

ARTIGO ORIGINAL

Clusterização de dados mistos para análise da atividade pesqueira artesanal na bacia Araguaia-Tocantins

Mixed data clustering for analysis of artisanal fishing activity in the Araguaia-Tocantins basin

Renato S. da Silva¹, Rogério R. da Silva¹, Hugo P. Kurabayashi^{1,2}, Cristiane V. da Cunha¹, Carlos R. L. Francês^{1,2}, Keid N. S. Sousa^{1,3}

¹Federal University of Southern and Southeastern Pará, ²Federal University of Pará, ³Federal University of Western Pará

* {renatosabino, rromulo, hugo, crisvieira_cunha}@unifesspa.edu.br; rfrances@ufpa.br; keid.ufopa@gmail.com

Recebido: 10/10/2019. Revisado: 15/10/2019. Aceito: 23/10/2019.

Abstract

This paper presents the application of an Adaptive Fisheries Monitoring model in conjunction with mixed data clustering techniques related to records of fishing activity in the Araguaia-Tocantins basin between 2016 and 2017. Data records of fishery landings were obtained through the Integrated Fishery Statistics System (SIEPE), which presents itself as a tool proposal capable of streamlining the process of collecting and analyzing data from Amazonian fishing basins. Through the SIEPE data exploration interface several categorical and numerical variables were extracted. From the application of the *k*-prototypes algorithm, it was revealed that the most significant numerical variables in the study were fishing yield and vessel engine power, while the most expressive categorical variables were popular name species and fishing environment. These variables should be taken into account in fisheries monitoring programs in the Araguaia-Tocantins basin, as well as the use of SIEPE to support fisheries management at different scales.

Keywords: Adaptive Fisheries Monitoring; Clustering; Araguaia-Tocantins Basin; SIEPE;

Resumo

Este trabalho apresenta a aplicação de um modelo de Monitoramento Adaptativo da Pesca em conjunto com técnicas de clusterização de dados mistos relativos à registros de atividade pesqueira na bacia Araguaia-Tocantins entre 2016 e 2017. Os registros de dados dos desembarques pesqueiros foram obtidos por meio do Sistema Integrado de Estatística Pesqueira (SIEPE), que se apresenta como uma proposta de ferramenta capaz de dinamizar o processo de coleta e análise de dados das bacias pesqueiras da Região Amazônica. Por meio da interface de exploração de dados do SIEPE diversas variáveis categóricas e numéricas foram extraídas. A partir da aplicação do algoritmo *k*-prototypes, revelou-se que as variáveis numéricas mais expressivas no estudo foram o rendimento da pesca e potência do motor da embarcação, enquanto que as variáveis categóricas mais expressivas foram, nome popular espécie e ambiente de pesca. Estas variáveis devem ser levadas em consideração em programas de monitoramento da pesca na bacia Araguaia-Tocantins, bem como o uso do SIEPE para apoiar a gestão pesqueira em diferentes escalas.

Palavras-Chave: Bacia Araguaia-Tocantins; Clusterização; Monitoramento Adaptativo de Pesca; SIEPE;

1 Introdução

As pescarias continentais mundiais produziram em 2012 cerca de 11 milhões de toneladas de pescado, que representam elevada importância devido à sua contribuição econômica e social para o rendimento das comunidades rurais (FAO, 2016). Em países em desenvolvimento, a pesca de pequena escala representa grande importância para a segurança alimentar e econômica (Kolding et al., 2014). No contexto mundial, o Brasil está entre os maiores produtores de pescado, com 235 mil toneladas de peixes desembarcados em 2014, posicionado na 11a. posição no ranking mundial (FAO, 2016). No entanto estes dados possivelmente são subestimados ou imprecisos dada sua desatualização e ainda observa-se escassez de iniciativas de monitoramento, de recursos e ferramentas para a coleta de informações relacionadas à pesca artesanal de pequena escala Oviedo and Bursztyn (2017), Kolding et al. (2014).

Na Amazônia brasileira, programas de monitoramento da pesca em escala local ou regional tem sido desenvolvidos nos últimos 50 anos de forma descontínua e com diferentes abordagens metodológicas, como censo diário em portos de desembarque pesqueiro, pescarias experimentais, estimativas de pesca por entrevistas e abordagens participativas, incluindo o uso de aplicativos em dispositivos móveis (Barthem et al., 1991, Batista et al., 1998, Cetra and Petrere, 2001, Isaac et al., 2008, Hallwass, 2015, Oviedo and Bursztyn, 2017). Dentre as metodologias utilizadas em programas de monitoramento na Amazônia as que tiveram melhor desempenho foram aquelas que foram baseadas na participação social. A exemplo, citamos o projeto Pro-Várzea/Ibama, que também avançou em instrumentos de armazenamento de informações sobre a pesca por meio do Sistema Integrado de Estatística Pesqueira para a Amazônia Ruffino (2008). Nas demais regiões da Amazônia as iniciativas de integração de dados sobre a pesca ainda não são uma realidade, representando portanto um cenário desafiador quanto à aplicação e manutenção de tais programas de monitoramento e acompanhamento da atividade pesqueira.

Cabe ainda ressaltar que grande parte das iniciativas de monitoramento da atividade pesqueira desenvolvidas na Amazônia Brasileira não considera a adoção de ferramentas de tecnologia da informação para coleta, armazenamento e análise das informações obtidas no processo de monitoramento. Assim, a abordagem utilizada neste trabalho toma como base, dados do Sistema Integrado de Estatística Pesqueira (SIEPE)¹, que foi desenvolvido para atender o Programa de Monitoramento Adaptativo da Pesca na média bacia Araguaia-Tocantins. Os dados utilizados do SIEPE foram coletados no Monitoramento Adaptativo da Pesca realizado entre setembro de 2016 à outubro de 2017 nas comunidades ribeirinhas de Itupiranga e São Geraldo do Araguaia, na bacia Araguaia Tocantins, Pará, Brasil.

Em adição ao exposto, ainda que existam programas de monitoramento que utilizem sistemas de informação, no melhor do nosso conhecimento, há falta de

iniciativas que combinem as informações armazenadas em tais sistemas, com técnicas de análise e mineração de dados, que em geral buscam explorar conjuntos de dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis, detectando assim novos subconjuntos de dados.

Desta forma, este trabalho tem como objetivo apresentar uma análise baseada em clusterização, como técnica de aprendizado de máquina não supervisionadas, que consiste em um conjunto de técnicas de prospecção de dados que buscam realizar agrupamentos automáticos de dados segundo o seu grau de semelhança, considerando critérios de semelhança e distância, para definição dos agrupamentos ou aglomerados. Assim, a partir da utilização de algoritmos de clusterização, é possível realizar a sumarização e extração de padrões e agrupamentos dos dados, de modo a fornecer informações complementares à estatística tradicional aplicada na literatura correlata.

