

FERRAMENTA BASEADA EM TÉCNICAS DE APRENDIZADO DE MÁQUINAS: QUAIS OS AVANÇOS PARA REABILITAÇÃO DE ECOSISTEMAS AQUÁTICOS?

Daniele Jovem-Azevêdo¹
Magnólia de Araújo Campos Pfenning²
Evaldo de Lira Azevêdo³
Joseline Molozzi⁴

RESUMO

Modelos preditivos frequentemente incorporam múltiplas variáveis e representam uma ferramenta útil para testar diferentes cenários de reabilitação. Aqui, objetivamos prever quais os efeitos sobre a comunidade de macroinvertebrados bentônicos podem ser esperados a partir da redução simulada no nível de eutrofização em reservatórios no semiárido. Um modelo preditivo foi construído a partir da seleção de três técnicas de aprendizado de máquinas: *Support Vector Machines*, *Multi-Layer Perceptron*, *K-nearest neighbor analysis*. O modelo foi elaborado a partir de dados ambientais e biológicos de 129 sites amostrados em seis reservatórios (bacia hidrográfica do rio Piranhas-Assu e rio Paraíba, Nordeste do Brasil). As variáveis associadas a eutrofização foram alteradas artificialmente para a simulação em dois cenários de reabilitação: D1 = menor nível de melhoria, 25% de alteração; e D2 = maior nível de melhoria, 75% de alteração. Três grupos de gradientes ambientais foram considerados previamente com base na PCA: sites com menor nível de impacto, nível de impacto intermediário e nível de impacto severo. As simulações indicam melhoria ambiental e aumento da riqueza total nos cenários D1 e D2 para todos os grupos de gradientes ambientais selecionados ($p = 0,0001$). Além disso, sob as novas condições simuladas, a expansão de *taxa* sensíveis (Ephemeroptera, Trichoptera e gênero de Chironomidae) também é esperada. As simulações indicam que medidas de melhorias aplicadas a sistemas artificiais, como reservatórios, podem de fato representar melhoria ambiental e biótica, mostrando ser um passo potencial para elaboração de propostas de manejo e restauração de recursos hídricos, sobretudo em região semiárida.

Palavras-chave: Semiárido, Reservatórios, Macroinvertebrados bentônicos, Modelos preditivos.

¹ Doutora em Ecologia, Conservação e Manejo da Vida Silvestre pela Universidade Federal de Minas Gerais – MG; Pós-Doutoranda no Programa de Pós-Graduação em Ciência Naturais e Biotecnologia da Universidade Federal de Campina Grande, *Campus* Cuité - PB, daniele.jazevedo@gmail.com;

² Doutora em Ciências Biológicas (Biologia Molecular) pela Universidade de Brasília – DF; Professora Adjunta da Universidade Federal de Campina Grande, *Campus* Cuité - PB, magnoliacp@gmail.com;

³ Doutor em Etnobiologia e Conservação da Natureza pela Universidade Federal Rural de Pernambuco – PE; Professor DI-2 do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, *Campus* Cajazeiras - PB, evaldo.azevedo@ifpb.edu.br;

⁴ Doutora em Ecologia, Conservação e Manejo da Vida Silvestre pela Universidade Federal de Minas Gerais – MG; Professora Adjunta da Universidade Estadual da Paraíba, *Campus* Campina Grande - PB, jmolozzi@gmail.com;

INTRODUÇÃO

Modelos preditivos são considerados ferramentas úteis tanto no contexto da bioavaliação de ecossistemas aquáticos, quanto a proposição de estratégias de manejo do recursos hídricos (STRACHAN; REYNOLDSON, 2014; ASHOFTEH; RAJAEI; GOLFAM et al., 2017). Frequentemente, esses modelos são baseados em análises discriminantes, as quais formam grupos com base em dados biológicos e ambientais a partir da semelhança entre os sites analisados, o que permite avaliar a qualidade biológica dos locais a partir do desvio da comunidade observada, comparada a uma comunidade esperada sob condições de referência (FEIO et al., 2009).

Os modelos mais comuns baseados na abordagem de referência são: RIVPACS (*River Invertebrate Prediction Classification System* - WRIGHT; SUTCLIFFE; FURSE, 2000), AUSRIVAS (*Australian River Assessment Scheme* - SIMPSON; NORRIS, 2000) e BEAST (*Benthic Assessment of Sediment* - REYNOLDSON et al., 1995), todos baseados em análises de funções discriminantes. Esses modelos já têm sido utilizados para avaliação de ecossistemas aquáticos em todo o mundo (ex. CLARKE; WRIGHT; FURSE et al., 2003; FEIO et al., 2007; NICHOLS; REYNOLDSON; HARRISON, 2014; STRACHAN; REYNOLDSON, 2014) e já integram as ferramentas aplicadas na Diretiva Quadro da Água da União Européia (EUROPEAN COMMISSION, 2000). Apesar de amplamente utilizadas, ferramentas com base em análise discriminante podem gerar agrupamentos artificiais e, além disso, requerem variáveis com distribuição normal, sem relação colinear e um grande número de dados para serem analisados.

Diferentemente, modelos baseados em técnicas de aprendizado de máquinas (*machine learning techniques*) são capazes de prever a distribuição de espécies sem a necessidade de um conjunto de dados de referências, pois são ferramentas hábeis na captura de relações não-lineares, sendo também pouco influenciadas por distribuição distorcida das variáveis analisadas (*outliers*), diferentemente das técnicas de análise discriminantes tradicionais (GEVREY et al., 2004). Entre as técnicas mais utilizadas em estudos ecológicos estão: *Support Vector Machines* (SVM, DRAKE et al., 2006), *Multi-Layer Perceptron* (MLP; GEVREY et al., 2004), *K-nearest neighbor analysis* (KNN; FEIO; VIANA-FERREIRA; COSTA 2014a). Todas as técnicas de aprendizado de máquinas baseiam-se no conhecimento preexistente (a partir do conjunto de dados observados) para o desenvolvimento de previsões. Assim, mostram habilidade em prever a condição esperada a partir do desvio da razão entre a condição observada e a esperada (FEIO et al., 2014a).

Apesar da grande evolução das ferramentas multivariadas, o emprego dessas no contexto da reabilitação de recursos hídricos ainda é tímido (ex., LANGHANS; REICHERT; SCHUWIRTH, 2014; MUHAR et al., 2016). Isso pode ser explicado pela ausência de objetivos que conecte as medidas de avaliação dos ecossistemas com as medidas de efeitos em programas de reabilitação, o que torna ainda mais urgente o desenvolvimento de ferramentas eficientes e de baixo custo que auxiliem tanto na avaliação, quanto na proposição de medidas eficientes para reversão do estado de degradação. Nesse contexto, ferramentas de simulação podem ser peças-chave no sucesso de reabilitação, pois incorporam informações de variáveis múltiplas, permitindo testar diferentes cenários de reabilitação (HERMOSO et al., 2011; BRUDVIG, 2017). Vale ressaltar que, em muitos casos, essas ferramentas ainda têm sido aplicadas de forma restrita, usando apenas táxons específicos ou métricas de diversidade (JOY; DEATH, 2004; MÉRIGOUX et al., 2015; ROSE et al., 2016; LAUGHLIN et al., 2017).

No presente estudo, nós empregamos simultaneamente três técnicas de aprendizado de máquinas: SVM, MLP e KNN, através da aplicação da ferramenta HYDRA (FEIO; VIANA-FERREIRA; COSTA 2014a, b). Nosso principal objetivo foi predizer quais os efeitos sobre a comunidade de macroinvertebrados bentônicos podem ser esperados a partir da redução simulada no nível de eutrofização em reservatórios no semiárido, buscando a proposição de medidas de reabilitação. Para isso, construímos um modelo a partir de dados bióticos e abióticos que incluem diferentes gradientes ambientais, opondo-se a modelos preditivos tradicionais, baseados em condições de referência. Essa abordagem segue o conceito de *dirty-water model* proposto por Hoang et al. (2003).

