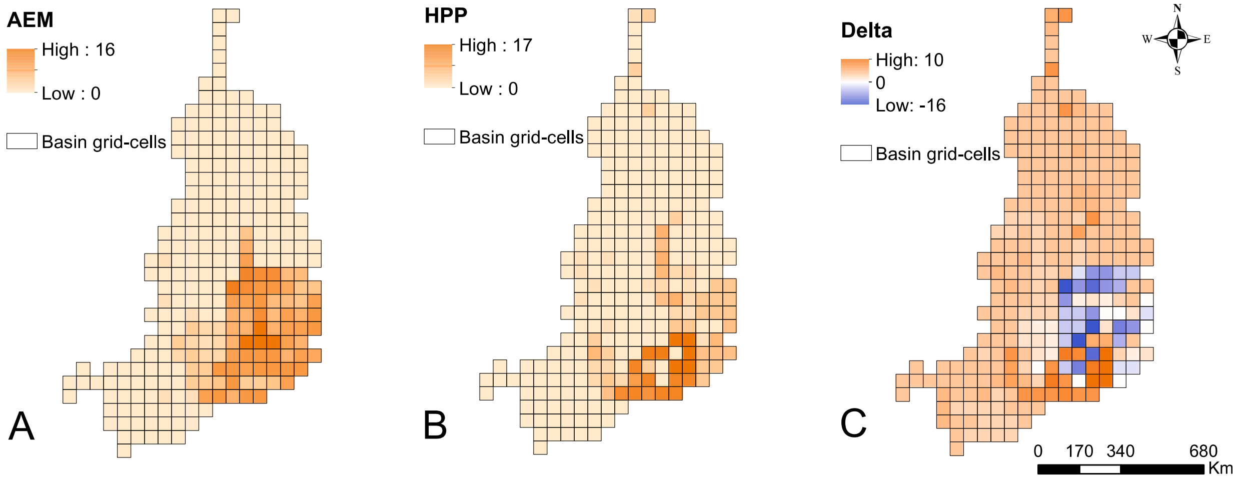


Figure 4 – Richness of species modeled using only AEM (A), using AEM with HPP restrictions built in (B), and Delta richness between both (C).

The beta diversity (i.e., change in species composition) analysis partitioned into turnover (i.e., species replacement) and nestedness (i.e., species loss) showed that most of the change in species composition was explained by nestedness, especially in the southeastern areas, meaning that the species in that location are more likely to disappear due to HPP establishment. The beta diversity in some southern areas (grid-cells) was explained by turnover, which means that the species in those areas will likely be replaced by other species that will be suitable for that area when considering the effect of HPPs (Figure 5).

