Universidade Estadual de Mato Grosso do Sul
Unidade Universitária de Dourados

Programa de Pós-Graduação em Recursos Naturais

MODELANDO A OCORRÊNCIA DE ESPÉCIES DE PEIXES EM RIOS NEOTROPICAIS UTILIZANDO MÉTODOS ESTATÍSTICOS BASEADOS NA CONECTIVIDADE HIDROLÓGICA

Julio César Jut Solórzano

Dourados – MS
Fevreiro/2018

Universidade Estadual de Mato Grosso do Sul
Unidade Universitária de Dourados
Programa de Pós-Graduação em Recursos Naturais

MODELANDO A OCORRÊNCIA DE ESPÉCIES DE PEIXES EM RIOS NEOTROPICAIS UTILIZANDO MÉTODOS ESTATÍSTICOS BASEADOS NA CONECTIVIDADE HIDROLÓGICA

Julio César Jut Solórzano
Orientador: Yzel Rondon Súarez

"Dissertação apresentada ao programa de pós-graduação em Recursos Naturais, área de concentração em Recursos Naturais, da Universidade Estadual de Mato Grosso do Sul, como parte das exigências para a obtenção do título de Mestre em Recursos Naturais".

Dourados – MS
Fevreiro/2018

J97m Jut Solórzano, Julio César

Modelando a ocorrência de espécies de peixes em rios neotropicais utilizandométodos estatísticosbaseados na conectividade hidrológica/ Julio César Jut Solórzano. Dourados, MS: UEMS, 2018.

28 p. ; 30cm.

Dissertação (Mestrado) – Recursos Naturais – Universidade Estadual de Mato Grosso do Sul, Unidade Universitária de Dourados, 2018.

Orientador : Prof. Dr. Yzel Rondon Súarez.

1. Interpolação 2. Métricas 3. Parâmetros físico-químicos.

I.Título. CDD 23. de. 639.3

UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL
UNIDADE UNIVERSITÁRIA DE DOURADOS
PÓS-GRADUAÇÃO EM RECURSOS NATURAIS

ACADÊMICO: JULIO CÉSAR JUT SOLÓRZANO ORIENTADOR: PROF. DR. YZEL RONDON SÚAREZ

APROVADO (___/___/___)

_____

Prof. Dr. Yzel Rondon Súarez (orientador)

_____

Prof. Dr. Julio César Pereira

_____

Profa. Dra. Elaine Antoniassi Luiz Kashiwaqui

“Se não puder voar, corra. Se não puder correr, ande. Se não puder andar, rasteje, mas continue em frente de qualquer jeito”

Marthin Luther King

Dedico este trabalho à minha família e amigos!!!

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me conceder saúde para a realização deste trabalho. Meus agradecimentos ao meu orientador Prof. Dr. Yzel Rondon Súarez, pelos conselhos, por todos os ensinamentos e sugestões que foram essenciais para a realização desta dissertação. Muito obrigado Ao Dr. Julio César Pereira pela disposição em ajudar, pelas sugestões valiosas. Obrigada por todo tempo dedicado a mim, pela paciência e por todos os ensinamentos. Obrigado a professora Silvana Moretti pelos conselhos e ajuda moral valiosa amizade e disposição em todos os processos burocráticos que tive que passar para me regularizar como estudante no Brasil. À Universidade Estadual de Mato Grosso do Sul – UEMS, pela oportunidade que esta instituição me proporcionou. Agradeço grandemente a Fundect pelo apoio financeiro a Organização dos Estados Americanos (OEA) pela oportunidade que me forneceu para continuar meus estudos de mestrado. A todos os professores do Programa de PósGraduação em Recursos Naturais da Universidade Estadual de Mato Grosso do Sul, pelos valiosos ensinamentos e por sempre estarem disponíveis quando solicitados. A todos os membros do Laboratório de Ecologia de Peixes - UEMS. E especial agradecimento aos meus colegas e amigos Djalma, Eliza, Roberta quem me deram a forca emocional e muitas vezes força intelectual para eu aguentar minha estadia no Brasil.

Meus sinceros agradecimentos, por toda valiosa ajuda a Dra. Lucilene Finoto e Fabiane pela ajuda com as minhas análises estatísticas, eu sei que não fosse possível ter concluído com sucesso sem a ajuda de todos vocês. Vocês foram demais! Valeu por tudo. A todos os meus amigos e amigas pelos momentos de descontração. Por último, mas não menos importante agradeço a toda minha família em especial minha mãe que me forneceu apoio e me acompanho nesta jornada, por todo o incentivo agradeço muito de coração, toda a força e por todo o carinho dedicado a mim neste tempo. Obrigado a todos, que de alguma maneira contribuíram para a realização deste trabalho! Foram dois anos maravilhosos e de muito conhecimento!

SUMÁRIO

RESUMO ix

ABSTRACT x

CAPITULO I – CONSIDERAÇÕES GERAIS 1

Modelos estatísticos espaciais utilizando redes de fluxo 2

Analises de interpolação 2

Verificação da formatação dos dados espaciais 3

Construção do cenário de redes de fluxo: 4

Obtenção do arquivo SSN 4

Análises de interpolação 5

REFERÊNCIAS BIBLIOGRÁFICAS 7

CAPÍTULO 2 - MODELING FISH SPECIES OCCURRENCE IN NEOTROPICAL RIVER BASINS USING NETWORK LINKAGE STATISTICAL METHODS 9

ABSTRACT: 9

INTRODUCTION: 10

MATERIAL AND METHODS: 11

Study area: 11

Data processing: 12

RESULTS 12

DISCUSSION 16

ACKNOWLEDGEMENT 17

REFERENCES 17

CAPÍTULO 3 - CONSIDERAÇÕES FINAIS 20

RESUMO

O presente trabalho foi desenvolvido na área do Alto Paraná no estado de Mato Grosso do Sul, Brasil, usando como base 232 pontos de coleta donde foram registrados os parâmetros físicoquímicos da agua. Foi gerada uma camada de pontos hipotéticos separados um de outro a 5 km, distribuídos homogeneamente nas dois microbacias, procurando interpolar os valores das métricas físico-químicas em locais não amostrados como objetivo final avaliar a probabilidade de ocorrência de quatro espécies de peixes (*Hyphessobrycon eques, Piabina argentea, Leporinus friderici e Prochilodus lineatus*) em relação aos valores das métricas (profundidade, velocidade, altitude, condutividade elétrica, largura, e oxigênio dissolvido). As amostragens foram realizadas entre julho/2001 e novembro/2016, utilizando peneiras, rede de arrasto e pesca elétrica foi utilizada a metodologia desenvolvida por Ver Hoef. Para a interpolação das métricas, baseadas em matrizes de covariância definidas pelas distâncias dos rios e não distâncias euclidianas (distâncias geográficas). Em relação aos resultados de probabilidade de ocorrência, as variáveis altitude e profundidade apresentaram uma maior importância pois foram as duas variáveis que estiveram presentes como significantes no modelo de regressão logística para as quatro espécies. Os resultados evidenciaram que três das quatro espécies preferem rios com grandes profundidades e larguras maiores, definidos como os rios principais das microbacias, sendo estas espécies *P. lineatus, H. eques, L. friderici* e somente *P. argentea* prefere rios com menores profundidades e maiores altitudes, definida como as cabeceiras das bacias amostradas.