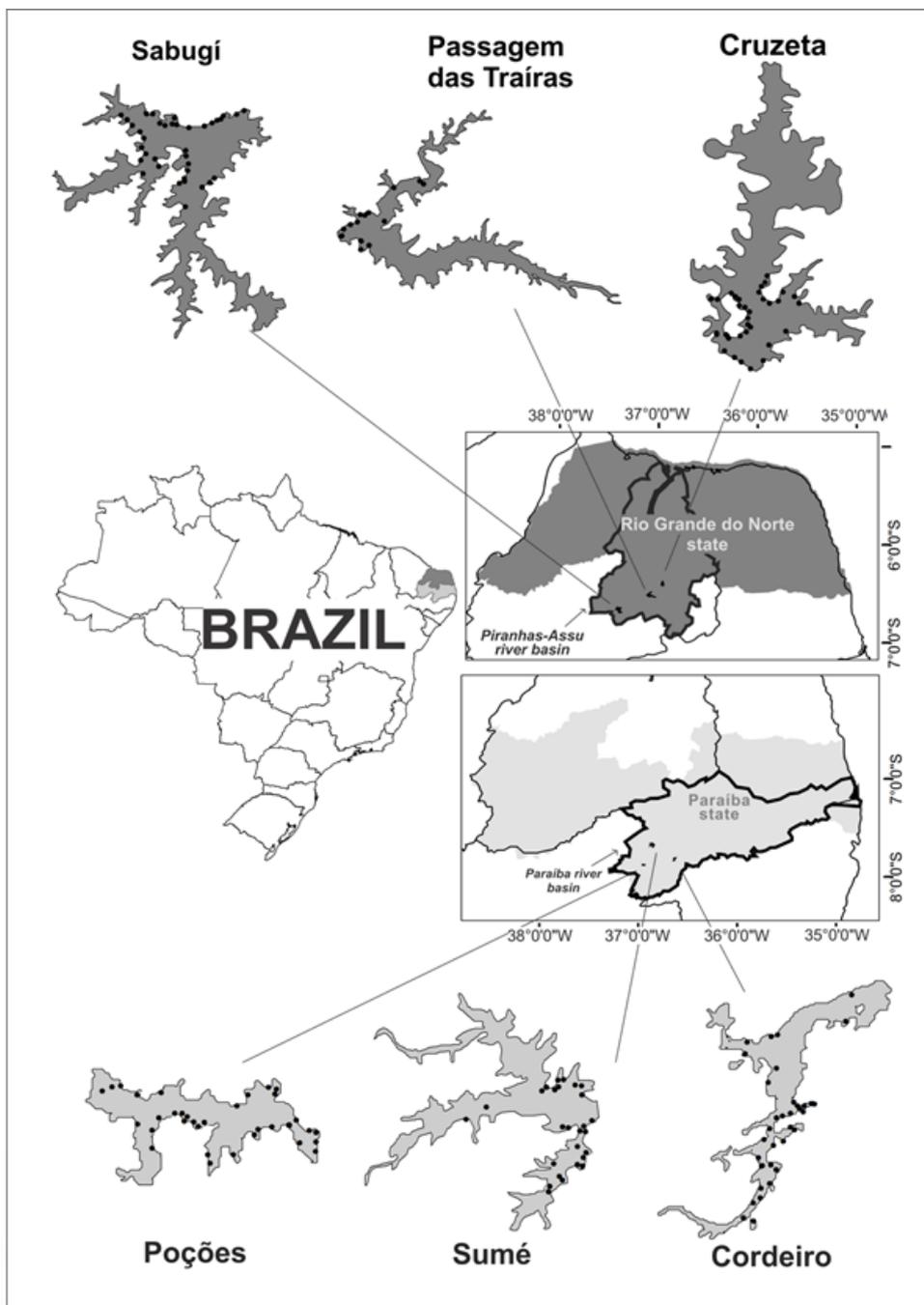
METODOLOGIA

Conjunto de dados

O conjunto de dados está baseado em 129 sites distribuídos na região litorânea de 6 reservatórios inseridos em 2 bacias hidrográficas localizadas no semiárido brasileiro: rio Piranhas-Assu (reservatórios Sabugí, Passagem das Traíras e Cruzeta - estado do Rio Grande do Norte) e rio Paraíba (reservatórios Cordeiro, Sumé e Poções - estado da Paraíba), Nordeste do Brasil (Figura 1). Todos esses reservatórios são destinados ao abastecimento público na região. As amostragens em cada site ocorreram em quatro períodos, sendo: Junho, Setembro e Dezembro de 2014 e Março de 2015. A região em que as duas bacias hidrográficas estão inseridas é classificada como BSh semiárido quente (Köppen–Geiger), com período seco

entre 9-10 meses/ano. Na região da bacia hidrográfica do rio Paraíba é registrado o menor nível pluviométrico (400mm/ano), quando comparada a bacia hidrográfica do rio Piranhas-Assu (800mm/ano). A temperatura mínima média em ambas as bacias hidrográficas é de 20°C (entre Julho e Agosto) e máxima média de 29,5°C (entre Novembro e Dezembro) (ALVARES et al., 2014).

Figura 1: Distribuição dos sites de amostragem nos reservatórios Sabugí, Passagem das Traíras, Cruzeta (bacia hidrográfica do rio Piranhas-Assu), Cordeiro, Sumé e Poções (bacia hidrográfica do rio Paraíba) (Figura a partir de Jovem-Azevêdo et al., 2019).



Cada local de amostragem foi caracterizado quanto aos aspectos físicos e químicos da água, aspectos do habitat e da paisagem. Entre os aspectos físicos e químicos da água foram medidas as seguintes variáveis: temperatura (°C), pH, turbidez (NTU), oxigênio dissolvido (mg/L) e sólidos totais dissolvidos, utilizando sonda multiparâmetro (Horiba-U50). A profundidade foi avaliada com auxílio de aparelho sonar. A transparência da água foi mensurada de acordo com o desaparecimento do disco de Secchi. Além disso, em cada local, um litro de água foi coletado na subsuperfície e, em laboratório, foram determinadas as concentrações de nutrientes fosfatados (fósforo total e fosfato solúvel reativo - µg/L) e nitrogenados (íons amônio, nitrato e nitrito - µg/L) (APHA, 2005). A concentração da clorofila foi avaliada através da extração do pigmento em acetona 90% (LORENZEN, 1967).

Para caracterização do habitat de colonização das espécies de macroinvertebrados, amostras do sedimento foram coletadas utilizando draga de Eckman-Birge (225cm²) e direcionadas para a determinação do teor de matéria orgânica pelo método gravimétrico e composição granulométrica do sedimento, seguindo o método proposto por Suguio (1973), modificado por Callisto & Esteves (1996). As partículas do sedimento classificadas de acordo com as categorias propostas por Jovem-Azevêdo et al. (2019): cascalho (>1mm), areia grossa (500-1000µm), areia média (250-500 µm), areia fina (125-250 µm), silte (63-125 µm) e argila (<63 µm).

A paisagem ao entorno também foi avaliada para determinação do uso e ocupação antrópica, seguindo o protocolo de avaliação de habitats *River Habitat Survey*, modificado por Rowan et al. (2006). Através desse protocolo foi possível analisar: 1) características associadas ao desenvolvimento urbano (ex. presença de residências, cerca de demarcação, linhas de transmissão) e 2) atividades agrícolas (ex. áreas de pastagem, plantio de leguminosas e fruteiras e criação de rebanho). Dados relacionados a localização, precipitação e geologia de cada área foram compilados a partir dos dados disponíveis nas agências oficiais de meio ambiente: Secretaria de Meio Ambiente e Recursos Hídricos no estado Rio Grande do Norte e Agência Executiva de Gestão das Águas no estado da Paraíba.