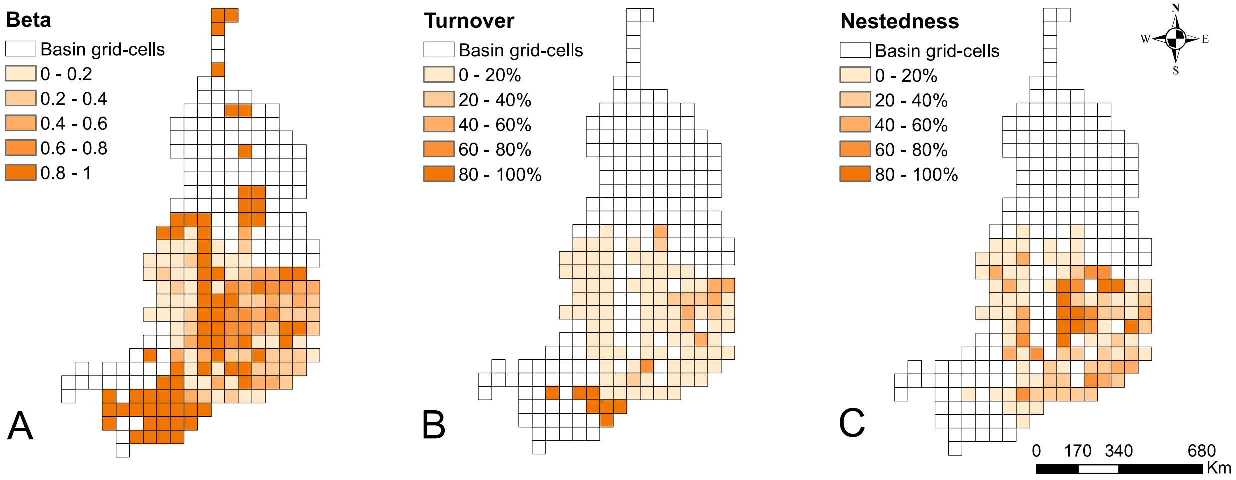


Figure 5 – Total beta (A; Median = 0.57) and partitioned into turnover (B; Median = 0) and nestedness (C; Median = 0.15), accounting for changes in species composition predicted by AEM predictors only and AEM predictors with HPP restrictions built in.

4. Discussion

Using AEM predictors to model the distribution of the native fish species allowed the insertion of dispersal constraints. Dam barriers (HPPs) could not be inserted using traditional climatic-based predictors in this case. Besides, by using those predictors with the barriers, the models had a slightly worse prediction performance and, most importantly, showed the disconnection pathways in the prediction. This result shows that the model predictions using HPPs better represented the scenario of river disconnection through anthropic barriers the species must undergo while trying to migrate to suitable areas (Barbarossa et al., 2020).

The use of AEM-based models itself already generated models with less overprediction to areas that could be climatically suitable for their survival but that are disconnected between sub-basins barriers (e.g., Parreira et al. 2022 in submission). Generally, species distribution models have some assumptions that disregard some intrinsic characteristics of the species or the environment where it inhabits, such as competition or resources (i.e., biotic interactions)(Anderson, 2017; Wisz et al., 2013) and connectivity and dispersal constraints (i.e., movement) (Vasudev et al., 2015). A way to overcome those problems is by adding species- or habitat-specific data either *a priori* (e.g., as predictors) or *a posteriori* (e.g., thresholds of dispersal) (Mendes et al., 2020). In our case, we use the hydropower plants' (HPP) barrier locations into the connectivity matrix of the AEM predictors as a dispersal constraint to be used when modelling the fish species potential distribution for the basin area.

The fish species had decreased suitability when modeled using the HPP barriers during modeling compared to using only AEM predictors without accounting for the anthropic barriers. The HPP models showed disconnections (lower or no suitability) within highly suitable areas for the species' survival, as shown by the AEM-based models. Hydropower plant dams act as dispersal barriers for freshwater species (Anderson et al., 2020), especially migratory species such as rheophilic fish, which use the fast-moving waters to migrate to different habitats seeking suitable areas for food resources and nurseries, for example (Stoffers et al., 2022). Therefore, using those barriers as dispersal constraints in SDM predictions better represent the species' potential distribution, avoiding overpredictions to areas outside their reach due to anthropic dispersal limitation.

The species richness maps showed that the basin's southeastern area holds the highest concentration of fish species, which is also the area with the most current and planned HPPs. Besides, the insertion of HPPs in the SDM projections showed a significant decrease in species richness for the basin area, especially for the downstream area (larger rivers) of the southeastern area. Many studies have shown that dam obstruction has several implications for freshwater fish biodiversity, including the loss of native species and the increase in exotic species favored in degraded habitats (Liermann et al., 2012). Besides, even though there are some hydropower plant models that, in theory, may allow the species to migrate through the dam barrier, such as HPP with diversion channel (Habit et al., 2007; Zhao et al., 2018) or ladderways (i.e., fish ladder) (Gowans et al., 1999; Takahashi et al., 2016). They can still cause significant changes (in the fish community through homogenization (e.g., species loss), especially in high-head HPPs, as in this study. Therefore, inserting the HPP locations as a directional effect for dispersal limitation can represent the effect of these constraints by reducing the predicted suitability and, as a consequence, the species richness for the HPP areas and the adjacent and downstream pathways.