Como organização do trabalho, na seção 2 são apresentados os materiais e métodos utilizados na pesquisa, dentre eles a descrição do processo de Monitoramento Adaptativo de Pesca, uma breve discussão sobre métodos de clusterização de dados mistos utilizados e o projeto e arquitetura do SIEPE. Na seção 3 são apresentados os resultados e discussão dos dados analisados por meio de técnicas de clusterização; e a seção 4 apresenta as considerações finais e discussões sobre trabalhos futuros.

2 Materiais e Métodos

2.1 Monitoramento Adaptativo Pesqueiro

O modelo de Monitoramento Adaptativo da Pesca (MAP) foi realizado entre os anos de 2015 a 2019. Este modelo buscou compreender a complexidade do sistema ecológico da porção média da bacia Araguaia-Tocantins adotando múltiplas metodologias com capacidade para evoluir à medida que novos problemas fossem surgindo (Lindenmayer and Likens, 2009, Lindenmayer et al., 2011). O objetivo central nos modelos de monitoramento adaptativo é proporcionar que as informações coletadas possam gerar políticas públicas e evoluir para programas de manejo adaptativo (Lindenmayer and Likens, 2009, Lindenmayer et al., 2011). O MAP foi realizado nas comunidades ribeirinhas de Santa Teresinha do Tauiri e Santo Antônio, situadas às margens do rio Tocantins em Itupiranga e nas comunidades de Santa Cruz dos Martírios e Ilha de Campo, localizadas às margens do rio Araguaia em São Geraldo do Araguaia, sendo que ambas comunidades estão localizadas na porção média da bacia Araguaia-Tocantins, no Estado do Pará (PA) conforme divisão sugerida pelo Programa de Estruturação Institucional para Consolidação da Política de Recursos Hídricos, do Ministério do Meio Ambiente MMA (2006). A Fig. 1 apresenta a representação da Bacia Araguaia-Tocantins.

Para realização do Mapeamento Adaptativo de Pesca (MAP) foram consideradas 03 (três) etapas, conforme a seguinte descrição:

¹SIEPE: <https://siepe.lcc.unifesspa.edu.br>

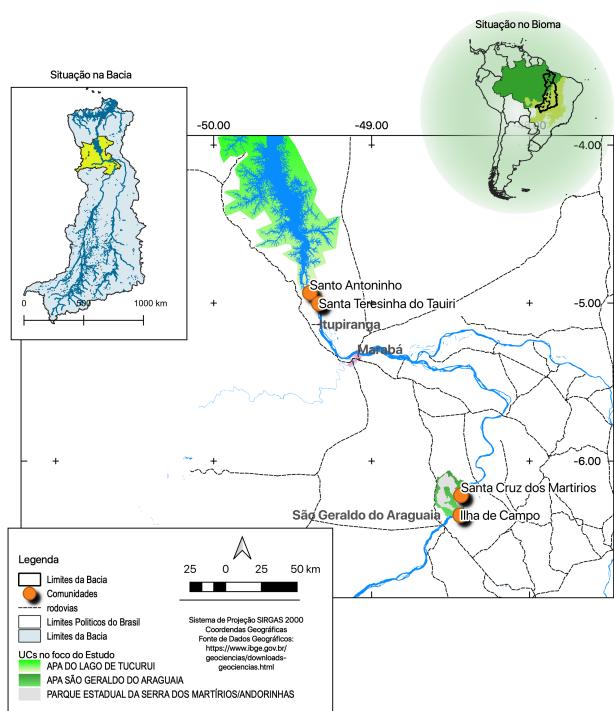


Figura 1: Representação da Bacia Araguaia-Tocantins.

- Etapa 1 – Construção de Confiança: Esta etapa esteve baseada no processo de estabelecimento de parcerias dentro das comunidades, entre instituições e pesquisadores para alinhamento coletivo dos objetivos e metodologia do monitoramento participativo da pesca e capacitação dos pescadores em processo contínuo que iniciou em 2015;
- Etapa 2 – Monitoramento Participativo: Nesta etapa foi desenvolvido um conjunto de atividades entre setembro de 2016 a outubro de 2017, que consistiram na aplicação do monitoramento da pesca, planejamentos coletivos e estruturação de um sistema de informação para cadastro dos dados coletados (SIEPE), que pudesse ser utilizado para aplicação de técnicas de clusterização e análise exploratória de dados;
- Etapa 3 – Consolidação de Dados: Esta etapa consistiu no projeto, desenvolvimento, teste funcional, implantação e avaliação do MAP e do SIEPE, além da aplicação de técnicas de análise de dados para análise e interpretação dos dados coletados resultante do processo de monitoramento de desembarques pesqueiros.

O monitoramento seguiu metodologia adaptada de Hallwass (2015), onde cada pescador voluntário preencheu cartões de desembarque de 05 (cinco) pescarias do mês durante 01 (um) ano. As informações preenchidas nos cartões de desembarque estiveram organizadas conforme os seguintes tipos de variáveis numéricas: data da pesca, quantidade de pescadores envolvidos na pescaria, tempo gasto na pescaria (horas), tamanho e quantidade de redes, potência do motor (hp), captura

por espécie (kg), captura total de espécies (kg), total de pescado vendidos (kg), total de pescado consumido (kg), remuneração total recebida pela venda (R\$ - Reais), custos associados à pescaria (R\$ - Reais).

Adicionalmente, o MAP também adotou um conjunto de variáveis categóricas, que descrevem características qualitativas presentes nos eventos de pesca, presentes nos cartões de desembarque. As seguintes variáveis categóricas foram consideradas no MAP: Nome do rio, Nome do Pescador, Tipo de Embarcação, Local de Pesca, Ambiente de Pesca, Nome popular de espécie e Tipo de Petrecho. Adicionalmente, a Tabela 1 apresenta de forma resumida a descrição das variáveis categóricas utilizadas e seus respectivos conjuntos de valores obtidos a partir dos dados coletados nos cartões de desembarque.

Os cartões de desembarques foram entregues a 25 (vinte e cinco) pescadores que se voluntariaram a participar da pesquisa. Os pescadores foram capacitados para o preenchimento dos cartões por meio de seminários e visitas de campo, que após preenchidos foram recolhidos pelos pesquisadores e as informações inicialmente armazenadas em planilhas eletrônicas, que posteriormente foram importadas por meio da interface de integração de dados do SIEPE.