A comunidade de macroinvertebrados bentônicos foi coletada com draga de Eckman-Birge (225cm²). Ainda *in situ*, as amostras foram fixadas com solução de formoldeído a 4%. Em laboratório, as amostras foram lavadas em peneiras com malha de abertura de 500µm e conservadas em álcool a 70%. Posteriormente, as amostras foram triadas para retirada dos organismos. Todos os indivíduos encontrados foram contados e identificados a nível de família, exceto as larvas da família Chironomidae, as quais foram identificadas a nível de gênero (BOFFI, 1979; PÉRES, 1988; TRIVINHO-STRIXINO, 2011).

Análise de Dados

Classificação dos gradientes ambientais

Os dados abióticos foram utilizados para determinar os gradientes ambientais a que estão submetidos os sites amostrados. Para isso, empregamos *Principal components analysis* (PCA) a fim de avaliar a segregação entre os sites pelos descritores ambientais. Previamente, nós analisamos a distribuição dos dados a fim de detectar aquelas variáveis com padrão não-normal de distribuição e aquelas que estivessem altamente correlacionadas (>0.75). As variáveis com distribuição não-normal foram transformadas em $\log(x+1)$ e as variáveis dos habitat transformadas em arco-seno. As variáveis altamente correlacionada foram excluídas. Posteriormente todas as variáveis foram normalizadas. Três grupos foram formados a partir da PCA, de acordo com a similaridade das variáveis ambientais, sendo classificados: 1) sites com menor nível de impacto ambiental (LI); 2) nível de impacto intermediário (ID); e 3) nível severo de impacto (SI). Para testar se os grupos formados diferem significativamente entre si, aplicamos *Permutational Multivariate Analysis of Variance* (PERMANOVA, com coeficiente de Distância Euclidian; 9999 permutações; ANDERSON; GORLEY; CLARKE, 2008).

Construção do modelo dirty-water

Para a construção do modelo, selecionamos três técnicas baseadas no aprendizado de máquinas (SVM, MLP e KNN) e as empregamos através da ferramenta HYDRA (FEIO; VIANA-FERREIRA; COSTA 2014a, b) desenvolvida em linguagem R (pacote *Rminer*, CORTEZ, 2010) e implementada na plataforma AQUAWEB (aquaweb.uc.pt). A partir do modelo, dados de referência são selecionados (dados teste), os quais são posteriormente utilizados para comparação com dados de sites mais próximos no espaço dimensional, de acordo com os preditores ambientais (dados de treinamento) (ver FEIO; VIANA-FERREIRA; COSTA 2014a, b para maiores detalhes). Para maior acurácia na predição, nós selecionamos apenas as probabilidade de ocorrência de um *taxa* esperado a partir do observado ≥ 0.5 . A razão entre o observado e esperado (O/E) foi utilizada para determinar o nível de comprometimento biótico e o desvio padrão de O/E para avaliar a acurácia das técnicas.

Simulação de reabilitação ambiental

Para testar um cenário de melhoria ambiental nos reservatórios estudados, nós alteramos artificialmente os valores das variáveis associadas a eutrofização, considerando um cenário hipotético de interrupção da entrada e/ou tratamento de esgoto urbano,

frequentemente descartados de forma imprópria em rios e reservatórios em todo o mundo (WHO, 2015; UNESCO, 2017). Nossas simulações foram projetadas em dois cenários de melhoria: D1 (menor nível de melhoria) = variáveis reduzidas em 25%; D2 (maior nível de melhoria) = variáveis reduzidas em 75%.

Assim, nós reduzimos (conforme D1 e D2) as concentrações de fósforo, fosfato solúvel reativo e clorofila-*a*. De forma contrária, nós aumentamos os valores de transparência, considerando que ambientes com menor nível de trofia apresentam águas claras. Adicionalmente, nós também reduzimos variáveis que estão diretamente associadas a eutrofização: oxigênio dissolvido, turbidez e sólidos totais dissolvidos. Todas essas variáveis seguem o cenário de melhoria proposta em D1 e D2, exceto o oxigênio dissolvido, o qual foi reduzido apenas em sites com valores acima de 7 mg/L, isso porque, a redução do oxigênio abaixo de 4 mg/L indica uma condição de anoxia (CONAMA 357/05). Posteriormente, nós aplicamos o novo conjunto de dados ao modelo construído previamente, a fim de avaliar as mudanças esperadas a partir das medidas de reabilitação simuladas. Para avaliar se há diferenças significativas nas variáveis ambientais observadas e esperadas (após as mudanças simuladas), empregamos uma série de PERMANOVA's (com coeficiente de Distância Euclidian; 9999 permutações).

Complementarmente, para avaliar se a melhoria biótica esperada resulta na expansão de colonização de táxons, nós selecionamos três *taxa* sensíveis mais abundantes: Caenidae (Ephemeropt com coeficiente de Distância Euclidian; 9999 permutações era), Polycentropodidae (Trichoptera) e *Zavreliella* sp (Diptera: Chironomidae) comparando o observado e o esperado após a aplicação dos dois cenários de melhoria. Mapas de distribuição foram elaborados, a fim de analisar a expansão na colonização esperada desses organismos após a reabilitação simulada. Por fim, empregamos análises de PERMANOVA's (com coeficiente de Distância Euclidian; 9999 permutações) para determinar se houve diferenças significativas na riqueza total de espécies observada e esperada a partir dos cenários de reabilitação.

RESULTADOS E DISCUSSÃO

Caracterização dos sites

Os três primeiros eixos da PCA, baseada nos dados abióticos, explicou um total 52,57% da variabilidade total dos dados. Baseada na distribuição dos dados, 30 sites foram incluídos no grupo LD, 54 no grupo ID e 45 sites no grupo SD. As maiores concentrações de fósforo ($313.5 \pm 107.9 \mu\text{g/L}$), clorofila-*a* ($58.0 \pm 51.3 \mu\text{g/L}$), turbidez ($104.6 \pm 68.9 \text{NTU}$), oxigênio dissolvido ($7.8 \pm 1.6 \text{mg/L}$) foram registradas entre os sites do grupo SI. Além disso, a menor transparência da água ($0.3 \pm 0.1 \text{m}$) também ocorreu nesses sites (Tabela 1), indicando que esses locais estão sob alta condição de degradação. Diferenças significativas ocorreram entre para as variáveis ambientais analisadas entre os grupos LD, ID e SI (PERMANOVA: Pseudo- $F_{2,128} = 25,7$; $p = 0,0001$).

Tabela 1: Média e desvio padrão (\pm) das variáveis ambientais analisadas nos reservatórios Sabugí, Passagem das Traíras e Cruzeta (bacia hidrográfica do Piranhas-Assu), Cordeiro, Sumé e Poções (bacia hidrográfica do rio Paraíba). Onde: LD = sites com menor nível de impacto ambiental; LI = nível de impacto intermediário; SI = nível de impacto severo.