Our results on the beta diversity analysis showed that most of the beta change (i.e., change in species composition) occurred in the southeastern region of the basin. They were mostly explained by nestedness, i.e., loss of current species in the areas instead of replacement of current species by new ones (i.e., turnover). This result corroborates the species richness loss shown for the same area with the insertion of HPPs in the modeled predictions. Many studies have used these analyses to show changes in species richness and composition from SDM predictions, especially underlying differences in predictions among present and future climate scenarios (RCPs) to show the effect of climate change on the species distribution (Duan et al., 2016; Ferreira et al., 2021). In our case, we show the differences in predictions between unconstrained SDMs using the asymmetrical dispersal predictors and SDMs with constrains (i.e., HPP barriers)o reduce species richness overpredictions (e.g., D’Amen et al., 2015) that do not account for the physical barriers. However, caution is necessary when evaluating species richness and composition from SDM predictions rather than observed data since they tend to obtain overestimated resultsompared to observed distribution (Toro et al., 2018).

In conclusion, we show a new method of inserting species dispersal constraints into *a priori* SMD modeling through the use of AEM predictors with the specific location of constraints (e.g., HPP barriers) into the grid cells of distribution maps that will act as constraints of predictions to areas outside the species reach or that may not be suitable if the barrier predicted is installed in place. Although the effect of hydropower plant dams is evident in how it affects the distribution of freshwater fish dispersed to suitable areas, other dispersal constraints could also be used within SMD predictors to improve the accuracy and reliability of predictions. These predictors include natural barriers (e.g., mountains) or known species dispersal range. In this study, by including the barriers as constraints in the SDM predictions, we showed that the predictions of freshwater fish distribution are biased if they do not account for dispersal constraints from natural or anthropic barriers. Furthermore, the species richness and composition were highly affected by the insertion of HPP dams in the AEM predictors since it informed disconnected areas in the species dispersal pathways that may affect their distribution to other suitable areas, leading to a loss in native species richness and diversity.

5. Acknowledgments

MRP thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – finance code 001) and FAPEG for the PhD scholarship received. Our work on SDMs has been continuously supported by different grants: CNPq, Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG/process 201710267000519), and National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation (MCTI/CNPq/FAPEG/465610/2014-5), and Brazilian Network on Global Climate Change Research (Rede CLIMA). JCN was supported by a CNPq productivity fellowship.

6. References

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Anderson, E.P., Freeman, M.C., Pringle, C.M., 2006. Ecological consequences of hydropower development in Central America: impacts of small dams and water diversion on neotropical stream fish assemblages. River Research and Applications 22, 397–411. https://doi.org/10.1002/rra.899

Anderson, R.L., Anderson, C.A., Larson, J.H., Knights, B., Vallazza, J., Jenkins, S.E., Lamer, J.T., 2020. Influence of a high-head dam as a dispersal barrier to fish community structure of the Upper Mississippi River. River Research and Applications 36, 47–56. https://doi.org/10.1002/rra.3534

Balian, E. V., Segers, H., Lévèque, C., Martens, K., 2008. The Freshwater Animal Diversity Assessment: An overview of the results. Hydrobiologia 595, 627–637. https://doi.org/10.1007/s10750-007-9246-3

Barbarossa, V., Schmitt, R.J.P., Huijbregts, M.A.J., Zarfl, C., King, H., Schipper, A.M., 2020. Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. Proc Natl Acad Sci U S A 117, 3648–3655. https://doi.org/10.1073/pnas.1912776117

Baselga, A., 2010. Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography 19, 134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F.L.V.B., Bastos, R.P., Pinto, M.P., 2006. Challenging Wallacean and Linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. Diversity and Distributions 12, 475–482. https://doi.org/10.1111/j.1366-9516.2006.00286.x

Blanchet, F.G., Legendre, P., Borcard, D., 2008. Modelling directional spatial processes in ecological data. Ecological Modelling 215, 325–336. https://doi.org/10.1016/j.ecolmodel.2008.04.001

Carneiro, L.R.D.A., Lima, A.P., Machado, R.B., Magnusson, W.E., 2016. Limitations to the use of species-distribution models for environmental-impact assessments in the Amazon. PLoS ONE 11, 1–17. https://doi.org/10.1371/journal.pone.0146543

D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L., Guisan, A., 2015. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. Journal of Biogeography 42, 1255–1266. https://doi.org/10.1111/jbi.12485

Domisch, S., Amatulli, G., Jetz, W., 2015. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data 2, 150073. https://doi.org/10.1038/sdata.2015.73

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2019. adespatial: Multivariate Multiscale Spatial Analysis.