PALAVRAS-CHAVE: Interpolação, métricas, parâmetros físico-químicos, probabilidade de ocorrência.

ABSTRACT

The present study was carried out in the Alto Paraná area in State of Mato Grosso do Sul, Brazil, using 232 collection points from which the physical-chemical parameters of the water were recorded. A layer of hypothetical points separated from each other at 5 km, homogeneously distributed in the two micro basins, was generated, trying to interpolate the values of the physico-chemical metrics in non-sampled sites as a final objective to evaluate the probability of occurrence of four species of fish (*Hyphessobrycon eques, Piabina argentea, Leporinus friderici and Prochilodus lineatus*) in relation to the values of the metrics (depth, velocity, altitude, electrical conductivity, width and dissolved oxygen). Samples were taken between July 2001 and November 2016 using sieves, (1), (2), (2), (2) and (2), respectively, are used to determine the probability of occurrence of the covariance matrices, the variables altitude and depth presented a greater importance because they were the two variables that were present as signifiers in the logistic regression model for the four species. The results showed that three of the four species prefer rivers with great depths and larger widths, defined as the main rivers of the micro basins, being these species *P. lineatus, H. eques, L. friderici* and only *P. argentea* prefers rivers with lower depths and higher altitudes, defined as the headwaters of sampled basins.

KEY WORDS: Interpolation, metrics, physical-chemical parameters, probability of occurrence.

CAPITULO I – CONSIDERAÇÕES GERAIS

Um número constante em crescimento de estudos em ecologia, biogeografia, biologia da conservação têm tentado construir modelos preditivos de distribuição de espécies que visa uma melhor proteção e gestão dos recursos naturais e dos ecossistemas. A maioria deles, usam técnicas de modelagem estatísticas usando a distribuição observada das espécies. (GUISAN; ZIMMERMANN, 2000; THUILLER, 2003).

Há muitas metodologias teóricas que permitem a modelagem ambiental de bacias hidrográficas que são escolhidos de acordo com o objetivo (BELMONTE; NUÑEZ, 2006) porém de acordo com VER HOEF et al (2006), um dos principais problemas com o desenvolvimento de modelos espaciais para as redes de drenagem nas bacias hidrográficas (redes hidrográficas) é que muito pouco se sabe sobre as propriedades da matriz de variância/covariância quando se utiliza a distância hidrológica e não a distância euclidiana (distância geográfica) entre os pontos.

Este estudo foi realizado com base na premissa de que os peixes têm comportamentos específicos de acordo com a variação nas caraterísticas do ambiente onde estão inseridos, por exemplo: de acordo com GILLOOLY et al., (2001), o comportamento, metabolismo e a sobrevivência depende da temperatura ambiente, baseado no fato que os peixes são ectotérmicos, o que significa que estas espécies não são capazes de manter sua temperatura constante, variando junto com a temperatura do ambiente. Assim como a temperatura é uma variável que pode alterar o comportamento em peixes, se pretende testar esta hipótese para outras variáveis, tais como: Turbidez, altitude, profundidade, largura, condutividade elétrica, oxigênio dissolvido, pH, buscando relacionar a distribuição dos peixes no Alto Rio Paraná em duas bacias: Rio Ivinhema e o Rio Amambai.

A bacia do rio Ivinhema ocupa total ou parcialmente 25 municípios, sendo: Anaurilandia, Angélica, Bataiporã, Deodápolis, Itaporã, Ivinhema, Jatei, Rio Brilhante, Novo Horizonte do Sul, Taquarussu, Vicentinha, Antônio Joao, Caarapó, Jutí, Laguna Caarapa, Maracaju, Nova Alvorada do Sul, Nova Andradina, Naviraí, Ponta Porã, e Sidrolândia (IMAP, 2006). Por outro lado, a bacia do rio Amambai, ocupa total ou parcialmente 10 municípios sendo: Amambai, Aral Moreira, Caarapo, Coronel Sapucaia, Iguatemi, Itaquirai, Jutí, Laguna Caarapa, Naviraí, Ponta Porã (BEREZUK et al, 2014).

De acordo com a Agência Nacional de Energia ANEEL (2016), existem 7 usinas hidrelétricas previstas ao longo do eixo principal do rio Amambai, com 6 pequenas centrais

hidrelétricas (PCH) e uma usina hidrelétrica (UHE), além de ter uma usina hidrelétrica localizada em um afluente do rio Amambai, totalizando 8 empreendimentos hidrelétricos nesta bacia. Para a bacia do rio Ivinhema existem duas PCHs em funcionamento no Alto Rio Dourados e outras duas UHEs previstas.

Modelos estatísticos espaciais utilizando redes de fluxo

Para a análise estatística com base no fluxo hidrológico foi necessária a utilização de duas metodologias que permitem o desenvolvimento de banco de dados específicos (arquivos SSN) criados através de softwares de processos geo-espaciais STAR® (PETERSON & VER HOEF. 2015); a segunda metodologia permite a leitura destes vetores no R® através do software de análise estatística espacial SSN® (VER HOEF, et al. 2014). O processo de criação do arquivo SSN no software R permitiu a obtenção de uma rede de fluxo (rede de rios no Paraná Alto Rio, incluindo a bacia do Rio Amambai e bacia do Rio Ivinhema), que foi obtido a partir da base de dados do Instituto brasileiro de Geografia e Estatística (IBGE), além de obter modelo digital de elevação (DEM's) e imagens Landsat 7 e 8, devido ao fato de que as amostragens foram obtidas em diferentes anos. Além dos diferentes pontos de recolha (232), que estão disponíveis no banco de dados do Laboratório de Ecologia Aquática da Universidade Estadual de Mato Grosso do Sul, estes contêm valores de diferentes métricas que foram usados para interpolação (altitude, inclinação, profundidade, largura, velocidade, condutividade eléctrica e oxigênio dissolvido), que foram utilizados para a geração da camada de pontos espaçados entre si por 5000 m (2901 pontos), a fim de permitir a estimativa do valor destas variáveis nos locais (interpolados) onde os valores não foram obtidos. O objetivo do presente trabalho foi analisar eventos de presença ou ausência de quatro espécies de peixes em 2901 pontos hipotéticos, com o uso da análise de interpolação de métricas em pontos de amostragem conhecidos.

Analises de interpolação

Para a análise de dados foi necessário o uso de aplicativos desenvolvidos independentemente por diferentes organizações. Spatial Tools for The Analysis Of River Systems (STAR): Aplicativo (extensão) desenvolvido pelo ARC Centre for Excellence in Mathematical and Statistical Frontiers (ACEMS) e do Institute For Future Environments (IFE) da Queensland Unversity of Technology.

Para a análise dos dados com o aplicativo STAR, foram necessários a atualização de acordo com os seguintes aplicativos e softwares.

Requerimientos do proceso:

- ArcGis 10.5 licença avançada (versão teste)
- Extensão Spatial Analyst
- STAR versão 2.0.4 para ArcGis
- Pyton versão 2.7.8
- Pyton Win Versao 2.7.8

Dados requeridos para o funcionamento do conjunto de ferramentas STAR para ArcGis:

Tabela 1. Requerimento de dados para o modelamento espacial de redes hídricas.