Variáveis	LD	ID	SI
Latitude	-37.2 \pm 0.2	-36.9 \pm 0.17	-36.9 \pm 0.09
Longitude	-6.7 \pm 0.4	-7.6 \pm 0.4	-6.98 \pm 0.7
Precipitação	33.7 \pm 7.5	12.5 \pm 12.2	31.3 \pm 3.9
Altitude	223.9 \pm 97.8	473.1 \pm 114.6	360.3 \pm 185.0
Litologia	1.33 \pm 0.6	3.2 \pm 0.9	2.4 \pm 0.5
Profundidade	0.6 \pm 0.1	0.7 \pm 0.1	0.6 \pm 0.2
Secchi	0.5 \pm 0.1	0.5 \pm 0.1	0.3 \pm 0.1
Turbidez	19.9 \pm 10.9	49.0 \pm 45.8	104.6 \pm 68.9
Oxigênio dissolvido	5.8 \pm 1.0	6.7 \pm 1.1	7.8 \pm 1.6
Sólidos totais dissolvidos	0.43 \pm 0.2	0.9 \pm 0.5	1.5 \pm 0.9
Nitrogênio	31.3 \pm 20.0	45.8 \pm 26.9	37.8 \pm 28.8
Íon amônio	32.7 \pm 16.0	53.8 \pm 29.1	79.8 \pm 48.6
Fósforo total	106.2 \pm 62.4	123.8 \pm 77.8	313.5 \pm 107.9
Fosfato solúvel reativo	39.9 \pm 48.7	43.1 \pm 33.2	111.7 \pm 65.9
Chlorofila- <i>a</i>	12.8 \pm 11.0	16.3 \pm 25.8	58.0 \pm 51.3
Matéria orgânica	0.2 \pm 0.1	0.3 \pm 0.1	0.2 \pm 0.1
Cascalho	1.3 \pm 0.2	1.4 \pm 0.1	1.5 \pm 0.1
Areia grossa	1.4 \pm 0.09	1.4 \pm 0.09	1.4 \pm 0.09
Areia média	1.39 \pm 0.08	1.3 \pm 0.07	1.4 \pm 0.08
Areia fina	1.33 \pm 0.07	1.3 \pm 0.07	1.3 \pm 0.08
Silte	1.44 \pm 0.05	1.4 \pm 0.07	1.3 \pm 0.07
Argila	1.51 \pm 0.04	1.5 \pm 0.04	1.5 \pm 0.04
Cobertura de residência	0.2 \pm 0.2	0.2 \pm 0.2	0.5 \pm 0.36
Cobertura de pastagem	0.3 \pm 0.2	0.2 \pm 0.2	0.4 \pm 0.3
Cobertura agrícola	0.1 \pm 0.2	0.2 \pm 0.2	0.4 \pm 0.3

Na comunidade de macroinvertebrados bentônicos, um total de 61 *taxa* foi identificado, com 77.400 indivíduos contatos. O grupo mais representativo foi Insecta (45 táxons), seguido de Mollusca (8 táxons), Annelida (2 táxons), Archnida (2 táxons), Crustacea (1 táxon), Nematoda (1 táxon) e Plathelminthes (1 táxon). O número da composição taxonômica entre os grupos dos gradientes ambientais, LI, ID e SI, foram similares: 45, 43 e 43 *taxa*, respectivamente (Tabela 2). No entanto, a composição dos táxons foi significativamente diferente entre os grupos (PERMANOVA: Pseudo- $F_{2,128} = 19,6$; $p = 0,0001$).

Modelo *dirty-water*

Do total de *taxa* identificado, 17 foram excluídos para a construção do modelo. Esses previamente retirados, correspondem aos táxons com baixa representatividade para a comunidade, abundância relativa $< 0,03$. O modelo construído mostrou alta predição e acurácia com médias similares entre as técnicas de aprendizado de máquinas selecionadas, sendo: SVM = 0,84. MLP = 0,73 e KNN = 0,82; e desvio padrão variando entre 0,14-0,18. A acurácia de precisão foi $<$ que 50% apenas para um *taxa* (Oligochaeta), nas técnicas SVM e KNN e para cinco *taxa* (*Melanoides tuberculata*, Oligochaeta, *Aedokritus* sp, *Coelotanypus* sp e *Goeldchironomus* sp), na técnica MLP. Para os demais *taxa* a acuraria foi $>$ que 50%. Nas técnicas SVM e KNN particularmente, a predição de acurácias foi acima de 0,9 para 12 *taxa*. Contrariamente, a técnica MLP mostrou-se menos sensível na predição de *taxa* com ocorrência $>0,58$ (Figura 2). A média entre O/E foi $0,79 (\pm 0,17)$ e $R^2 = 0,89$.

A aplicação simultânea das três técnicas de aprendizado de máquinas mostrou alta precisão de predição, considerando que a acurácia de predição foi $> 0,7$ para todas as técnicas selecionadas. No entanto, temos observado que especialmente na técnica MLP houve baixa acurácia para cinco táxons. A baixa acurácia encontrada nesse caso, pode ser relacionada a frequência de ocorrência dos *taxa*, considerando que o alto ou baixo nível de ocorrência de um táxon pode afetar rede de arquitetura, especialmente para técnicas de aprendizado baseadas em redes simples (DEDECKER et al., 2004).

Assim, para os casos de táxons raros ou muito frequentes, modelos de rede complexa fornecem predições melhores ou mais robustas, assim como temos observado para as técnicas SVM e KNN. Estudo prévio de Feio et al. (2014a), analisando de forma comparativa as três técnicas de aprendizado utilizadas no presente estudo (SVM, MLP e KNN) também demonstra que a técnica MLP é menos acurada para um conjunto de dados com sites limitados e *taxa* com alta frequência de ocorrência, corroborando com nossos achados.

Simulação dos cenários de reabilitação

Os cenários simulados (D1 e D2) indicam que as mudanças nas variáveis associadas a eutrofização resultou na melhoria da qualidade da água e, conseqüentemente, na qualidade biótica dos sites analisados, como podemos observar pelo aumento da riqueza total esperada (Tabela 2). A melhoria na riqueza total da comunidade macroinvertebrados foi observada tanto para o cenário D1 quanto para o cenário D2, no entanto, o maior nível de melhoria foi encontrado para o cenário D2 em todos os grupos de gradientes ambientais analisados (LD, ID e SI). Diferenças significativas foram observadas para a comunidade sob o cenário D1 e D2 (PERMANOVA: Pseudo- $F_{2,386} = 127,9$, $p = 0,0001$).

Analisando a cobertura dos táxons sensíveis, temos encontrado que Caenidae aumentou consideravelmente a sua distribuição para os sites LD (sem registro de ocorrência observada para 15 sites com presença esperada), sites ID (de 5 sites observados para 22 sites esperados) e SD (de 1 site observado para 8 sites esperados). A expansão na colonização também é esperada para Polycentropodidae: em LD (de 5 sites observados para 24 esperados), ID (0 sites observados para 3 esperados) e SD (de 1 site observado para 7 esperados). Contrariamente, o aumento na colonização do gênero *Zavreliela* sp é esperado apenas para os sites LD e SI (de 2 sites observados para 7 esperados; de 2 sites observados para 12 esperados, respectivamente), não tendo sido observada nem esperada a colonização para os sites com nível intermediário de impacto (ID) (Figuras 3-5).

Nossos resultados mostram que sob a condição de redução da eutrofização, melhorias na riqueza são esperadas tanto em sites sob menor nível de impacto (LD), quanto para sites sob alto nível de impacto (SI), mesmo que nesses últimos as alterações ocorram em menor escala. Isso sugere que para sites sujeitos a intensa condição de impacto, como temos observado em nosso conjunto de dados, outras medidas associadas possam ser empregadas para alcance de um estado de reabilitação mais satisfatório (PALMER; HONDULA; KOCH, 2014; MUHAR et al., 2016).

Além disso, os cenários de simulações também demonstram que o quadro de melhoria no nível de eutrofização fornece condições ambientais favoráveis para a colonização de táxons com menor amplitude de nicho, antes com distribuição espacial restrita. Temos observado, que no maior nível de melhoria (D2) ocorre o aumento na riqueza total e expansão da distribuição esperada de táxons sensíveis, a exemplo de Caenidae, Polycentropodidae e *Zavreliela*. Esses resultados são reforçados pelo fato de que distúrbios ambientais de alta frequência e intensidade impõe riscos as espécies especialistas, considerando a degradação do habitat e a redução na disponibilidade de recursos, afetando o crescimento populacional.

Enquanto que espécies generalistas são tolerantes a diferentes gradientes ambientais, com menor nível de exigência de recursos, capazes de colonizar habitats vizinhos e atingir altas proporções numéricas mesmo em sites sob condição de impacto (BÜCHI; VUILLEUMIER, 2014).

Tabela 2: Composição da comunidade de macroinvertebrados nos reservatórios da bacia hidrográfica do rio Piranhas-Assu e Paraíba, semiárido brasileiro. Onde: LD = sites com menor nível de impacto; ID = nível de impacto intermediário; SI = nível de impacto severo; N° = número de sites em que foi registrada ocorrência do táxon.