Duan, R.Y., Kong, X.Q., Huang, M.Y., Varela, S., Ji, X., 2016. The potential effects of climate change on amphibian distribution, range fragmentation and turnover in China. PeerJ 2016, e2185. https://doi.org/10.7717/PEERJ.2185/SUPP-5

Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Ferreira, R.B., Parreira, M.R., Nabout, J.C., 2021. The impact of global climate change on the number and replacement of provisioning ecosystem services of Brazilian Cerrado plants. Environmental Monitoring and Assessment 193, 731. https://doi.org/10.1007/s10661-021-09529-6

Frey, G.W., Linke, D.M., 2002. Hydropower as a renewable and sustainable energy resource meeting global energy challenges in a reasonable way 30, 1261–1265. https://doi.org/10.1016/S0301-4215(02)00086-1

Gowans, A.R.D., Armstrong, J.D., Priede, I.G., 1999. Movements of adult Atlantic salmon in relation to a hydroelectric dam and fish ladder. Journal of Fish Biology 54, 713–726. https://doi.org/10.1111/j.1095-8649.1999.tb02028.x

Habit, E., Belk, M.C., Parra, O., 2007. Response of the riverine fish community to the construction and operation of a diversion hydropower plant in central Chile. Aquatic Conservation: Marine and Freshwater Ecosystems 17, 37–49. https://doi.org/10.1002/aqc.774

Hall, C.J., Jordaan, A., Frisk, M.G., 2011. The historic influence of dams on diadromous fish habitat with a focus on river herring and hydrologic longitudinal connectivity. Landscape Ecology 26, 95–107. https://doi.org/10.1007/s10980-010-9539-1

Hijmans, R.J., Phillips, S.J., Leathwick, J.R., Hortal, J., 2016. dismo: Species Distribution Modeling. R package version 1.1-1.

Jackson, D.A., 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology 74, 2204–2214. https://doi.org/10.2307/1939574

Kaygusuz, K., 2016. Hydropower as clean and renewable energy source for electricity generation. Journal of Engineering Research and Applied Science 5, 359–369.

Kuriqi, A., Pinheiro, A.N., Sordo-Ward, A., Bejarano, M.D., Garrote, L., 2021. Ecological impacts of run-of-river hydropower plants—Current status and future prospects on the brink of energy transition. Renewable and Sustainable Energy Reviews 142, 110833. https://doi.org/10.1016/J.RSER.2021.110833

Liermann, C.R., Nilsson, C., Robertson, J., Ng, R.Y., 2012. Implications of Dam Obstruction for Global Freshwater Fish Diversity. BioScience 62, 539–548. https://doi.org/10.1525/bio.2012.62.6.5

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

Mendes, P., Velazco, S.J.E., Andrade, A.F.A. de, De Marco, P., 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

Nelson, J.S., Grande, T.C., Wilson, M.V.H., 2016. Fishes of the World, 5th ed. John Wiley & Sons, Inc., Hoboken, New Jersey.

Nilsson, C., 1997. Long-Term Responses of River-Margin Vegetation to Water-Level Regulation. Science (1979) 276, 798–800. https://doi.org/10.1126/science.276.5313.798

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. Ecological Niches and Geographic Distributions. Princenton University Press, Princenton.

Pfauserová, N., Brabec, M., Slavík, O., Horký, P., Žlábek, V., Hladík, M., 2022. Effects of physical parameters on fish migration between a reservoir and its tributaries. Scientific Reports 12, 8612. https://doi.org/10.1038/s41598-022-12231-3

Pringle, C.M., Freeman, M.C., Freeman, B.J., 2000. Regional Effects of Hydrologic Alterations on Riverine Macrobiota in the New World: Tropical–Temperate Comparisons. BioScience 50, 807. https://doi.org/10.1641/0006-3568(2000)050[0807:REOHAO]2.0.CO;2

Stoffers, T., Buijse, A.D., Nagelkerke, L.A.J., Verreth, J.A.J., 2022. Environmental requirements and heterogeneity of rheophilic fish nursery habitats in European lowland rivers : Current insights and future challenges. Fish and Fisheries 23, 162–182. https://doi.org/10.1111/faf.12606