Nome	**Formato**	**Descrição**	**Requerido**
Locais de coleta	Vetor	Localização e dados dos locais de amostragem	Sim
Área de estudo	Vetor	Rede hídrica	Sim
RCAs	Vetor ou raster	Área de contribuição para cada segmento da rede hídrica	Opcional
DEM	Raster	Modelo digital de elevação	Opcional
Corpos de água	Raster	Lagos, lagoas, estuários, etc.	Opcional
Dados auxiliares	Raster	Geologia, clima, uso e cobertura, etc.	Opcional

Verificação da formatação dos dados espaciais

A direção de fluxo na rede hídrica, uma informação importante para a modelagem de dados espaciais com caraterísticas de rede, foi digitalizada, uma vez que nem sempre os dados disponíveis nos vetores disponibilizados pelo IBGE apresentam a vetorização seguindo a direção do fluxo da agua, partindo das nascentes até a foz da bacia (Figura 1).

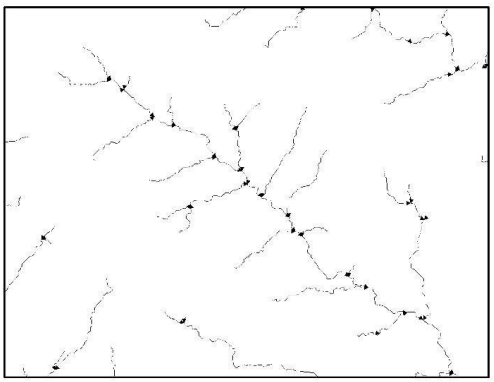

Figura 1. Segmentos de rios digitalizados corretamente na direção abaixo.

Construção do cenário de redes de fluxo:

O arquivo LSN (Landscape Network), é um tipo de arquivo utilizado para representar o contexto espacial e as relações com a informação geográfica (THEOBALD, et al., 2006).

A rede hídrica como insumo foi obtida do servidor do Instituto Brasileiro de Geografia e Estatísticas (IBGE), o qual foi adequado e re-digitalizado para conferir todos os possíveis erros topológicos potenciais que interferem no processo do analises (redes não conectadas, sobreposição de vetores, direção de fluxo e confluência de um ou mais trechos).

Obtenção do arquivo SSN

Após de a utilização dos diferentes aplicativos e conferir os erros topológicos dos insumos, foi obtido o arquivo Spatial Sistem Network (SSN) o qual foi utilizado para o analises estadístico para a precisão das métricas espaciais, paisagísticas e da qualidade da agua, o qual tem a seguinte estrutura:

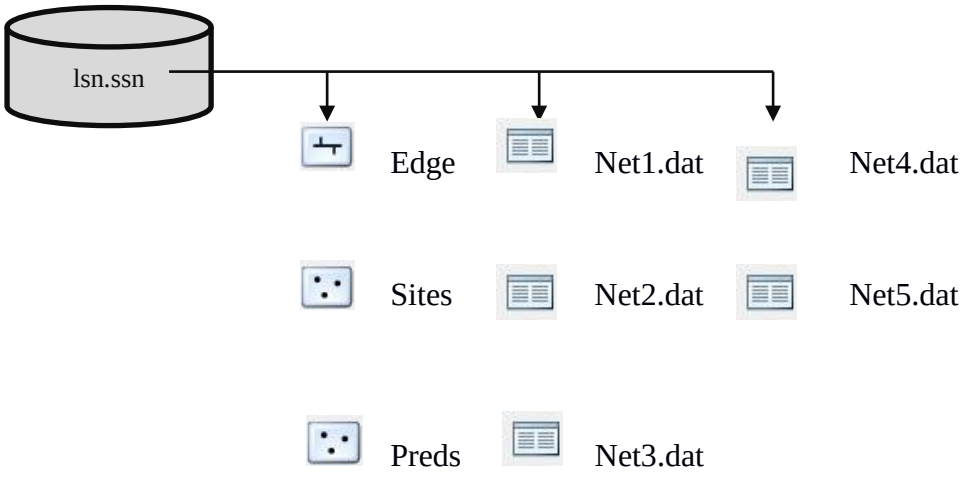

Figura 4. Estrutura do arquivo comprimido SSN para a utilização no software R.

Onde o arquivo “Edge” contém a estrutura de redes que será analisada; o arquivo “Sites” contém as informações dos locais de coleta que contém as informações das métricas que serão utilizadas além das informações estatísticas calculadas com o pacote STAR®; “Preds” contém uma camada de pontos com distância de 5km e as mesmas estatísticas espaciais calculadas para a camada sites. Nos quais foi realizada a interpolação para cada uma das métricas; os arquivos de texto (Net1.dat, Net2.dat, Net3.dat, Net4.dat e Net5.dat) contém todas as informações numéricas que serão utilizadas para as análises no R, como distâncias entre os pontos e os valores da função aditiva.

Análises de interpolação

Para cada uma das métricas foi analisado um método de interpolação não espacial para permitir a comparação da variação com os métodos espaciais. Neste caso é avaliado até que ponto a análise pela rede de drenagem difere da análise pela distância euclidiana.

Altitude: Esta métrica não foi interpolada, devido ao fato de ser obtida facilmente através do modelo digital de elevação (MDE) o qual foi obtido do servidor do Serviço Geológico dos Estados Unidos (USGS) por sua sigla em inglês, de onde foram extraídos os valores para cada um dos pontos de predição fazendo um total de 2901 valores os quais vão de 209 até 642 m.a.n.m.

Declividade: Para a obtenção dos dados de declividade não foi necessário o processo de interpolação, devido à disponibilidade do modelo digital de elevação para os pontos de predição e a distância entre eles foi facilmente calculado, os valores obtidos variam desde 0 até 6.85 graus.

Profundidade: Para a obtenção de dados das profundidades em cada um dos pontos de predição foi utilizado o processo de interpolação utilizando o pacote SSN do software R. Foi analisada a dependência espacial após a análise para identificar os valores de range, nuggett e sill do semi variograma, que é a representação da autocorrelação espacial, para a variável de interesse.

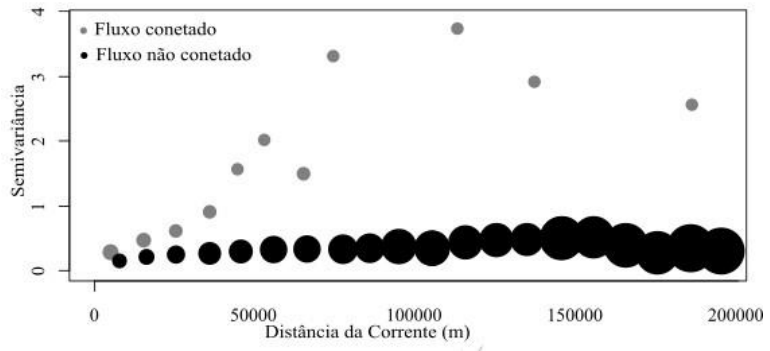

Figura 2. Torguegrama da distribuição da profundidade indicando a estabilização no fluxo não conectado e fluxo conectado em função da distância entre pontos.