Taxa	LD	N° sites	ID	N° sites	SI	N° sites
ANNELIDA						
Hirudinea	*	0	1	25	1	10
Oligochaeta	1	30	1	54	1	45
ARACNÍDEO						
Acarina	*	0	1	1	*	0
Oxidae	*	0	1	2	*	0
CRUSTÁCEO						
Decapoda	*	0	1	11	1	5
Ostracoda	1	18	1	2	1	7
INSECTA						
Chironomidae						
Chironominae						
<i>Aedokritus</i> (Roback, 1958)	1	25	1	15	1	36
<i>Asheum</i> (Sublette, 1964)	1	29	1	6	1	23
<i>Chironomus</i> (Meigen, 1803)	1	26	1	8	1	25
<i>Cladopelma</i> (Kieffer, 1921)	1	3	*	0	1	2
<i>Dicrotendipes</i> (Kieffer, 1913)	1	21	*	0	1	16
<i>Fissimentum</i> (Cranston and Nolte, 1996)	1	23	1	6	1	6
<i>Goeldichironomus</i> (Fittkau, 1965)	1	29	1	25	1	35
<i>Parachironomus</i> (Lenz, 1921)	1	6	1	8	1	6
<i>Pelomus</i> (Reis, 1989)	1	23	1	5	1	11
<i>Polypedilum</i> (Kieffer, 1912)	1	30	1	17	1	31
<i>Saetheria</i> (Jackson, 1977)	1	1	*	0	*	0
<i>Tanytarsus</i> (Van der Wulp, 1874)	1	30	1	13	1	28
<i>Zavreliella</i> (Kieffer, 1920)	1	2	*	0	1	2
Tanypodinae						
<i>Ablabesmyia</i> (Johannsen, 1905)	1	10	1	7	1	4
<i>Brundiniella</i> (Roback, 1978)	1	2	*	0	*	0
<i>Clinotanypus</i> (Kieffer, 1913)	*	0	1	2	1	1
<i>Coelotanypus</i> (Kieffer, 1913)	1	28	1	37	1	24
<i>Denopelopia</i> (Roback and Rutter, 1988)	*	0	1	1	*	0
<i>Djalmabatista</i> (Fittkau, 1968)	1	14	1	3	1	12
<i>Larsia</i> (Fittkau, 1962)	1	6	1	4	1	4
<i>Monopelopia</i> (Fittkau, 1962)	*	0	1	1	*	0
<i>Parapentaneura</i> (Fittkau and Serrano, 2006)	*	0	1	1	*	0
<i>Procladius</i> (Skuse, 1889)	1	3	*	0	1	4
<i>Tanypus</i> (Meigen, 1803)	1	9	*	0	1	2
Coleoptera						
Districidae	1	1	*	0	*	0
Elmidae	*	0	*	0	1	1
Gyrinidae	1	7	1	1	1	8
Diptera						
Chaoboridae	1	5	1	1	1	8
<i>Chaoborus</i> (Lichtenstein, 1980)	1	4	1	3	1	5

Ceratopogonidae	1	28	1	16	1	25
Ephemeroptera						
Baetidae	1	1	1	2	1	1
Caenidae	*	0	1	5	1	1
Polymitarcyidae	*	0	1	10	1	1
Hemiptera						
Belostomatidae	*	0	1	1	*	0
Corixidae	1	5	1	2	1	5
Heteroptera	1	1	*	0	*	0
Odonata						
Coenagrionidae	1	5	1	7	1	3
Dicteriadidae	1	1	*	0	*	0
Gomphidae	1	10	1	5	1	4
<i>Phyllocycla</i>	1	3	1	2	1	1
<i>Progomphus</i>	1	1	*	0	*	0
Libellulidae	1	2	1	9	1	2
Trichoptera						
Leptoceridae	1	1	*	0	*	0
Polycentropodidae	1	5	*	0	1	1
MOLLUSCA						
Gastropoda						
Ancylidae	1	7	*	0	1	1
Bulimidae	1	1	*	0	1	6
Planorbidae	1	12	1	25	1	9
Ampullariidae -						
<i>Pomacea</i>	1	17	1	8	1	7
Thiaridae -						
<i>Melanoides tuberculatus</i> (Müller, 1774)	1	29	1	54	1	42
Lymnaeidae	*	0	1	1	*	0
Bivalve						
Corbiculidae -						
<i>Corbicula largillierti</i> (Philippi, 1844)	*	0	1	19	*	0
Sphaeriidae	*	0	1	5	*	0
NEMATODA	1	1	*	0	*	0
PLATYHELMINTHES						
Planariidae	1	10	1	19	1	2
Mean richness/site						
Observed		6.7±2.0		5.8±2.0		4.7±1.9
D1		7.8±1.9		6.5±2.0		5.3±2.2
D2		10.6±1.5		8.6±2.7		7.7±2.5

Figura 2: Média da acurácia de predição para cada táxon identificado nos reservatórios Sabugí, Passagem das Traíras, Cruzeta (bacia hidrográfica do rio Piranhas-Assu), Cordeiro, Sumé e Poções (bacia hidrográfica do rio Paraíba). Linha pontilhada indica a acurária > 0,5 para as técnicas de aprendizado de máquinas selecionadas: SVM, MLP e KNN.