2.1.1 Captura por Unidade de Esforço - CPUE

O setor pesqueiro exige modelos onde o homem esteja incluso para realizar o monitoramento da pesca (Gartside and Kirkegaard, 2010) em uma abordagem socioecológica levando em consideração a teoria ecológica que é baseada em conceitos de ecologia de sistemas, ecologia cultural e ecologia evolutiva. A ecologia evolutiva faz uma relação entre a interação predador-presa que influi diretamente na dinâmica das populações, neste sistema, os predadores são os pescadores e as presas os recursos pesqueiros (da Silva Batista et al., 2004).

Nesta relação, historicamente o homem aplica o conhecimento acumulado sobre a natureza para minimizar o esforço e maximizar a confiabilidade de captura (Gartside and Kirkegaard, 2010), este princípio foi determinante para o desenvolvimento de modelos de monitoramento dos recursos aquáticos.

A relação predador-presa pode ter implicações na gestão pesqueira, por este motivo as análises sobre os recursos pesqueiros tendem a avaliar as estratégias de pesca (tempo de pesca, número de pescadores, petrechos, embarcações e a quantidade de peixes capturados) para entender qual o esforço tem sido empregado pelos pescadores sobre o recurso e como este esforço pode ser otimizado em busca de uma sustentabilidade. Neste ínterim as análises estatística utilizadas em programas de monitoramento da pesca são pautadas no cálculo da Captura por Unidade de Esforço (CPUE) que tem como objetivo entender como as ações dos pescadores interagem na dinâmica populacional dos recursos pesqueiros, ou mesmo para avaliar a produtividade do sistema e das estratégias de pesca, sobretudo das espécies explotadas comercialmente (da Silva Batista et al., 2004).

O uso de cálculos matemáticos representa uma simplificação da realidade, no caso específico dos ecossiste-

Tabela 1: Descrição de variáveis categóricas utilizadas no MAP

Variável Categórica	Possíveis Valores
Ambiente de Pesca	Rio, lago, pedral e igarapé
Local de Pesca	Porto de Santa Cruz dos Martírios e Ilha de Campo
Nome do Pescador	-
Nome Popular das Espécies	Avoador, braquinha, curimatá, mapará, pacu, piau e tucunaré
Nome do Rio	Araguaia ou Tocantins
Tipo de Embarcação	Canoa, lancha ou rabeta
Tipo de Petrecho	Malhadeira, armadilha fixa e linha de mão

mas aquáticos, a representação de uma realidade bem complexa, uma vez que é difícil determinar o estoque de determinada espécie de peixes e desta forma avaliar qual a produção máxima sustentável. Mesmo assim a CPUE tem sido a mais comumente utilizada na avaliação de estoques pesqueiros (Petrere Jr and Peixer, 2007).

O uso da CPUE supõe uma relação linear entre a captura (C), o esforço (f) e (B) biomassa do estoque, $C = q \cdot f \cdot B$, onde q representa o coeficiente de capturabilidade, que deve ser constante (Nunes, 2015, da Silva Jr. et al., 2017). O esforço neste contexto pode ser aquele fornecido pelo pescador, como tamanho da embarcação, tamanho da malhadeira ou a quantidade de lances de tarrafa, que neste caso deverá ser analisada como produto da quantidade de horas pescando e quantidade de pescadores participantes. O cálculo de CPUE tem sido amplamente utilizado nos estudos de estatística pesqueira como a principal ferramenta para avaliação de estoques de peixes e de medidas de gestão dos recursos aquáticos da Silva Jr. et al. (2017).

A CPUE é embasada a partir da Teoria do Forrageamento Ótimo, onde é pressuposto que a aptidão de um animal provém de sua eficiência ao procurar e obter recursos (Begossi, 1993), tendo como resultado uma razão positiva entre o tempo gasto pela energia adquirida através da obtenção/consumo do recurso. Este retorno pode ser traduzido de forma monetária, onde pode ser utilizado para satisfazer outras necessidades do forrageador (Silva et al., 2010).

Porém, para este trabalho, o cálculo do índice CPUE utilizado no MAP, é realizado considerando a definição da Eq. (1), a partir de adaptação da proposição de Petrere Jr (1978), que propõem que este seja o melhor modelo a ser utilizado, pois a unidade de CPUE gerada pelo número de pescadores tomados como unidade de esforço de pesca, apresentam o menor coeficiente de variação nos cálculos estatísticos. Quando não for possível calcular pela variável hora é utilizada a variável dia, como utilizado por Hallwass (2015) em estudo realizado no rio Tocantins, pois algumas pescarias no rio Tocantins podem durar mais de um dia, sobretudo no período de maior produção do Mapará (*Hypophthalmus marginatus*) e do Avoador (*Hemiodus unimaculatus*). Conforme a Eq. (1), temos que:

$$C = \frac{p}{n \cdot t}, \quad (1)$$

onde p representa a produção total de pescado (kg), n

representa o quantitativo de pescadores envolvidos na pescaria e t representa o tempo total investido na pesca (horas). Para garantir a análise da CPUE, os únicos atributos que não puderam ficar sem preenchimento no SIEPE foram: rio, data, pescador, tempo gasto na pesca (horas), quantidade de pescadores envolvidos na pesca e captura em kg total. Estas informações são consideradas essenciais pois dão subsídio para a estatística pesqueira amplamente utilizada por pesquisadores na Amazônia (Ruffino, 2004, Petrere Jr and Peixer, 2007, Hallwass and Silvano, 2015, Sci et al., 2014, da Silva Jr. et al., 2017).

2.2 Arquitetura e Modelagem de Dados do SIEPE

O Sistema Integrado de Estatística Pesqueira (SIEPE) é um sistema que possui arquitetura cliente-servidor, baseado em tecnologias Web e foi desenvolvido especialmente para atender o Programa de Monitoramento Adaptativo da Pesca na média bacia Araguaia-Tocantins. Ao ser baseado em plataformas Web, ele permite o acesso por meio de navegadores em dispositivos fixos e móveis, facilitando desta forma o processo de interação com a interface do sistema. O SIEPE ainda adota um conjunto diverso de tecnologias assistivas e realiza o atendimento à padronizações internacionais de usabilidade e acessibilidade para os usuários do sistema. Entre as metodologias de desenvolvimento de software utilizadas no processo de desenvolvimento de software, podemos destacar a *eXtreme Programming* (XP) (Beck, 1999). Com relação às ferramentas e plataformas de desenvolvimento utilizadas, o SIEPE foi desenvolvido integralmente com ferramentas de software livre das quais podemos destacar a linguagem PHP, o sistema gerenciador de Banco de Dados PostgreSQL, o servidor Web Apache e o framework Selenium para execução de testes funcionais.