Para o modelo não espacial, as métricas que deram significância (p=< 0.05) foram velocidade (p=0,017), largura (< 0,001), altitude (p<0,001) e declividade (p=0.009). Para o modelo espacial combinado os modelos tailup + taildown + e o modelo euclidiano as métricas significativas foram largura (p<0,001), altitude (p<0,001) e declividade (p=0.037),

mas a explicação da variabilidade dos dados foi major com os modelos Tail-Up + TailDown, então foi necessário ajustar os modelos a uma combinação que fosse a mais adequada para a natureza dos mesmos. O ajuste de resíduos foi necessário para garantir uma homogeneização dos dados, pois o que se procura é que a dependência espacial seja a menor possível, os valores atípicos identificados foram de -3.

Após a análise dos resíduos foi gerado uma nova análise de significância já com o modelo adequado à natureza dos dados. De acordo com os dados obtidos, as métricas largura, altitude e declividade ainda apresentam correlação com a profundidade apresentando valores $p<0,05$. No entanto, o efeito "partial sill", associado com o modelo exponencial Tail-Down e o modelo euclidiano foi baixo em comparação com o modelo Exponencial Tail-up. Porém o modelo tail-down e euclidiano foram excluídos do modelo ajustado. Contudo, a largura apresentou correlação significativa em relação a profundidade, mas não foi possível adicionar ao modelo, pois tinham se disposição só das dois métricas restantes (elevação e declividade) as quais foram obtidas do modelo digital de elevação para todos os pontos de predição, pois para que uma métrica seja inserida no modelo deverá estar adicionada em todos os pontos de predição.

Ao reduzir a quantidade de modelos na combinação e a quantidade de métricas, a colinearidade entre as mesmas aumenta a sua significância, podendo gerar um viés no resultado estatístico.

A validação cruzada Leave-one-out (LOOCV) gera um bom diagnóstico do rendimento do modelo, definida pela equação:

$$E\{Perror^{l-1}\} = \frac{E\{l(x1, y1, x2, y2, \dots, xl, yl)\}}{l}$$

Onde $Perror^{l-1}$ é a probabilidade de erro do teste para um classificador treinado numa amostra de tamanho $l - 1$, $l(x1, y1, x2, y2, \dots, xl, yl)$permite identificar o número de erros "leveone-out" para um classificador treinado num conjunto de pares de dados $\{(x_i, y_i)\}^l_{i=1}$ é o tamanho de l (CRAWLEY; ALBOT, 2003). A função GR2 calcula o r^2 para o objeto "glmssn" ajustado utilizando o modelo clássico $y* = X*\beta + \epsilon*$ assim a função "varcomp" assume aquela proporção da variação que não é explicada pelos efeitos fixos (sejam estes chamados de covariáveis) na função GR2 e o ajusta para cada "partial sill" de cada covariância para cada componente espacial.

Assim os efeitos fixos explicam perto de 12.7% da variação dos dados sendo um 84.6% por parte do modelo exponencial "tail-up" e pequeno efeito pepita "nuggett". Para a seleção do modelo, foi utilizado o critério de informação de Akaike (IAC) pelas suas siglas em inglês. O qual estabelece que os menores valores garantem uma major confiança na hora de realizar a interpolação em estruturas de covariância (CAVANAUGH, 1997). Para ter segurança com o modelo ideal selecionado, analisou-se os valores de IAC para vários modelos existentes para estruturas de covariância citados por VER HOEF, et al. (2006). Baseado nos menores valores de IAC e o baixo erro de predição quadrático médio, dois modelos foram adequados para a estrutura de dados, no entanto decidimos utilizar o modelo combinado Spherical.tailup + Spherical.taildown + Nugget pois uma variação menor de 2 IAC não é significativa, e os modelos tailup e taildown são os mais adequados para este analise.

Assim foram obtidos os resultados anteriores, observando que as métricas altitude e declividade não foram interpoladas. Foi realizado um histograma para análise de resíduos e identificar outliers em relação a predições que não aparecem como grandes desvios do ajuste.

Para o processo de predição das probabilidades de ocorrência foi utilizada a função "predict" do software R®. O mesmo processo de análise de interpolação foi utilizado para as outras métricas, levando em consideração os resultados das significâncias e confiabilidade dos métodos de interpolação espacial utilizados.

REFERÊNCIAS BIBLIOGRÁFICAS

BELMONTE, S., & NÚÑEZ, V. Desarrollo de modelos hidrológicos con herramientas SIG. **GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica** v. 6, p. 15-27. 2006.

BEREZUK, A. G., MARTINS, J. H. P., DO NASCIMENTO RIBEIRO, A. F., & DE ALCÂNTARA LIMA, P. Análise morfométrica linear e areal da Bacia Hidrográfica do Amambai-Mato Grosso Do Sul-Brasil. **Revista Eletrônica AGB-TL**, v. 1, n. 20, p. 8-38. 2014.

CAVANAUGH, J. E. Unifying the derivations for the Akaike and corrected Akaike information criteria. **Statistics & Probability Letters**, v. 33 n. 2, p 201-208. 1997.

CAWLEY, G. C., & TALBOT, N. L. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. **Pattern Recognition**, v. 36, n. 11, p. 2585-2592. 2003.

GILLOOLY, J. F., BROWN, J. H., WEST, G. B., SAVAGE, V. M., & CHARNOV, E. L. Effects of size and temperature on metabolic rate. **Science**, v. 293 n. 5538, p. 2248-2251. 2001.

GUISAN, A., & ZIMMERMANN, N. E. Predictive habitat distribution models in ecology. **Ecological modelling**, v. 135 n. 2-3, p. 147-186. 2000.

HOEF, J. V., PETERSON, E., CLIFFORD, D., & SHAH, R. SSN: An R package for spatial statistical modeling on stream networks. **Journal of Statistical Software**, v. 56, n.3, p. 145. 2014.

IMAP, M., & DO RIO IVINHEMA, B. Diagnóstico hidroambiental e socioeconômico. Campo Grande: **Imap/MS**. 2006.

MAGNUSON, J. J., CROWDER, L. B., & MEDVICK, P. A. Temperature as an ecological resource. **American Zoologist**, v. 19 n. 1, p. 331-343. 1979.

PETERSON E.E. AND VER HOEF J.M. STARS: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. **Journal of Statistical Software**, v. 56, n. 2. 2014.

THEOBALD D.M., NORMAN J.B., PETERSON E., FERRAZ S., WADE A., AND SHERBURNE M.R. Functional Linkage of Water basins and Streams (FLoWS) v1 Users Guide: ArcGIS tools for Network-based analysis of freshwater ecosystems. **Natural Resource Ecology Lab, Colorado State University, Fort Collins**, **CO.** p. 43. 2006.

THUILLER, W. BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. **Global change biology,** v. *9, n.* 10, p. 13531362. 2003.

VER HOEF, J. M., PETERSON, E., & THEOBALD, D. Spatial statistical models that use flow and stream distance. **Environmental and Ecological statistics**, v. 13, n.4, p. 449-464. 2006.