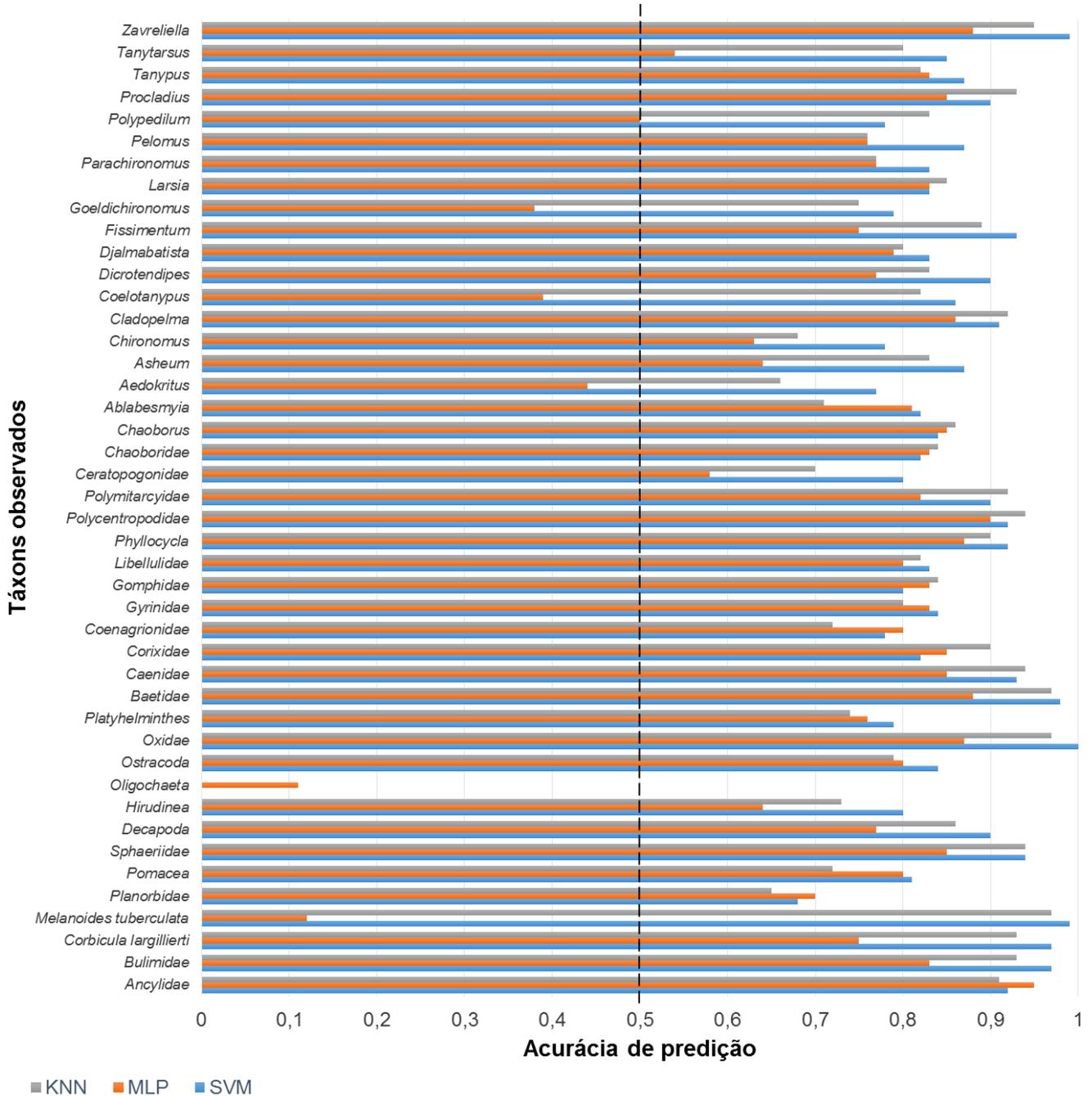


Figura 3: Distribuição espacial esperada de organismos da família Caenidae baseado no cenário D2 de melhoria nos reservatórios Sabugí, Passagem das Traíras, Cruzeta (bacia hidrográfica do rio Piranhas-Assu), Cordeiro, Sumé e Poções (bacia hidrográfica do rio Paraíba). Círculos pretos representam os sites de amostragem; círculos amarelos os sites onde o táxon foi observado; círculos vermelhos os sites onde o táxon é esperado.

Taxa: Caenidae

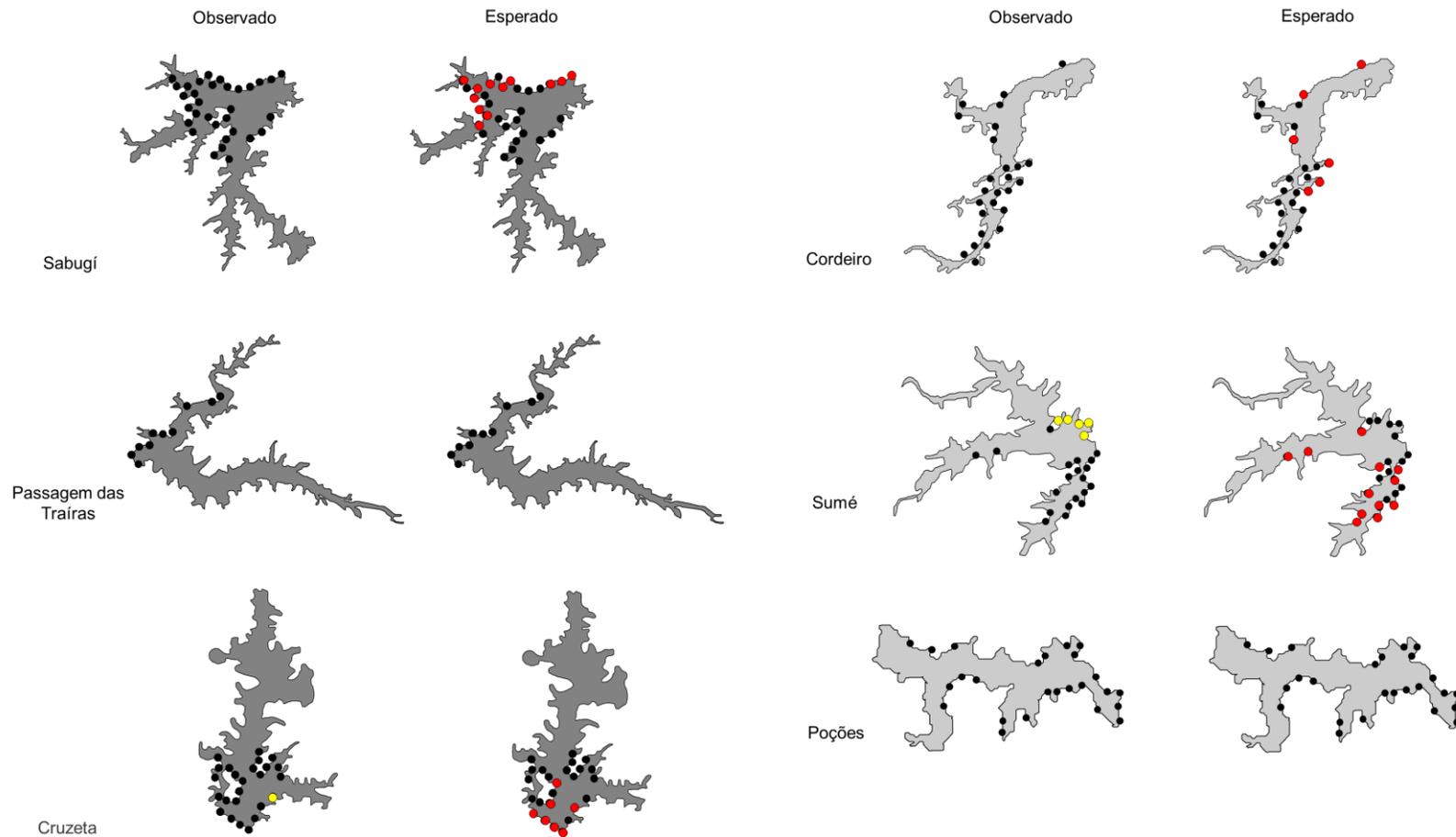


Figura 4: Distribuição espacial esperada de organismos da família Polycentropodida baseado o cenário D2 de melhoria nos reservatórios Sabugí, Passagem das Traíras, Cruzeta (bacia hidrográfica do rio Piranhas-Assu), Cordeiro, Sumé e Poções (bacia hidrográfica do rio Paraíba). Círculos pretos representam os sites de amostragem; círculos amarelos os sites onde o táxon foi observado; círculos vermelhos os sites onde táxon é esperado.