Strayer, D.L., Dudgeon, D., 2010. Freshwater biodiversity conservation: recent progress and future challenges. J North Am Benthol Soc 29, 344–358. https://doi.org/10.1899/08-171.1

Takahashi, Y., Rashid, M.H., Sarkar, A., Asaeda, T., Isono, M., Omura, T., Koga, K., 2016. Use of ladderways in fragmented habitat to aid the movement of Japanese giant salamander (Andrias japonicus). International Journal of River Basin Management 14, 233–241. https://doi.org/10.1080/15715124.2016.1159572

Toro, I. del, Ribbons, R.R., Hayward, J., Andersen, A.N., 2018. Are stacked species distribution models accurate at predicting multiple levels of diversity along a rainfall gradient? https://doi.org/10.1111/aec.12658

Vasudev, D., Fletcher, R.J., Goswami, V.R., Krishnadas, M., 2015. From dispersal constraints to landscape connectivity: lessons from species distribution modeling. Ecography 38, 967–978. https://doi.org/10.1111/ecog.01306

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 467, 555–561. https://doi.org/10.1038/nature09440

Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M., Willis, K.J., 2005. Conservation Biogeography: assessment and prospect. Diversity and Distributions 11, 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Zhao, T., Chen, S., Fu, C., Zhong, Q., 2018. Influence of diversion channel section type on landslide dam draining effect. Environmental Earth Sciences 77, 1–9. https://doi.org/10.1007/s12665-017-7217-1

Supplementary Material

Table S1 – Accuracy table with mean evaluation metrics for each species modeled using the two different types of variables.

ID	Species	AUC	TSS	Variable
1	*Acestrocephalus maculosus*	0.6763889	0.5011905	AEM
2	*Ammoglanis diaphanus*	0.3564484	0.1543651	AEM
3	*Ancistrus aguaboensis*	0.7789187	0.6099206	AEM
4	*Ancistrus minutus*	0.6935847	0.5198413	AEM
5	*Apareiodon cavalcante*	0.6418651	0.4940476	AEM
6	*Apteronotus camposdapazi*	0.8725694	0.7515873	AEM
7	*Aspidoras eurycephalus*	0.7593585	0.5719577	AEM
8	*Astyanax courensis*	0.9105159	0.8503968	AEM
9	*Astyanax unitaeniatus*	0.8411706	0.7404762	AEM
10	*Corumbataia tocantinensis*	0.8308036	0.6944444	AEM
11	*Corydoras maculifer*	0.5811508	0.4099206	AEM
12	*Creagrutus atrisignum*	0.789881	0.5989418	AEM
13	*Creagrutus saxatilis*	0.7654762	0.6460317	AEM
14	*Gymnotocinclus anosteos*	0.7261905	0.6297619	AEM
15	*Hemiancistrus cerrado*	0.8	0.702381	AEM
16	*Hemiancistrus micrommatos*	0.6785714	0.5198413	AEM
17	*Hemiancistrus spinosissimus*	0.7685847	0.621164	AEM
18	*Hypostomus asperatus*	0.8873677	0.7903439	AEM
19	*Hypostomus ericae*	0.6446032	0.3845238	AEM

20	*Loricaria lata*	0.3460317	0.1333333	AEM
21	*Moenkhausia pyrophthalma*	0.500248	0.2274802	AEM
22	*Moenkhausia tergimacula*	0.7201389	0.5269841	AEM
23	*Pimelodus joannis*	0.5373016	0.4253968	AEM
24	*Pimelodus stewarti*	0.6741071	0.502381	AEM
25	*Rineloricaria osvaldoi*	0.7611111	0.5850529	AEM
26	*Steindachnerina gracilis*	0.5118717	0.2583333	AEM
27	*Steindachnerina notograptos*	0.7535714	0.6515873	AEM
28	*Sternarchorhynchus mesensis*	0.8170635	0.743254	AEM
29	*Xenurobrycon coracoralinae*	0.5279762	0.4361111	AEM
1	*Acestrocephalus maculosus*	0.6763889	0.5011905	AEM+HPP
2	*Ammoglanis diaphanus*	0.3564484	0.1543651	AEM+HPP
3	*Ancistrus aguaboensis*	0.7789187	0.6099206	AEM+HPP
4	*Ancistrus minutus*	0.6935847	0.5198413	AEM+HPP
5	*Apareiodon cavalcante*	0.6418651	0.4940476	AEM+HPP
6	*Apteronotus camposdapazi*	0.8725694	0.7515873	AEM+HPP
7	*Aspidoras eurycephalus*	0.7593585	0.5719577	AEM+HPP
8	*Astyanax courensis*	0.9105159	0.8503968	AEM+HPP
9	*Astyanax unitaeniatus*	0.8411706	0.7404762	AEM+HPP
10	*Corumbataia tocantinensis*	0.8308036	0.6944444	AEM+HPP
11	*Corydoras maculifer*	0.5811508	0.4099206	AEM+HPP
12	*Creagrutus atrisignum*	0.789881	0.5989418	AEM+HPP
13	*Creagrutus saxatilis*	0.7654762	0.6460317	AEM+HPP