CAPÍTULO 2 - MODELING FISH SPECIES OCCURRENCE IN NEOTROPICAL RIVER BASINS USING NETWORK LINKAGE STATISTICAL METHODS

Julio César Jut Solórzano[1] & Yzel Rondon Súarez[2]

1-Graduate Program in Natural Resources, Universidade Estadual de Mato Grosso do Sul/Centro de Estudos em Recursos Naturais, Dourados-MS, Brazil. 79804-970.
2-Universidade Estadual de Mato Grosso do Sul/Centro de Estudos em Recursos Naturais/Laboratório de Ecologia. Dourados-MS, Brazil. 79804-970

ABSTRACT:

1. The probability of occurrence of fish species is strongly related to the conditions that occur in the drainage networks, however, these conditions are not stationary because within these scenarios (rivers) the water has movement and the conditions vary in relation with its course.

2. Hydrological, spatial and water quality metrics (altitude, slope, depth, velocity, dissolved oxygen and river width) were interpolated in places not sampled (every 5 km), to establish metrics throughout the river network and evaluate the probability of occurrence of four fish species (*Hyphessobrycon eques, Piabina argentea, Leporinus friderici and Prochilodus lineatus*).

3. A distribution pattern strongly related to altitude was identified. Three of the four species had a distribution related to the main networks, and only the species *P. argentea* showed a distribution pattern in the upper and upper areas

4. The main descriptors of distribution of species were associated with hydrological characteristics based on altitude and depth because the species have behaviors in terms of distinctive preference such as water velocity, dissolved oxygen, which are directly linked to different altitudes and position of the currents.

Key words: stream distance, interpolation, Ivinhema, Amambai, logistic regression.

INTRODUCTION:

Fishes are the main diversified vertebrates, with aproximately 34700 species recognized in the world (Eschemeyer & Fong 2018). These higher species diversity is accomplished by a great ecological diversity with adaptations to live in some greater variations of ecological conditions. Neotropical icthyofauna have at least 13000 species (Nelson et al., 2016) and La Plata River had 943 fish species (Reis et al., 2016).

The knowledge of the role of environmental determinants of species distribution is one the main aim in neotropical ecology, despite this, some basins still have not a credible species list. In this context, the modeling of fish distribution has always been a matter of great importance for ecology (Guisan & Thuiller 2005). It is well known that the variation in spatial scale of the physical and chemical characteristics of water can influence habitat connectivity in fundamental ecological processes, community structure and nutrient flow (Ten & Burbano 2007).

A steadily growing number of studies in ecology, biogeography, and conservation biology have attempted to construct predictive models of species distribution aiming a better protection and management of natural resources and ecosystems. Most of them apply statistical modeling techniques using the observed species distribution (Guisan & Zimmermann, 2000; Thuiller, 2003). When working with flow networks, it is possible that the "euclidean distance" is not very helpful since the spatial scenarios in a terrestrial environment do not obey the same distribution rules as the scenarios in river network environments. The distance of the current in this case is defined as the shortest distance between two points where this distance is calculated along the current, it is for this reason that the models of self-covariance for Euclidean distance are not very helpful for the flow distances (Gardner, et al. 2003).

There are many theoretical methodologies that allow the environmental modeling of watersheds that are defined for each target need (Belmonte & Nuñez 2006), however, one of the main problems with the development of spatial models for the networks (hydrographic networks) is that very little is known about certain self-covariance models when using stream distances instead of geographic distance (Ver Hoef et al., 2006). The objective of this work was to analyze the probability of occurrence of four fish species (*Hyphessobrycon eques*, *Piabina argentea*, *Leporinus friderici* and *Prochilodus lineatus*) in a neotropical river basin using a network statistical procedure.

MATERIAL AND METHODS:

Study area*:*

Upper Paraná River basin are located to central Brazil, with an area of approximately 900000 km^2 (Langeani et al., 2007). The sampled basins (Amambai and Ivinhema) are located to western portion of the Upper Paraná River basin in Mato Grosso do Sul State (Figure 1).

The fish samples in Ivinhema River Basin were made from 2001 to 2016 with 179 stream portions and for Amambai river basin were 59 streams portions from 2007 to 2015. The samples were made using mainly a rectangular metal frame with dimensions of 0.8x1.2 m and a seine net with 1.5x5m, both with a mesh of approximately 2 mm. In rivers portions or streams with larger volume are also used trawl nets of different mesh sizes (15, 20, 30, 40 and 50 mm between adjacent nodes). Some samples sites were sampled with electrofishing. Both rivers as free of larger impoundments with only two smalls hydroelectric central (Small hydroelectrics) in upper Dourados River, Ivinhema basin.

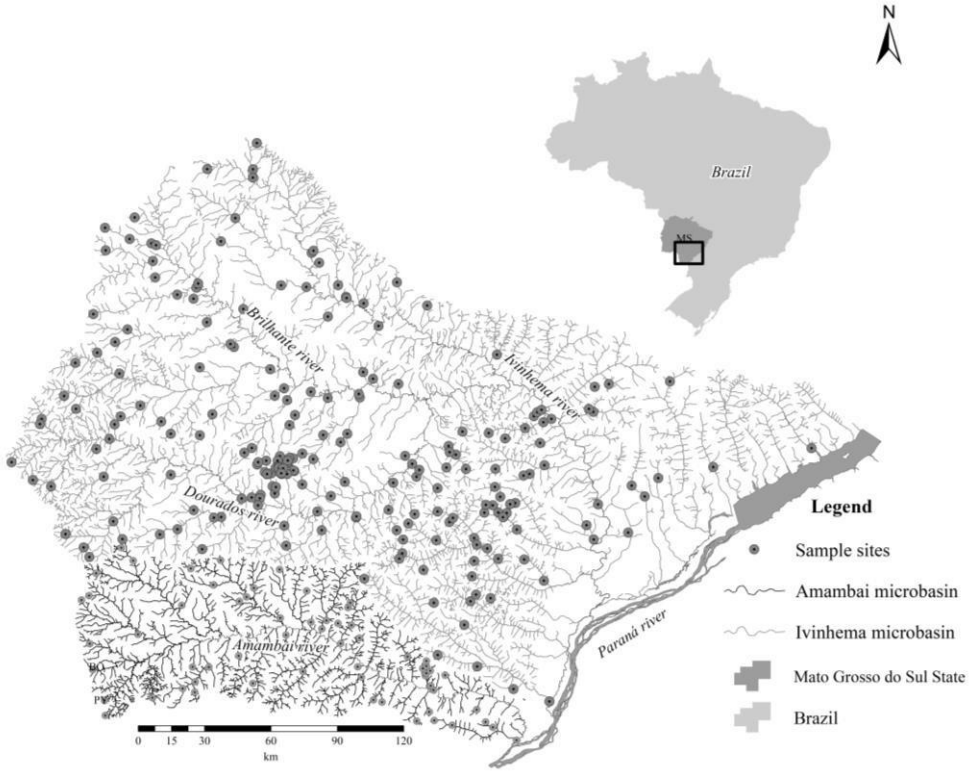

Figure 1. Location of study area and sampled sites of fish species in the basins of Amambai and Ivinhema rivers in Upper Paraná area.