Com relação ao desenho arquitetural do SIEPE, este está baseado no padrão de projeto *Model-View-Controller* (MVC). A camada *Model*, é responsável pela representação dos dados da aplicação e pelo gerenciamento de estados das entidades do sistema. A camada *View* gerencia as informações apresentadas na interface gráfica do sistema, sendo responsável por apresentar as informações para o usuário através de uma combinação de gráficos e textos e relatórios associados às métricas de pesca adotadas. Por fim, a camada *Controller*, busca gerenciar o fluxo de dados entre os componentes do sistema, aplicando as regras de negócio, bem como validando as informações necessárias para o cálculo do

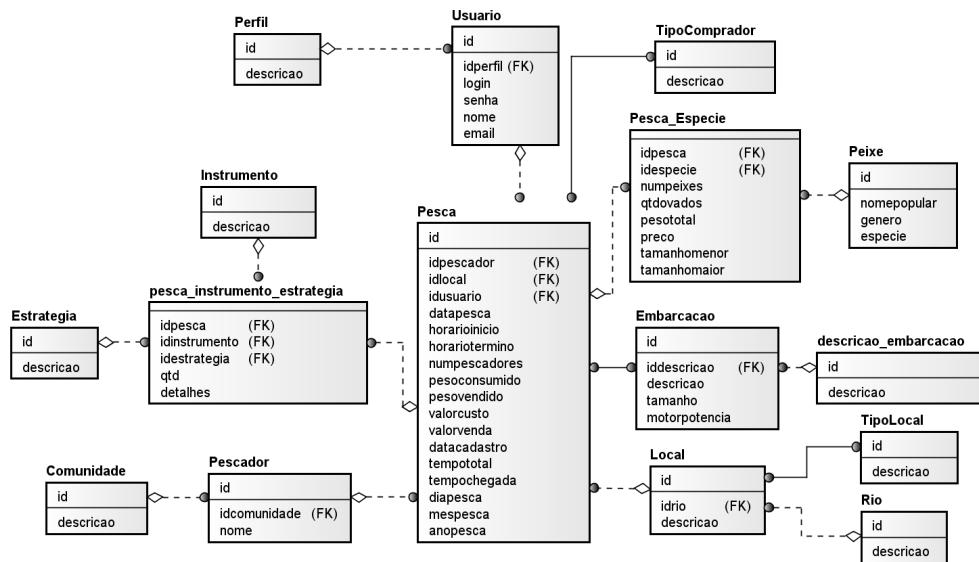


Figura 2: Diagrama de Entidade-Relacionamento do SIEPE. Fonte: Autoria Própria.

CPUE, por exemplo. O SIEPE ainda adota o padrão arquitetural de acesso de dados *Data Access Object* (DAO) e as suas funcionalidades estão organizadas em: Registrar Pesca, Gerenciar Cadastros Auxiliares da pesca, Importar Dados, Realizar Consulta, Gerar Relatórios, Visualizar Dados da Pesca e Exportar Dados.

O SIEPE busca promover o armazenamento, tratamento e transformação dos dados do monitoramento da pesca, tendo o potencial de integrar informações de diversas bacias pesqueiras, dado que o seu banco de dados foi concebido de forma a se ajustar a diversos tipos de dados e atributos, de modo a considerar a inclusão futura de novos atributos advindos do MAP. Assim, o SIEPE pode possibilitar a análise da dinâmica da pesca de diferentes escalas, em diferentes bacias podendo auxiliar inclusive em iniciativas de cruzamento e correlação de dados entre diversas bacias.

Ao se considerar os atributos numéricos e categóricos adotados pelo processo de Mapeamento Adaptativo de Pesca, a Fig. 2 apresenta o Diagrama de Entidade-Relacionamento (DER) concebido para o SIEPE, no processo de modelagem relacional do banco de dados, onde é possível visualizar os diversos atributos que compõem a base de dados do SIEPE.

2.3 Clusterização de Dados Mistos

O particionamento do conjunto de dados em grupos homogêneos ou clusters é uma operação fundamental do processo de análise de dados. É uma operação útil em diferentes situações, como classificação (não supervisionada), agregação, segmentação, por exemplo. Clusterização é uma abordagem comum na literatura, para particionamento de conjuntos de dados, de tal forma que os objetos classificados sob um mesmo cluster apresentam maior similaridade entre si, em relação à objetos de outros agrupamentos (Huang, 1998).

Em termos de clusterização, a literatura apresenta

diversas abordagens de agrupamento, relacionados à dados numéricos e categóricos. Em ambos os casos é desejável a adoção de técnicas que realizem tal agrupamento de modo eficiente, em geral para grandes bases de dados. Como os dados apresentados no SIEPE consistem de dados numéricos e dados categóricos, o algoritmo a ser utilizado deveria ser capaz de trabalhar de modo eficiente com os ambos tipos de dados. Assim, durante a análise em questão optou-se pela utilização do algoritmo *k-prototypes* (Huang, 1998). Este algoritmo combina a capacidade de agrupamento de dados numéricos do algoritmos *k-means* (Anderberg, 1973), com o algoritmo *k-modes*, que realiza agrupamento de dados categóricos por meio da medição de dissimilaridade entre os objetos.

Assim, dado um conjunto de dados X , contendo n objetos de dados descritos por m_r atributos numéricos e m_c atributos categóricos, o objetivo do algoritmo k -prototypes é minimizar a seguinte função objetivo:

$$J = \sum_{l=1}^k \sum_{i=1}^n p_{il} d(x_i, Q_l), \quad (2)$$

onde p_{il} é uma variável binária, isto é $p_{il} \in \{0, 1\}$, indicando a pertinência do objeto de dados x_i no cluster l ; Q_l é o *prototype* (centro) do cluster l e $d(x_i, Q_l)$ é a dissimilaridade medida, que é definida como:

$$d(x_i, Q_l) = \sum_{r=1}^{m_r} (x_{ir} - q_{lr})^2 + \gamma_l \sum_{c=1}^{m_c} \delta(x_{ic}, q_{lc}), \quad (3)$$

onde x_{ir} representa os valores dos atributos numéricos e x_{ic} os valores dos atributos categóricos para cada objeto de dados x_i . $Q_l = q_{l1}, \dots, q_{lm}$ representa os centroides do cluster l , onde q_{lr} é a média do atributo numérico r

e do cluster l , enquanto q_{jc} é a amostra mais comum (moda) do atributo categórico c e cluster l . Para atributos categóricos, $\delta(p, q) = 0$ para $p \equiv q$ e $\delta(p, q) = 1$ para $p \neq q$. γ_l é um peso para atributos categóricos, de modo a balancear a influência entre as componentes da equação. A formulação geral do algoritmo k -prototypes é descrita pelo algoritmo 1.