14	*Gymnotocinclus anosteos*	0.7261905	0.6297619	AEM+HPP
15	*Hemiancistrus cerrado*	0.8	0.702381	AEM+HPP
16	*Hemiancistrus micrommatos*	0.6785714	0.5198413	AEM+HPP
17	*Hemiancistrus spinosissimus*	0.7685847	0.621164	AEM+HPP
18	*Hypostomus asperatus*	0.8873677	0.7903439	AEM+HPP
19	*Hypostomus ericae*	0.6446032	0.3845238	AEM+HPP
20	*Loricaria lata*	0.3460317	0.1333333	AEM+HPP
21	*Moenkhausia pyrophthalma*	0.500248	0.2274802	AEM+HPP
22	*Moenkhausia tergimacula*	0.7201389	0.5269841	AEM+HPP
23	*Pimelodus joannis*	0.5373016	0.4253968	AEM+HPP
24	*Pimelodus stewarti*	0.6741071	0.502381	AEM+HPP
25	*Rineloricaria osvaldoi*	0.7611111	0.5850529	AEM+HPP
26	*Steindachnerina gracilis*	0.5118717	0.2583333	AEM+HPP
27	*Steindachnerina notograptos*	0.7535714	0.6515873	AEM+HPP
28	*Sternarchorhynchus mesensis*	0.8170635	0.743254	AEM+HPP
29	*Xenurobrycon coracoralinae*	0.5279762	0.4361111	AEM+HPP

Table S2 – Built and planned hydropower plants (HPP) located in the Tocantins-Araguaia River basin.

Name	Municipality	State	River	Capacity (KW)	Stage
Ponte Nova	Rio Sono	TO	Sono	73000	Planned

Peixe Angical	Peixe	TO	Tocantins	498750	Built
Cana Brava	Cavalcante	GO	Maranhão and Tocantins	450000	Built
Marabá	Marabá	PA	Tocantins	2160000	Planned
Laguna	Barro Alto	GO	Maranhão	40000	Planned
Araguanã	Araguanã	TO	Araguaia	960000	Planned
Natividade	Chapada da Natividade	TO	Manuel Alves	29500	Planned
Araguaia	Montes Claros de Goiás	GO	Claro	19900	Planned
São Domingos	São Domingos	GO	São Domingos	12000	Built
Foz do Atalaia	Nova Roma	GO	Paranã	72000	Planned
Cachoeira da Velha	Mateiros	TO	Novo	81000	Planned
São Salvador	Paranã	TO	Tocantins	243200	Built
Vermelho	Rio Sono	TO	Sono	58900	Planned
Ipueiras	Ipueiras	TO	Tocantins	480000	Planned
Mortes 2	Nova Xavantina	MT	Mortes	310400	Planned
Jatobá	Novo São Joquim	MT	Mortes	41800	Planned
Nova Roma	Nova Roma	GO	Paranã	49790	Planned
Porteiras	Barro Alto	GO	Maranhão	86000	Planned
Macaúba Alta	Jussara	GO	Claro	13800	Planned
Mirador	Niquelândia	GO	Tocantinzinho	106000	Planned