For each sample site, a set of environmental variables was obtained, including the geographical coordinates, stream width, stream depth, water velocity, dissolved oxygen and water electrical conductivity, using a digital flowmeter and using a digital portable condutivimeter. As result of a long samples period others limnological variables are not available for all sites. Stream depth, width and water velocity were measured at least 5 times, in different portions of the sampling area, and a simple arithmetic mean were used as a descriptor of the stream stretch; the altitude (m.a.s.l.) and the slope (degrees) were obtained by extracting the values of the digital elevation models (DEM's) available on the United States Geological Survey (USGS), the samples correspond to the years 2007 to 2016, which are in the database of the Aquatic Ecology Research Laboratory of the State University of Mato Grosso do Sul.

Data processing:

The sampled rivers were vectorized using an elevation digital models available in the database of the Geological Service of the United States (USGS). The metrics interpolated with the Spatial Tools for the Analysis of River Systems (STAR®) and Spatial Stream Network (SSN®) (Ver Hoef et al., 2014). Methodologies to be used in the logistic regression analysis were: Altitude, depth, width, dissolved oxygen, electrical conductivity and speed. To perform the interpolation of the variables, linear models with spatial dependence structure were adjusted. For this, a selection of the available models was made, where the covariables of the model were selected, as well as the structure of the spatial dependence, in accordance with (Ver Hoef et al., 2006)

The interpolated data for each point distributed in the two watersheds were used to perform the probability of occurrence analysis with the results of logistic regression methodology.

Logistic regression was used to estimate the importance of environmental variables to predict species occurrence. The final model was selected using a stepwise procedure. After this procedure, the retained model was used to predict probability of occurrence of each species in the whole studied hydrographic basins (using the 2901 created points).

$$Logist(\pi_i) = log\left(\frac{\pi_i}{1-\pi_i}\right)$$

Where π_i is the probability of observing the category or event to be predicted and 1-π_i is the probability of not observing the category or event to be predicted (Alderete, 2006).

RESULTS

The results of logistic regression suggest that altitude was the environmental descriptor that presented significant influence on the occurrence of all analyzed species, acting negatively on species occurrence probability for *P. lineatus*, *H. eques* and *L. friderici* and positively for *P. argentea*. Stream depth was also an important variable to occurrence of *P. lineatus*, *H. eques* and *L. friderici* that have probability of occurrence positively influenced by stream depth. Stream width was not selected as important variable for none evaluated fish species, Table 2. Results of stepwise logistic regression for probability of occurrence of studied fish species in Upper Paraná River basin.

The retained model showed that for *P. lineatus* three variables can describe spatial distribution (depth, altitude and water conductivity), in this importance order (Table 2 and Figure 2). *H. eques* have your occurrence predicted by altitude and depth (Table 2 and Figure 3). *P. argentea* occurrence are mainly influenced by water velocity, dissolved oxygen and altitude (Table 2 and Figure 4).

Table 1. Environmental descriptors selected by stepwise logistic regression for Upper Paraná River basin

Species	Variable	Estimated	t	p value
P. lineatus	Intercept	0.843	0.639	0.523 ns
	Elevation	-0.013	-3.163	0.002 **
	Depth	1.127	4.396	1.1e-05 ***
	Conductivity	0.006	2.913	0.003 **
H. eques	Intercept	-0.052	-0.040	0.967 ns
	Elevation	-0.009	-2.255	0.024 *
	Depth	0.628	2.625	0.009 **
P. argentea	Intercept	-6.133	-5.756	8.61e-09 ***
	Elevation	0.004	2.523	0.012 *
	velocity	1.960	3.756	0.0002 ***
	Dissolved oxygen	0.412	3.490	0.0005 ***
L. friderici	Intercept	0.087	0.082	0.93502 ns
	Elevation	-0.008	-2.693	0.00708 **
	Depth	1.175	4.599	4.24e-06 ***

Significance codes: p<0.001 '***', p< 0.01 '**' p<0.05 '*' ns=non significant

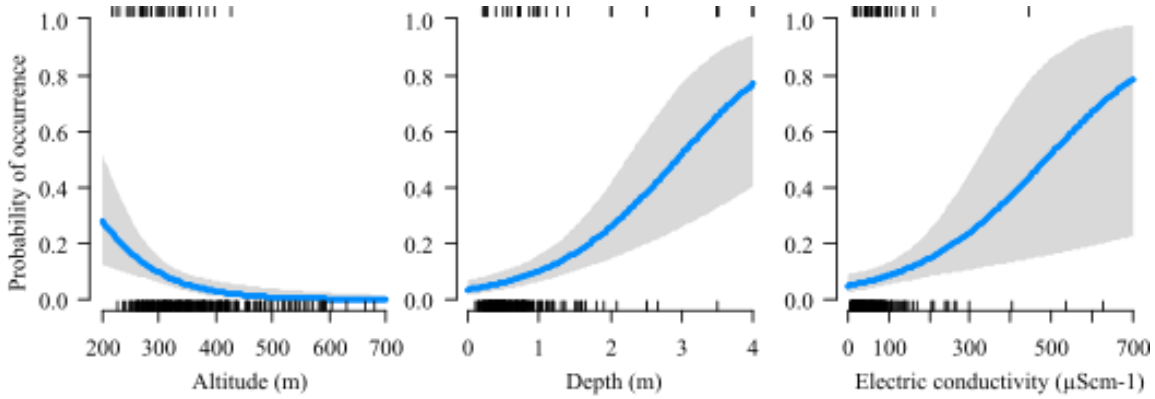

Figure 2. Relation among probability of occurrence of *P. lineatus* and environmental descriptors selected by stepwise logistic regression and the confidence interval (in grey) for Upper Paraná River basin.

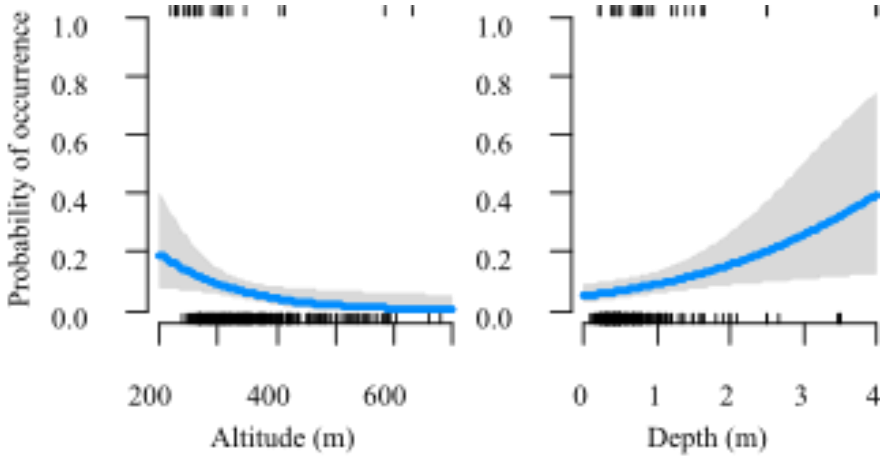

Figure 3. Relation among probability of occurrence of *H. eques* and environmental descriptors selected by stepwise logistic regression and the confidence interval (in grey) for Upper Paraná River basin.