Algoritmo 1: Formulação geral do k -prototypes

```

1 Dados: Conjunto de Dados  $X = \{x_1, \dots, x_n\}$ ; valor  $k$ ;
  Result: Centroides dos Clusters  $Q = \{Q_1, \dots, Q_k\}$ 
2 begin
3   Escolher  $k$  centros de Clusters a partir de  $X$ ;
4   repeat
5     Computar a distância entre os objetos de
       dados e os centros dos clusters utilizando
       Eq. (2);
6     Atualizar os centros dos clusters  $Q$  (Salva
       o conjunto prévio de centroides em  $Q'$ 
       para analisar a convergência);
7   until  $Q' = Q$ ;
8 end

```

3 Análise de Dados e Resultados

O desenvolvimento do MAP registrou 477 desembarques de pesca entre setembro de 2016 e outubro de 2017 que foram incluídos na base de dados do SIEPE. O SIEPE permitiu a inclusão de 100% dos registros de desembarque do MAP e não foram observados dados obrigatórios ausentes ou que necessitassem de prévia transformação, a partir dos formulários utilizados pelos pescadores.

Todas as funcionalidades da arquitetura do SIEPE tiveram bom funcionamento, e através do SIEPE utilizou-se a interface de exploração de dados, utilizada com o objetivo de visualizar e combinar diversos atributos, bem como obter informações relacionadas à estatística básica das variáveis em questão, como moda, média, mediana ou desvio padrão. É importante destacar, que a partir da interface em questão foi possível a geração de um arquivo de dados em formato aberto, que continha todos os registros de desembarques, discretizados e transformados para facilitar o processo de análise de dados.

A partir do arquivo de dados obtido através do SIEPE, buscou-se dar continuidade às atividades de pré-processamento de dados, onde algumas das variáveis foram suprimidas das análise, como nome do pescador, Estado ou cidade dado o aparente baixo ganho de entropia para a análise. Assim, após a finalização das atividades de pré-processamento de dados, a configuração dos dados relacionados ficou composta por um total de 09 (nove) variáveis categóricas (Comunidade, Local de Pesca, Rio, Tipo de Embarcação, Arte de Pesca, Ambiente de Pesca e Nome Popular da Espécie) e 08 (oito) variáveis numéricas (Mês da pesca, Número de

Pescadores, Peso Consumido, Peso Vendido, Valor de Custo da Pesca, Valor de Venda da Pesca e Potência da Embarcação).

Ao utilizar o algoritmo k -prototypes, há a necessidade de determinação do parâmetro k , que representa o quantitativo de clusters cujos registros de desembarque devem ser agrupados. Assim, para este estudo foram considerados inicialmente de 01 a 06 clusters, e realizou-se uma análise do número de clusters em função da soma das áreas dos clusters propostos. A partir da análise buscou-se um quantitativo k de clusters que minimizasse a área total dos clusters, conforme apresentado pela Fig. 3. Quando $k > 3$, observa-se que a área dos clusters não apresenta significativa redução, e desta forma optou-se por prosseguir com o parâmetro k configurado como 03 (três) clusters.

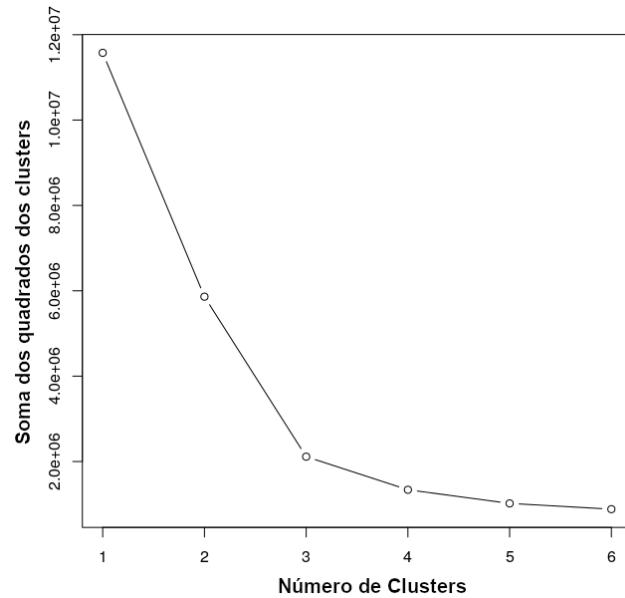


Figura 3: Representação do quantitativo de clusters analisados.

Para execução do algoritmo k -prototypes, utilizou-se como parâmetros $\gamma_l = 0.1$, e com base em sua execução, 338 registros foram classificados no cluster 1, 07 registros no cluster 2 e 132 registros foram agrupados no cluster 3. Os clusters encontrados possuem características similares ao se considerar parte dos atributos categóricos. No entanto, ao se considerar os centroides dos clusters 1, 2 e 3, em especial a variável numérica, *valor de venda do pescado*, observa-se que os clusters obtiveram valores expressivamente distintos, indicando uma possível finalidade da atividade pesqueira monitorada. A partir da variável *valor de venda*, a Fig. 4 apresenta uma representação gráfica do agrupamento obtido por meio do k -prototypes.

As variáveis numéricas foram as que mais influenciaram na análise, em especial a variável *valor de venda* e *potência da embarcação*. As variáveis categóricas que mais influenciaram foram *nome popular espécie* e *ambi-*

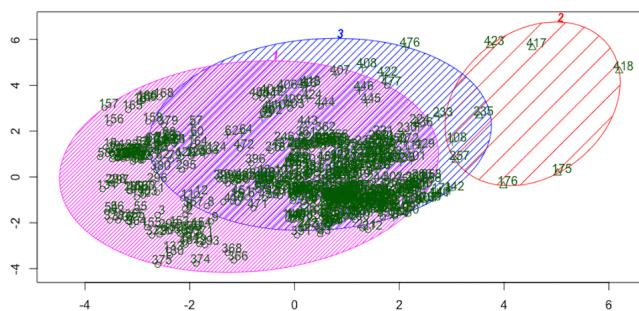


Figura 4: Representação gráfica em 2D demonstrando agrupamento em cluster de dados numéricos e categóricos do MAP

ente de pesca. A variável que representa o valor de venda da pesca foi responsável por cerca de 33,98% da variabilidade dos dados, além de representar a variável que melhor divide os dados em questão, conforme dados apresentados pela Tabela 2.