Taxa: Polycentropodidae

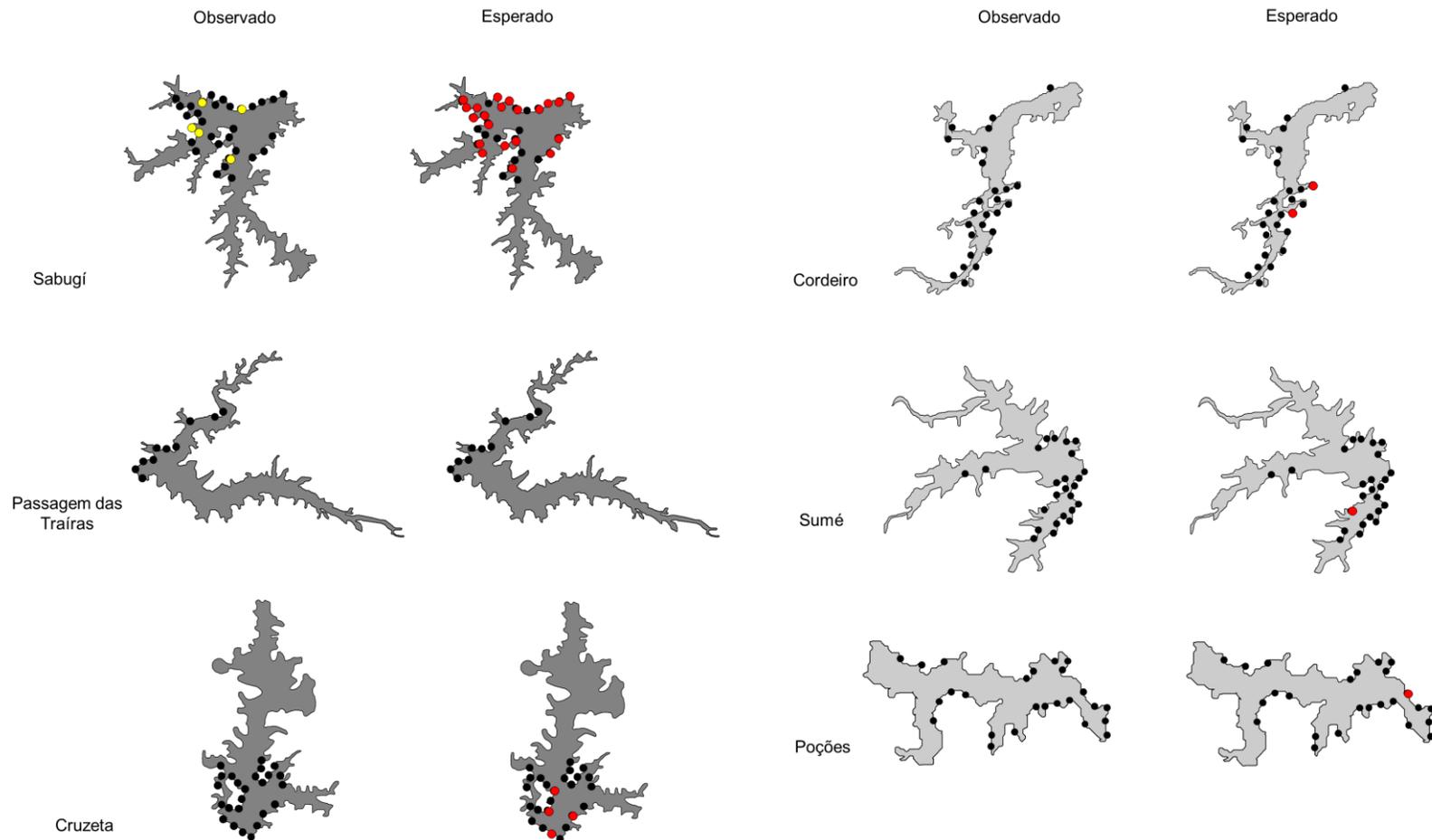
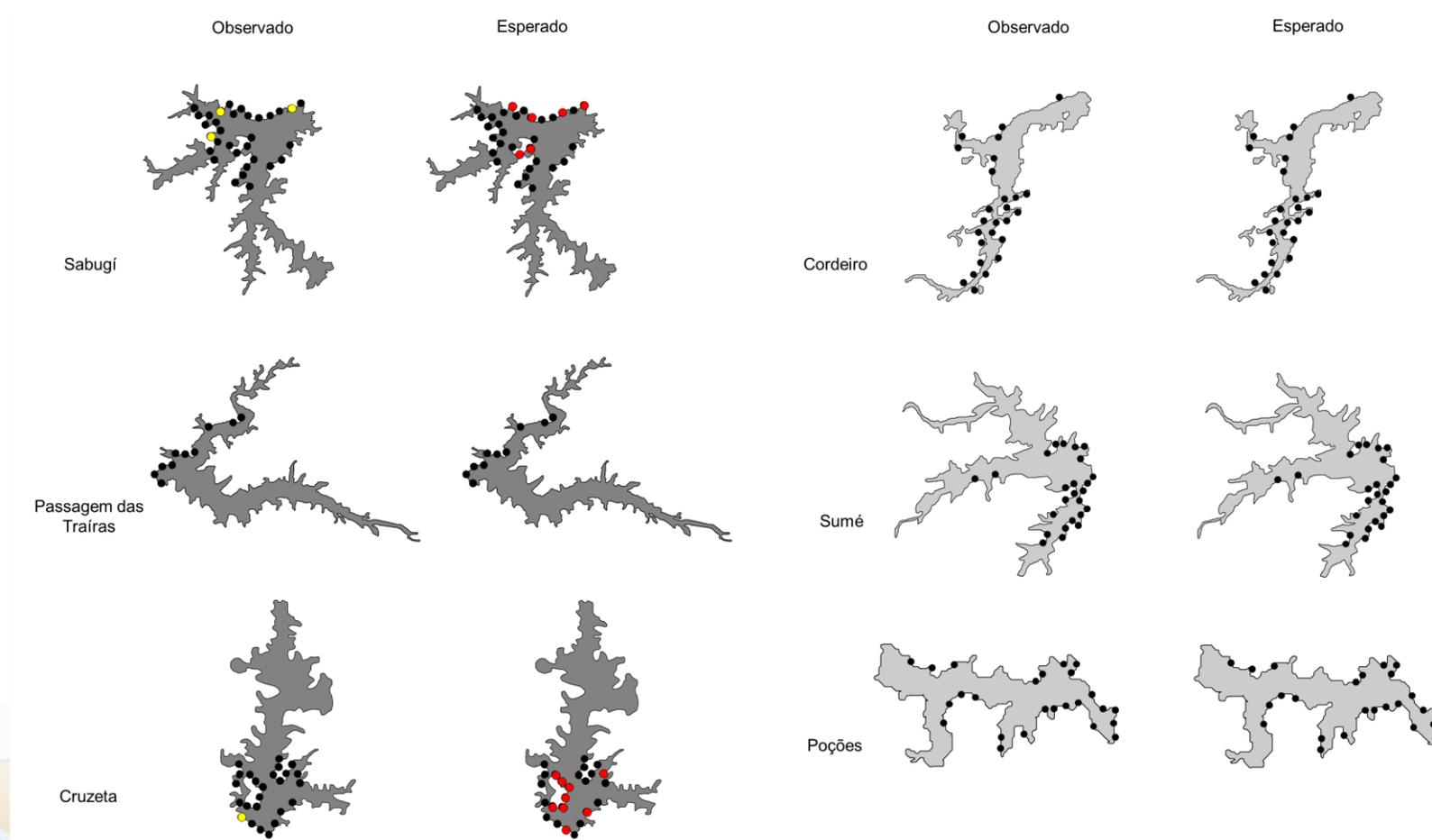


Figura 5: Distribuição espacial esperada de organismos do genero *Zavreliella* baseado o cenário D2 de melhoria nos reservatórios Sabugí, Passagem das Traíras, Cruzeta (bacia hidrográfica do rio Piranhas-Assu), Cordeiro, Sumé e Poções (bacia hidrográfica do rio Paraíba). Círculos pretos representam os sites de amostragem; círculos amarelos os sites onde o táxon foi observado; círculos vermelhos os sites onde táxon é esperado.

Taxa: *Zavreliella*



CONSIDERAÇÕES FINAIS

Nossos resultados fornecem importantes informações para o contexto de reabilitação de recursos hídricos:

- i) A diminuição do enriquecimento de nutrientes e contaminação orgânica, considerando a suspensão da entrada de esgoto doméstico e o tratamento de águas residuais, representa uma medida que resulta na melhoria da condição ambiental, aumento da riqueza total e expansão de táxons sensíveis. Condição que pode indicar também a melhoria dos serviços ecossistêmicos (pesca, consumo humano e irrigação).
- ii) Sites sob alto nível de degradação ambiental necessitam do emprego de medidas associadas para garantia de um quadro de melhoria satisfatório;
- iii) Se os cenários de melhoria tornarem-se uma realidade, as novas condições podem favorecer a chegada de táxons antes limitados pela baixa qualidade ambiental, como táxons comuns a zonas lóxicas. Estudos futuros podem considerar a inclusão de espécies comuns a essas zonas junto aos dados de treinamento, a fim de modelar o padrão de distribuição dessas espécies sobre o novo gradiente ambiental.