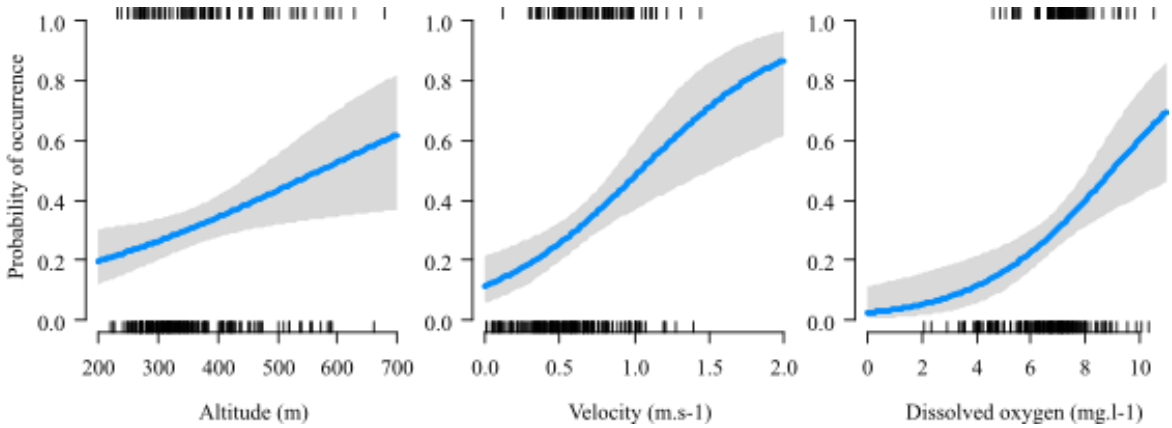

Figure 4. Relation among probability of occurrence of *P. argentea* and environmental descriptors selected by stepwise logistic regression and the confidence interval (in grey) for Upper Paraná River basin.

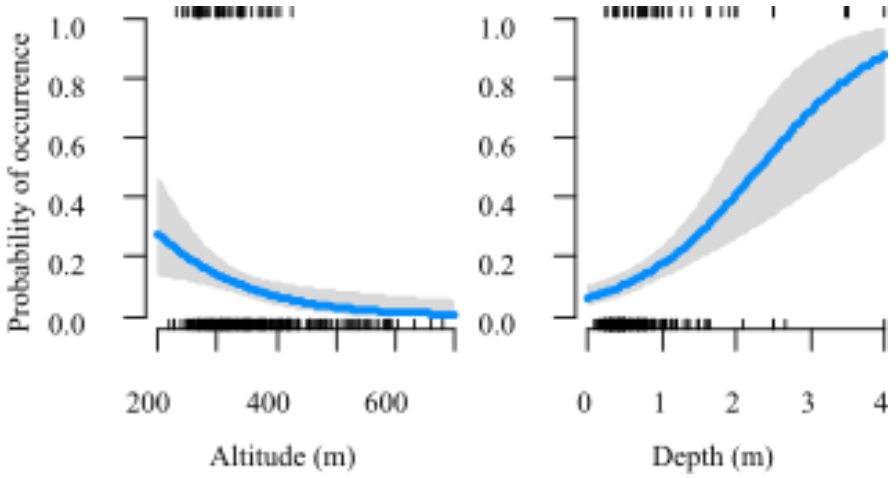

Figure 5. Relation among probability of occurrence of *L. friderici* and environmental descriptors selected by stepwise logistic regression and the confidence interval (in grey) for Upper Paraná River basin

The estimated probability of occurrence for the four fish species in Amambai and Ivinhema rivers showed that *P. lineatus* occur predominantly in main rivers, in middle and lower portions of the basins as also *H. eques* and *L. friderici*, however the last two species have a broad distribution in these basins when compared to *P. lineatus*. On the other hand, *P. argentea* have an inverse distribution pattern with higher occurrence probabilities in headwaters and faster streams (Figure 6).

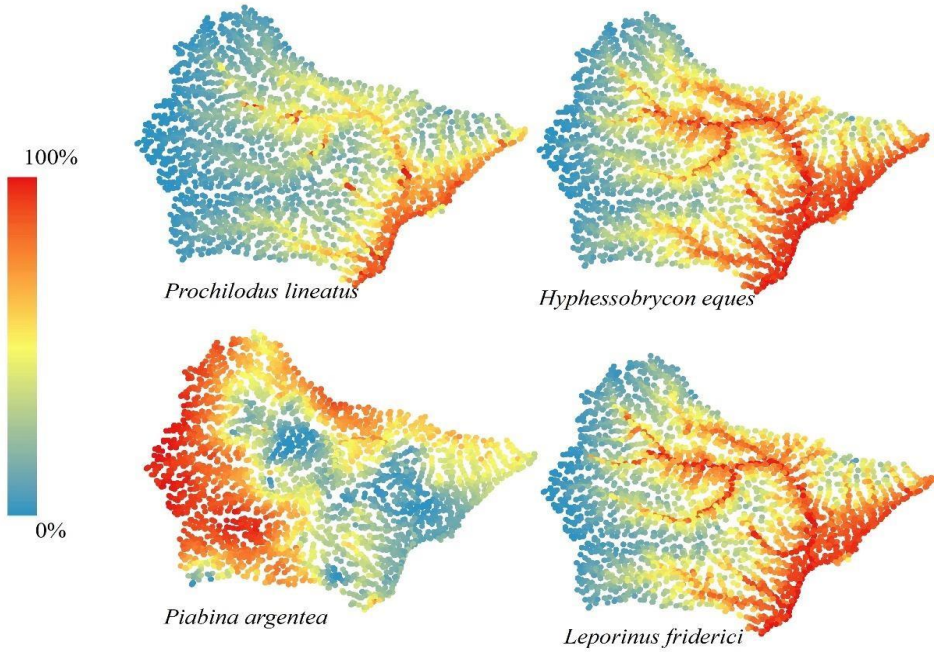

Figure 6. Estimated probability of occurrence of studied fish species on the Upper Paraná River Basin.

DISCUSSION

Prochilodus lineatus, commonly known as curimba, presented higher probabilities of occurrence in the main river areas. Our results suggest that the main determinant of your occurrence as river depth, and this result can be explained by two factors, the size of this species, that reaching to 80 cm and weighing more than 7 kg (Corbetta, 2018), and to a detritivore feeding, then your occurrence as also limited to availability of organic matter in sediment, more frequent in deeper and slower rivers. Besides the little, but not limited probability of occurrence of this species in upstream rivers is because this species is of migratory nature, and at certain times of the year it migrates to be able to spawn in the floodplains of the Upper Paraná basin.

Hyphessobrycon eques also commonly known as Mato Grosso and occur mainly in lower portions of the basins, living associated to aquatic macrophytes (Dias *et al*., 2017) as a strategy to avoid predators. Our results suggest that the occurrence is not restrict to floodplain areas and streams with larger volume can present adequate conditions to this species occurrence and probably to complete your life cycle (Lima, 2003; Nelson, *et al*. 2016).