Em adição ao exposto, a variável *valor de venda* apresenta valores bem característicos para cada um dos clusters, conforme apresentado pela Fig. 5. É importante destacar que para o cluster 2, o primeiro quartil e a mediana possuem valores aproximados de R\$1.000,00, enquanto que para o cluster 3, estes valores aproximam-se de R\$400,00. Para melhor explicar a diferença financeira entre os clusters, buscou-se observar inicialmente outras variáveis numéricas, como *número de pescadores*, *peso consumido* e *potência da embarcação*. Para a variável *potência da embarcação*, observa-se que o cluster 2 apresenta como valor 7,57 (hp), indicando que as embarcações utilizadas por pescadores agrupados no cluster 2, são de maior porte do que as do cluster 1 e 3 (Tabela 2).

Pescarias com barcos motorizados em geral apresentam maior quantidade de peixes capturados, como já observado no médio Araguaia (da Silva and de Farias, 2017), no baixo Tocantins (Hallwass et al., 2013) e outras regiões da Amazônia (Isaac et al., 2008). O comportamento do pescador (número de pescadores e potência da embarcação), aqui analisado, foi capaz de influenciar em diferenças no rendimento da pesca artesanal. Esta constatação pode estar relacionada à necessidade dos pescadores, que tem motores mais potentes, necessitarem aumentar o esforço, devido ao custo mais elevado com petrechos de pesca e combustível, que pode estar relacionado com a escolha de espécies alvos.

A influência da potência da embarcação e sua aparente influência no rendimento de pesca artesanal, pode indicar a necessidade de adaptar a métrica de esforço que de modo a considerar tais constatações, de forma a criar uma relação que possivelmente considere a potência da embarcação como um grandeza inversa à métrica de esforço (*f*).

Ao analisar os dados categóricos, em especial, as variáveis *nome popular espécie* e *ambiente de pesca*, as diferenças observadas entre os clusters ficam ainda mais evidentes e sugerem que as pescarias na média bacia

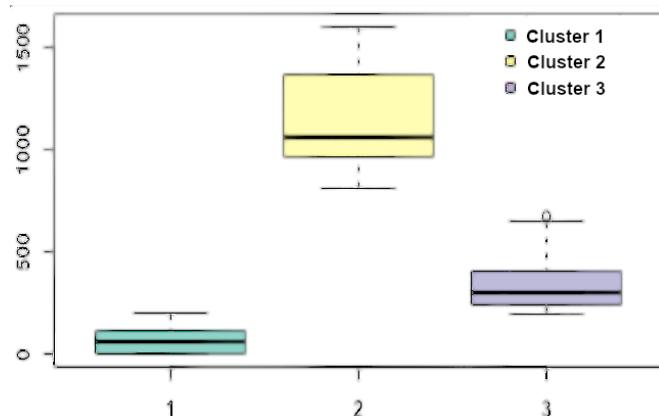


Figura 5: Diagrama boxplot da variável *valor de venda*

Araguaia-Tocantins, mesmo sendo multi-específicas (Hallwass and Silvano, 2015, Sci et al., 2014), apresentam moderada seletividade para espécies com maior valor comercial, ou maior abundância na bacia, relacionado ainda à escolha do ambiente de pesca.

Com relação à variável *nome popular espécie*, observa-se que no cluster 1, o principal grupo de peixes capturados é o Tucunaré (*Cichla spp.*, *C. kelberi*, *C. piquiti*), em ambientes de pedral e meio do rio. O Tucunaré é considerado um peixe nobre e que representa elevada importância econômica no rio Araguaia (Sci et al., 2014) e no rio Tocantins (Alves and Barthem, 2008, Prysthon et al., 2019). Devido ser um dos peixes de maior valor econômico no Araguaia-Tocantins, espera-se que o rendimento dos pescadores em relação a esta espécie seja alto, no entanto, não foi o observado na análise. Desta forma sugere-se que os pescadores agrupados no cluster 1, estejam direcionando a venda do pescado a atravessadores, que compram por um valor subestimado, fazendo com que o rendimento caia. Situações semelhantes a esta já foram observadas no rio Tocantins e outras regiões da Amazônia, onde o pescador torna-se dependente de um sistema de exploração (de Camargo and Petrere Jr, 2004, Fraxe et al., 2007, Petrere Jr and Peixer, 2007). Neste caso, é importante observar se os rendimentos capturados nos clusters 2 e 3 se refere à quantia justa da produção do pescador.

No cluster 2, estão agrupados amostras de pescarias em ambientes no meio do rio e na beira do rio, para captura de grupos de espécies variadas, como Avoador (*Hemiodus unimaculatus*, *H. microlepis*, *H. argenteus*), Piau (*Leporinus spp.*, *Leporinus fasciatus* e *Schizodon vittatus*), Curimatá (*Prochilodus nigricans*), Pacu (*Myleus setiger*, *Myloplus rubripinnis*, *Myloplus asterias*, *Mylesinus paucisquamatus*), Branquinha (*Curimata spp.*, *C. cyprinoides*, *C. kneri*) e Mapará (*Hypophthalmus spp.*, *H. marginatus*, *H. edentatus*, *H. fimbriatus*), que representam espécies de maior expressão comercial ou maior abundância na bacia (Hallwass and Silvano, 2015, Sci et al., 2014, Juras et al., 2004, da Silva and de Farias, 2017).

A diversidade de espécies capturadas talvez seja o fator mais relevante para as pescarias do cluster 2, garantindo assim, melhores condições de negociação,

Tabela 2: Dados de centroides encontrados por meio do k-prototypes

	Cluster 1	Cluster 2	Cluster 3
Amostras	338	7	132
Valor Venda	R\$ 66,79	R\$ 1.161,00	R\$ 335,93
N. Pescadores	1,85	2,0	2,02
Peso Consumido	3,12 kg	2,57 kg	2,39 kg
Potência do Motor	5,45 hp	7,57 hp	5,41 hp
Nome Popular Espécie	Tucunaré	Avoador, Piau, Curimatá, Pacu, Branquinha, Mapará	Avoador
Ambiente de Pesca	Pedral, Meio do Rio	Meio do Rio, Beira do Rio	Pedral, Meio do Rio

elevando o rendimento. É importante enfatizar, que a pesca do Avoador e Mapará (espécies mais expressivas nos clusters 2 e 3), representam grande biomassa desembarcada no rio Tocantins (Juras et al., 2004, Hallwass, 2011, Prystyon et al., 2019). Porém, o valor de venda desses peixes é de no máximo R\$2,50/kg (verificação de campo). Pela grande quantidade capturada, os pescadores necessitam de motores mais potentes para auxiliar na propulsão da embarcação, além de que, os petrechos utilizados em pescarias de Mapará e Avoador, necessitam de alto investimento financeiro. Neste caso, especial atenção deve ser dada ao cluster 2, que apresenta volume financeiro médio superior em torno de 345% em relação ao cluster 3. Uma pesquisa mais apurada poderia ser realizada com o propósito de verificar se esses rendimentos estão sendo direcionados para os pescadores, ou para atravessadores, em um modelo de exploração do trabalho.