REFERÊNCIAS

- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; MORAES, G.; LEONARDO, J.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, n 22, p. 711-728, 2013.
- ANDERSON, M.J.; GORLEY, R. N.; CLARKE, K. R. **Permanova + for Primer: Guide to Software and Statistical Methods**. Primer-E Ltd, Plymouth, 2008.
- APHA, A. **Standard methods for the examination of water and wastewater**, n 21, p. 258-259, 2005.
- ASHOFTEH, P. S.; RAJAEI, T.; GOLFAM, P. Assessment of Water Resources Development Projects under Conditions of Climate Change Using Efficiency Indexes (EIs). **Water Resources Management**, n 12, p. 3723-3744, 2017.
- BOFFI, A.V. **Moluscos brasileiros de interesse médico e econômico São Paulo**. HVCITEC 181, 1979.
- BRUDVIG, L. A. Toward prediction in the restoration of biodiversity. **Journal of Applied Ecology**, n 4, p. 1013-1017, 2017.

- BÜCHI, L.; VUILLEUMIER, S. Coexistence of specialist and generalist species is shaped by dispersal and environmental factors. *The American Naturalist*, n 5, p. 612-624, 2014.
- CALLISTO, M.; ESTEVES, F. A. Composição granulométrica do sedimento de um lago amazônico impactado por rejeito de bauxita e um lago natural (Pará, Brasil). *Acta Limnologica Brasiliensia*, n 8, p. 115-126, 1996.
- CLARKE, R. T.; WRIGHT, J. F.; FURSE, M. T. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, n 3, p. 219-233, 2003.
- CONAMA (Conselho Nacional do Meio Ambiente). Resolução CONAMA N° 357/2005 -- Publicação DOU n° 053, de 18/03/2005, págs. 58-63, 17/03/2005.
- DEDECKER, A.; GOETHALS, P. L.; GABRIELS, W.; PAUW, N. Optimisation of Artificial Neural Network (ANN) model design for prediction of macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium). **9th International Congress on Environmental Modelling and Software**, n 74, p. 141-147, 2004
- EUROPEAN COMMISSION. **Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy**. Official Journal of the European Communities, n 15, p. 1-5, 2000.
- FEIO M.; NORRIS R.; GRAÇA M.; NICHOLS S. Water quality assessment of Portuguese streams: regional or national predictive models? **Ecological Indicators**, n 9, p. 791–806, 2009.
- FEIO, M. J.; VIANA-FERREIRA, C.; COSTA, C. Combining multiple machine learning algorithms to predict taxa under reference conditions for streams bioassessment. **River research and applications**, n 9, p. 1157-1165, 2014b.
- FEIO, M. J.; VIANA-FERREIRA, C.; COSTA, C. Testing a multiple machine learning tool (HYDRA) for the bioassessment of fresh waters. **Freshwater Science**, n 4, p. 1286-1296, 2014a.
- GEVREY, M.; RIMET, F.; PARK, Y. S.; GIRAUDEL, J. L.; ECTOR, L.; LEK, S. Water quality assessment using diatom assemblages and advanced modelling techniques. **Freshwater biology**, n 2, p. 208-220, 2004.
- HERMOSO, V.; PANTUS, F.; OLLEY, J. O. N.; LINKE, S.; MUGODO, J.; LEA, P. Systematic planning for river rehabilitation: integrating multiple ecological and economic objectives in complex decisions. **Freshwater Biology**, n 1, p. 1-9, 2012.

- HOANG, H.; RECKNAGEL, F.; MARSHALL, J.; CHOY, S. **Elucidation of hypothetical relationships between habitat conditions and macroinvertebrate assemblages in freshwater streams by artificial neural networks.** In *Ecological Informatics* (pp. 239-251). Springer, Berlin, Heidelberg, 2003.
- JOVEM-AZEVEDO, D.; BEZERRA-NETO, J. F.; AZEVEDO, E. L.; GOMES, W. I. A.; MOLOZZI, J.; FEIO, M. J. Dipteran assemblages as functional indicators of extreme droughts. **Journal of arid environments**, n 164, p. 12-22, 2019.
- JOY, M. K.; DEATH, R. G. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. **Freshwater Biology**, n 8, p. 1036-1052, 2004.
- LANGHANS, S. D.; REICHERT, P.; SCHUWIRTH, N. The method matters: a guide for indicator aggregation in ecological assessments. **Ecological indicators**, n 45, p. 494-507, 2014.
- LAUGHLIN, D. C.; STRAHAN, R. T.; MOORE, M. M.; FULÉ, P. Z.; HUFFMAN, D. W.; COVINGTON, W. W. The hierarchy of predictability in ecological restoration: are vegetation structure and functional diversity more predictable than community composition? **Journal of Applied Ecology**, n 4, p. 1058-1069, 2017.
- LORENZEN, C. J., Determination of chlorophyll and pheo-pigments: spectrophotometric equations. **Limnology and oceanography**, n 12, p. 343-346, 1967.
- MÉRIGOUX, S.; FORCELLINI, M.; DESSAIX, J.; FRUGET, J. F.; LAMOUREUX, N.; STATZNER, B. Testing predictions of changes in benthic invertebrate abundance and community structure after flow restoration in a large river (French Rhône). **Freshwater Biology**, n 6, p. 1104-1117, 2015.
- MUHAR, S.; JANUSCHKE, K.; KAIL, J.; POPPE, M.; SCHMUTZ, S.; HERING, D.; BUIJSE, A. D. Evaluating good-practice cases for river restoration across Europe: context, methodological framework, selected results and recommendations. **Hydrobiologia**, n 1, p. 3-19, 2016.
- NICHOLS, S. J.; REYNOLDSON, T. B.; HARRISON, E. T. Evaluating AUSRIVAS predictive model performance for detecting simulated eutrophication effects on invertebrate assemblages. **Freshwater Science**, n 4, p. 1212-1224, 2014.
- PALMER, M. A.; HONDULA, K. L.; KOCH, B. J. Ecological restoration of streams and rivers: shifting strategies and shifting goals. **Annual Review of Ecology, Evolution, and Systematics**, n 45, p. 247-269, 2014.

- PÉRES, G.P. **Guía para el estudio de los macroinvertebrados acuáticos del departamento de Antioquia**. Editorial Presencia Bogotá, 1988.
- REYNOLDSON, T. B.; BAILEY R. C.; DAY, K. E.; NORRIS, R. H. Biological guidelines for freshwater sediment based on Benthic Assessment of Sediment (the BEAST) using a multivariate approach for predicting biological state. **Australian Journal of Ecology**, n 20, p. 198–219, 1995.
- ROSE, P. M.; KENNARD, M. J.; MOFFATT, D. B.; SHELDON, F.; BUTLER, G. L. Testing three species distribution modelling strategies to define fish assemblage reference conditions for stream bioassessment and related applications. **PloS One**, n 1, p. 1-23, 2016.
- ROWAN, J. S.; CARWARDINE, J.; DUCK R. W.; BRAGG, O. M.; BLACK, A. R.; CUTLER, M. E. J.; SOUTAR, I.; BOON, P. J. Development of a technique for lake habitat survey (LHS) with applications for the European Union Water Framework Directive. **Aquatic Conservation: Marine and Freshwater Ecosystems**, n 16, p. 637-657, 2006.
- SIMPSON, J. C.; NORRIS, R. H. **Biological assessment of river quality: development of AUSRIVAS models and outputs**. In WRIGHT, J. F.; SUTCLIFFE, D. W.; FURSE, M. T. (eds) *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques*. Freshwater Biological Association, Ambleside, Cumbria, U.K, p. 125-142, 2000.
- STRACHAN, S. A.; REYNOLDSON, T. B. Performance of the standard CABIN method: comparison of BEAST models and error rates to detect simulated degradation from multiple data sets. **Freshwater Science**, n 4, p. 1225-1237, 2014.
- SUGUIO, K. **Introdução à sedimentologia**. Edgard Blücher, São Paulo, 1973.
- TRIVINHO-STRIXINO, S. **Larvas de Chironomidae: guia de identificação**. São Carlos: UFSCar, 2011.
- WRIGHT, J. F.; SUTCLIFFE, D. W.; FURSE, M. T. Assessing the biological quality of fresh waters: RIVPACS and other techniques. In: *Assessing the biological quality of fresh waters: RIVPACS and other techniques*. **Freshwater Biological Association**, 2000.