Piabina argentea, due to the fact that it is a small-sized species measuring near to 7.3 cm of standard length (Lima, et al., 2003), are bentopelagic species, preferably in small rivers. of being species strongly dependent on allochthonous organic material that they use as a resource for their feeding (Gomiero & Braga, 2008) and is a strong species threatened by environmental disturbances such as deforestation of riparian vegetation (Kintopp & Avilhoa, 2017). Our data showed a preference for upper portion of the basins, these streams portions probably present more adequate conditions to this species occurrence, with larger riparian cover on streams, offering feeding resources (Sica, 2016) and higher oxygen concentration as response of higher water flow and altitude. *Piabina argentea* are frequently sampled in streams with moderate water velocity and located to higher altitudes (Valério *et al.*, 2007), and our results corroborate this result.

Leporinus friderici have certain similarity with the species *P. lineatus* in the spatial distribution; despite the differences in feeding ecology and size (the maximum registered size was 40cm and 1.5 kg) (Froeze & Pauly, 2018). Then this species presented a high probability of occurrence in main rivers and some secondary ones located to the middle and lower basins portions.

Considering all selected variables, the hydrological variables are considered the main descriptors of distribution of all analyzed species, excepting dissolved oxygen and water conductivity. The role of hydrological characteristics in species distribution in a hydrographic watershed is usually recognized on assemblage scale (Poff, 1997; MartinSmith, 1998; Hued & de los Angeles, 2002; Valério et al., 2007; Súarez *et al*., 2011). However, few studies in a neotropical region were made about species distribution in a hydrographic watershed scale and more effort are necessary to better understand the role of hydrological and limnological characteristics on species distribution.

ACKNOWLEDGEMENT

To the Organization of the American States (OAS) and FUNDECT by graduate scholarship. To FUNDECT, CNPq and Green Farm for the financial and logistical support. To IBAMA (SISBIO # 13458-1) for authorization to scientific collection.

REFERENCES

Alderete, A. M. (2006). Fundamentos del análisis de regresión logística en la investigación psicológica. *Revista Evaluar*, *6*.

Corbetta, D. D. F., Veronezzi, A. L., & Benedito, E. (2016). Body condition and energy density of juvenile streaked prochilod Prochilodus lineatus (Valenciennes, 1837) in a Neotropical floodplain. *Neotropical Ichthyology*, *14*(4).

Dias, R. M., da Silva, J. C. B., Gomes, L. C., & Agostinho, A. A. (2017). Effects of macrophyte complexity and hydrometric level on fish assemblages in a Neotropical floodplain. *Environmental Biology of Fishes*, *100*(6), 703-716.

Gardner B, Sullivan PJ, Lembo Jr, AJ (2003) Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. Can J Fish Aquat Sci 60:344–351.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters*, *8*(9), 993-1009.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, *135*(2-3), 147-186.

Gomiero, L. M., & Braga, F. M. D. S. (2008). Feeding habits of the ichthyofauna in a protected area in the state of São Paulo, southeastern Brazil. *Biota Neotropica*, *8*(1), 41-47.

Hued, A. C., & de los Ángeles Bistoni, M. (2002). Effects of water quality variations on fish communities in the Central Part of Argentina, South America. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen*, *28*(3), 1476-1481.

Kintopp, I., & Abilhoa, V. (2017). Ecologia alimentar de Piabina argentea Reinhardt, 1867 (Teleostei, Characidae) no Rio das Almas, São Paulo, Brasil. Estudos de Biologia, 31(73/75).

Langeani, F., Macedo Corrêa e Castro, R., Takeshi Oyakawa, O., Akio Shibatta, O., Simone Pavanelli, C., & Casatti, L. (2007). Diversidade da ictiofauna do Alto Rio Paraná: composição atual e perspectivas futuras. *Biota Neotropica*, *7*(3).Lima, F. C. T. (2003). Genera incertae sedis in Characidae. Check list of the freshwater fishes of South and Central America, 106-169.

Martin-Smith, KM. 1998. Relationships between fishes and habitat in rainforest streams in Sabah, Malaysia. Journal of Fish Biology, vol. 52, n. 3, p. 458-482

Nelson, J. S., Grande, T. C., & Wilson, M. V. (2016). *Fishes of the World*. John Wiley & Sons.

Poff, N. L. (1997). Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the north american Benthological society*, *16*(2), 391-409.

Reis. R. E, S.O. Kullander and C.J. Ferraris, Jr. (eds.) Checklist of the Freshwater Fishes of South and Central America. Porto Alegre: EDIPUCRS, Brasil.

Súarez, Y. R. (2008). Spatial and temporal variation in fish species diversity and composition in streams of Ivinhema River basin, upper Paraná River. Biota Neotropica, 8(3), 197204.

Súarez, Y. R., Souza, M. M. D., Ferreira, F. S., Pereira, M. J., Silva, E. A. D., Ximenes, L. Q. L., ... & Lima Júnior, S. E. (2011). Patterns of species richness and composition of fish assemblages in streams of the Ivinhema River basin, Upper Paraná River. *Acta Limnologica Brasiliensia*, *23*(2), 177-188.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369-373.

Valério, SB., Súarez, YR., Felipe, TRA., Tondato, KK. and Ximenes, LQL. 2007. Organization patterns of headwater-stream fish communities in the Upper ParaguayParaná basins. Hydrobiologia, vol. 583, 22. n. 1, p. 241-250.

Ver Hoef, J. M., Peterson, E., & Theobald, D. (2006). Spatial statistical models that use flow and stream distance. Environmental and Ecological statistics, 13(4), 449-464.

Ver Hoef, J. M., Peterson, E., Clifford, D., & Shah, R. (2014). SSN: An R package for spatial statistical modeling on stream networks. Journal of Statistical Software, 56(3), 1-45.

Sica, Y. V. (2016). Cambios en el uso del suelo y sus efectos a diferentes escalas espaciales y temporales sobre la diversidad de aves en el Bajo Delta del río Paraná (Doctoral dissertation, Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires).

CAPÍTULO 3 - CONSIDERAÇÕES FINAIS

No capítulo 1, de acordo com o levantamento bibliográfico sobre as metodologias utilizadas para a interpolação das métricas utilizadas, foi apresentada de forma mais detalhada a metodologia complementar de construção dos diferentes layers necessários ao procedimento analítico adotado e que geram informações da paisagem e locais a serem utilizadas, além de corrigir os pontos de amostragem para reduzir o erro causado pela falha na precisão do GPS. Os resultados obtidos com base nas metodologias do capítulo 1, para a interpolação das métricas, indicam que as profundidades e as altitudes influenciam os parâmetros físico-químicos da água e permitiram as análises utilizadas no capítulo 2.

No capítulo 2, as espécies avaliadas apresentaram um padrão de distribuição vinculado às variáveis hidrológicas e espaciais, como altitude, profundidade e largura dos riachos/rios, uma vez que são principalmente espécies de tamanho médio a grande (*L. friderici* e *P. lineatus*) e espécies que preferem baixa velocidade (*H. eques*). Com exceção de *P. argentea*, que é uma espécie de pequeno porte com preferência por correnteza.

Os resultados obtidos neste projeto ainda devem permitir a modelagem da ocorrência de outras espécies de peixes e a inserção de outras variáveis ambientais, gerando modelos de distribuição e diversidade potencial nas bacias avaliadas e permitindo identificar porções com elevada diversidade de espécies e com elevado potencial para conservação das espécies.