No cluster 3, o grupo de espécies mais capturada é o Avoador, em ambientes de pedrais e meio do rio. O rendimento deste grupo de pescadores não é tão alto como no cluster 2, e nem a potência dos motores de suas embarcações. É possível dizer que os pescadores que estão investindo na pesca do voador, empregam maior esforço durante a safra desta espécie (junho a setembro), que mesmo tendo valor comercial baixo, compensa pela quantidade capturada. No decorrer do ano o esforço é direcionado para outras espécies, o que explicaria a sobreposição dos clusters 1, 2 e 3, conforme apresentado pela Fig. 4.

Cabe ressaltar, que há similaridades entre os dados observados, em especial ao se considerar os dados numéricos. Entende-se que os clusters 1, 2 e 3 estão relacionados à atividade pesqueira de pequena escala e para subsistência familiar, mesmo que alguns grupos se sobressaiam em relação ao rendimento. A constatação de que o tipo de pescaria é de pequena escala e subsistência, se dá em consideração pela quantidade de pescado consumido por dia de pesca, além da potência do motor das embarcações e a tripulação embarcada (da Silva and de Farias, 2017). Mediante esta constatação, a aplicação de políticas de proteção dos recursos hídricos e pesqueiros representa questão de fundamental importância dos pontos de vista ambiental, da economia local e de segurança alimentar na bacia Araguaia-Tocantins.

4 Considerações Finais

Um dos grandes entraves para a governança dos recursos pesqueiros é falta de informações sobre o setor. Ainda há carência de iniciativas que gerem informação agregada para definição de políticas públicas para o melhor manejo e defesa das espécies.

Neste contexto, este trabalho propõe a implantação de um modelo de Monitoramento Adaptativo da Pesca, o desenvolvimento do Sistema Integrado de Estatística Pesqueira e aplicação de técnicas de clusterização de dados para gerar entendimento complementar à literatura relacionada. Com relação a avaliação das contribuições realizadas por este trabalho para melhoria nos processos de gestão pesqueira, podemos considerar:

- **MAP:** Tem viabilidade financeira, se comparado aos programas de monitoramento da pesca tradicionais, pois adota como princípio, o estabelecimento de parcerias com instituições e comunidades locais, bem como o estabelecimento de elo de confiança entre pesquisadores e pescadores; Tem retorno social; Quando consolidado poderá ser a base para o manejo e gestão adaptativa de ecossistemas altamente alterados e que ainda não tenham iniciativas de monitoramento; O modelo pode ser adotado em locais que já tenham medidas de monitoramento da pesca sem prejuízos aos dados pretéritos coletados;
- **SIEPE:** garantiu um retorno de 100% das informações de pesca armazenadas, gerando múltiplas variáveis categóricas e numéricas para avaliação do MAP, acelerando as atividades de pré-processamento de dados. É um recurso de armazenamento e tratamento de informações da pesca inovador e promissor; Tem capacidade de integrar e relacionar dados pesqueiros de diferentes bacias hidrográficas, por meio de sua interface de exploração de dados, para apoiar a gestão pesqueira. artesanal;
- **Clusterização de Dados Mistas:** Foi capaz de expressar uma visão mais atualizada sobre o rendimento da pesca na bacia Araguaia-Tocantins; Indicou padrões por meio do agrupamento de dados utilizando variáveis numéricas e categóricas; Indicou que há pressão de pesca pelo Mapará, voador e tucunaré, desta forma são necessárias medidas de manejo para estes grupos de espécies; A potência do motor da embarcação influencia no rendimento da pesca e deve ser levada em consideração nos planos de manejo para a bacia Araguaia-Tocantins.

Em relação a trabalhos futuros, consideramos que

o desenvolvimento de programas de monitoramento da pesca na bacia Araguaia-Tocantins, devem levar em consideração as variáveis que se destacaram neste estudo: valor de venda do pescado, potência do motor da embarcação, peixes capturados e ambientes de pesca, como adicional a estatística pesqueira tradicional.

Para o melhor funcionamento do SIEPE é necessário realizar o desenvolvimento de versão para dispositivos móveis, dado que grande parte das localidades não possui oferta constante de serviços de banda larga, dificultando portanto o acesso ao SIEPE.

Há ainda a pretensão de desenvolvimento de modelos de análise, complementares às técnicas de clusterização, de forma a ainda relacionar os dados da pesca com recursos de geoprocessamento, de modo apoiar a gestão pesqueira em diferentes escalas.

Agradecimentos

Agradecemos às comunidades ribeirinhas de Santa Telesinha do Tauiri (PA) e Santo Antoninho (PA) pela colaboração e estabelecimento de parcerias.