Couto Magalhães	Santa Rita do Araguaia	GO	Araguaia	150000	Planned
Paranã	Paranã	TO	Paranã	90000	Planned
Boaventura	Pontal do Araguaia	MT	Garças	32100	Planned
Santa Tereza	Santa Tereza do Tocantins	TO	Balsas	74500	Planned
Arraias	Paranã	TO	Palma	70000	Planned
Perdida 2	Rio Sono	TO	Perdida	49000	Planned
Barra do Palma	Paranã	TO	Palma	85000	Planned
Buritizal	Novo São Joaquim	MT	Mortes	42800	Planned
Água Limpa	General Carneiro	MT	Mortes	320000	Planned
São Domingos	Paranã	TO	Paranã	50000	Planned
Serra da Mesa	Colinas do Sul	GO	Tocantins	1275000	Built
Estreito	Palmeiras do Tocantins	TO	Tocantins	1087000	Built
Aparecida do Rio Claro Alta	Montes Claros de Goiás	GO	Claro	14400	Planned
Monte Santo	Rio Sono	TO	Sono	47000	Planned
Santa Isabel	Ananás	TO	Araguaia	1080000	Planned
Serra Grande	Padre Bernardo	GO	Verde	9000	Planned
Buriti Queimado	Hidrolina	GO	Almas	142000	Planned
Toricoejo	Barra do Garças	MT	Mortes	76000	Planned
Tucuruí	Tucuruí	PA	Tocantins	8535000	Built

Torixoréu	Torixoréu	MT	Araguaia	408000	Planned
Maranhão	Mimoso de Goiás	GO	Maranhão	125000	Planned
Luís Eduardo Magalhães (Lajeado)	Miracema do Tocantins	TO	Tocantins	902500	Built

Figure S1 – All species occurrences (black dots) plotted along with the randomly generated pseudo-absences (red Xs) within the gridded Tocantins-Araguaia River basin divided by the two main sub-basins (Dark grey area = Araguaia portion; Light grey area = Tocantins portion). Figure numbers correspond to the species ID available in Table S1.

Conclusão Geral

Neste trabalho nós adaptamos e desenvolvemos a metodologia de mapas de autovetores assimétricos (AEM) para o uso em modelos de distribuição de espécies. Sendo que para os modelos, esses filtros espaciais atuam como preditores de dispersão assimétrica seguindo o fluxo de direcionamento da dispersão das espécies pelos canais fluviais, inserindo a informação de movimento direcional das espécies de peixes de água doce. Aqui nós evidenciamos que os filtros, quando usados como preditores a partir da criação da matriz binária direcional, melhoram a acurácia estatística dos modelos quando usados como único conjunto de variáveis ou em conjunto com as variáveis tradicionais climáticas. Além disso, eles evitam extrapolações nas predições para áreas não acessíveis ou de difícil acesso para migração, mesmo que sejam áreas climaticamente similares às da distribuição atual da espécie. Por exemplo, distribuição para sub-bacias desconectadas ou com ponto de conexão longitudinal distante para a dispersão, como é o caso da bacia Tocantins-Araguaia usada como modelo nos artigos. Posteriormente, o uso abordagem permite também a inserção de desconexões na matriz de direcionamento das espécies representando barreiras de dispersão. Nesse sentido, essas barreiras foram inseridas *a priori* por meio dos preditores assimétricos de dispersão no momento da modelagem. Estas barreiras por sua vez nos modelos aumentam a acurácia dos mesmos, representando a realidade da fragmentação da paisagem das espécies que afetam a sua dispersão para outras áreas adequáveis para sua existência.

Assim, neste trabalho nós evidenciamos a importância do uso de preditores de dispersão assimétrica e suas variações para o uso em modelos de distribuição de espécies e como estes afetam e melhoram as predições de distribuição potencial destes modelos. Visto que a área de modelagem de distribuição de espécies é recente, porém altamente mutável, nós esperamos que novas técnicas sejam desenvolvidas nesse sentido de dispersão assimétrica *a priori*. Por exemplo, essa metodologia ainda é muito manual na criação da matriz de direcionamento, sendo que a automatização da criação da mesma permitiria o seu uso em escalas maiores, testando hipóteses de dispersão entre bacias e não somente dentro da mesma bacia. Desta forma, esperamos que essa metodologia possa ser ajustada e melhorada para implementar ferramentas de distribuição de espécies, auxiliando nas tomadas de decisões conservacionistas.