Este trabalho foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- Alves, M. C. B. and Barthem, R. B. (2008). A Pesca Comercial dos "Tucunarés" Cichla spp. (Perciformes, Cichlidae) no Reservatório da UHE-Tucuruí, Rio Tocantins, PA, *Inst. Pesca* 34(4): 553–561.
- Anderberg, M. R. (1973). Cluster analysis for applications, *Probability and mathematical statistics*, 19 CN - QA278 .A5 1973 .
- Barthem, R. B., de Brito Ribeiro, M. C. L. and Petrere, M. (1991). Life strategies of some long-distance migratory catfish in relation to hydroelectric dams in the amazon basin, *Biological Conservation* 55(3): 339 – 345. [https://doi.org/10.1016/0006-3207\(91\)90037-A](https://doi.org/10.1016/0006-3207(91)90037-A).
- Batista, V. S., Inhamuns, A. J., Freitas, C. E. C. and Freire-Brasil, D. (1998). Characterization of the fishery in river communities in the low-Solimões/high-Amazon region, *Fisheries Management and Ecology* 5(5): 419–435. <https://doi.org/10.1046/j.1365-2400.1998.550419.x>.
- Beck, K. (1999). Embracing change with extreme programming, *Computer* 32(10): 70–77.
- Begossi, A. (1993). Ecologia humana: um enfoque das relações homem-ambiente, *Interciência* 18(3): 121–132.
- Cetra, M. and Petrere, M. (2001). Small scale fisheries in the middle river tocantins, imperatriz (ma), brazil, *Fisheries Management and Ecology* 8: 153 – 162. <https://doi.org/10.1046/j.1365-2400.2001.00233.x>.
- da Silva, A. P. and de Farias, E. G. G. (2017). Caracterização participativa da frota pesqueira do Rio Araguaia – Tocantins, Brasil, *Magistra* 29(1): 80–90.
- da Silva Batista, V., Nahum, V. J. I., Fabré, N. N., de Almeida, O. T., Gonzalez, J. C. A., Ruffino, M. L. and da Silva, C. O. (2004). O estado da pesca na amazônia.
- da Silva Jr., U. L., Raseira, M. B., Ruffino, M. L., Batista, V. d. S. and Leite, R. G. (2017). Estimativas do Tamanho do Estoque de algumas Espécies de Peixes Comerciais da Amazônia a partir de Dados de Captura e Esforço, *Biodiversidade Brasileira* 7(1): 105–121.
- de Camargo, S. A. F. and Petrere Jr, M. (2004). Análise de risco aplicada ao manejo precaucionário das pescarias artesanais na região do reservatório da UHE-Tucuruí (Pará, Brasil), *Acta Amazonica* 34(3): 473–485. <https://doi.org/10.1590/s0044-59672004000300013>.
- FAO (2016). The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all, *Technical report*, Food and Agriculture Organization of the United Nations.
- Fraxe, T. d. J. P., Witkoski, A. C. and Pereira, H. d. S. (2007). Comunidades ribeirinhas amazônicas: memória, ethos e identidade, *Manaus*: EDUA p. 224.
- Gartside, D. and Kirkegaard, I. (2010). A history of fishing, *Interactions: Food, Agriculture And Environment* 2: 70–80.
- Hallwass, G. (2011). Ecologia Humana da Pesca e Mudanças Ambientais no Baixo Rio Tocantins , Amazônia Brasileira.
- Hallwass, G. (2015). *Etnoecologia e Pesca: Influência de unidades de conservação e aplicação do conhecimento ecológico local de pescadores no manejo e conservação dos recursos pesqueiros no Baixo Rio Tapajós, Amazônia brasileira*, PhD thesis, Universidade Federal do Rio Grande do Sul.
- Hallwass, G. and Silvano, R. A. M. (2015). Patterns of selectiveness in the Amazonian freshwater fisheries: implications for management, *Journal of Environmental Planning and Management* 59(9): 1537–1559. <https://doi.org/10.1080/09640568.2015.1081587>.
- Hallwass, G. et al. (2013). Behavioral and environmental influences on fishing rewards and the outcomes of alternative management scenarios for large tropical rivers, *Journal of environmental management* 128C: 274–282. <https://doi.org/10.1016/j.jenvman.2013.05.037>.
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values, *Data Mining and Knowledge Discovery* . <https://doi.org/10.1023/A:1009769707641>.
- Isaac, V. J., Da Silva, C. O. and Ruffino, M. L. (2008). The artisanal fishery fleet of the lower amazon, *Fisheries Management and Ecology* 15(3): 179–187. <https://doi.org/10.1111/j.1365-2400.2008.00599.x>.

- Juras, A. A., Cintra, I. H. A. and Ludovino, R. M. R. (2004). A pesca na área de influência da usina hidrelétrica de tucuruí, estado do pará. *Boletim técnico-científico do cepnor*, 4(1): 77–88.
- Kolding, J., Béné, C. and Bavinck, M. (2014). *Small-scale fisheries*, John Wiley & Sons, Ltd, chapter 22, pp. 317–331. <https://doi.org/10.1002/9781118392607.ch22>.
- Lindenmayer, D. B. and Likens, G. E. (2009). Adaptive monitoring: a new paradigm for long-term research and monitoring, *Trends in Ecology and Evolution* . <https://doi.org/10.1016/j.tree.2009.03.005>.
- Lindenmayer, D. B., Likens, G. E., Haywood, A. and Miezis, L. (2011). Adaptive monitoring in the real world: Proof of concept. <https://doi.org/10.1016/j.tree.2011.08.002>.
- MMA (2006). Programa de Estruturação Institucional para Consolidação da Política Nacional de Recursos Hídricos – BRA/OEA/01/002, *Technical report*, Ministério do Meio Ambiente, Brasília–DF.
- Nunes, J. L. G. (2015). Estimador da produtividade para as pescarias artesanais do rio xingu.
- Oviedo, A. F. P. and Bursztyn, M. (2017). Decentralization and Fisheries Management in The Brazilian Amazon: Resource Rights and Accountability, *Ambiente & Sociedade* . <https://doi.org/10.1590/1809-4422asoc0029r1v2042017>.
- Petrere Jr, M. (1978). Pesca e esforço de pesca no Estado do Amazonas I – Esforço e captura por unidade de esforço, *Acta Amaz* 8(3): 439–454. <http://dx.doi.org/10.1590/1809-43921978083439>.
- Petrere Jr, M. and Peixer, J. (2007). O setor pesqueiro na amazônia: situação atual e tendências, *Ibama/PróVárzea*, 122 pp. . <https://doi.org/10.1590/s1519-69842007000300023>.
- Prystyon, A., Cunha, V. C. and Dias, C. R. G. (2019). The Fishing Productivity Assessment Upstream and Downstream of Tucuruí Hydroelectric Dam, Tocantins–Araguaia basin, Brazil, *International Journal of Advanced Engineering Research and Science* 6(4): 85–92. <https://doi.org/10.22161/ijaers.6.4.10>.
- Ruffino, M. L. (2004). Pesca e os recursos Pesqueiros na Amazônia brasileira, *IBAMA/PróVárzea, Manaus-AM*, 272pp. .
- Ruffino, M. L. (2008). Sistema integrado de estatística pesqueira para a amazônia, *Pan-American Journal of Aquatic Sciences* 3(3): 193–204.
- Sci, F. et al. (2014). The panorama of artisanal fisheries of the araguaia river, brazil, *Fisheries Science* 81. <https://doi.org/10.1007/s12562-015-0853-z>.
- Silva, C., Costa, R. d. and Lopes, P. (2010). A exploração de anomalocardia brasiliiana (gmelin, 1791)(mollusca, bivalvia, veneridae) por marisqueiros de grossos (rn) sob a perspectiva da teoria do forrageamento ótimo, *13º Simpósio de Biologia Marinha* .