

**Universidade de São Paulo  
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelagem de redes neurais artificiais MLP para previsão de vazões na  
bacia do rio Miranda afluente do Pantanal**

**Christian Pascal Silva Bouix**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área de concentração: Engenharia  
de Sistemas Agrícolas

**Piracicaba  
2024**

**Christian Pascal Silva Bouix**  
**Engenheiro Agrônomo**

**Modelagem de redes neurais artificiais MLP para previsão de vazões na bacia do rio  
Miranda afluente do Pantanal**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:

Profa. Dra: **PATRICIA ANGÉLICA ALVES MARQUES**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área: Engenharia de Sistemas  
Agrícolas

**Piracicaba**  
**2024**

**Dados Internacionais de Catalogação na Publicação  
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Silva, Christian Pascal Bouix

Modelagem de redes neurais artificiais MLP para previsão de vazões na  
bacia do rio Miranda afluente do Pantanal / Christian Pascal Silva Bouix - -  
versão revisada de acordo com a resolução CoPGr 6018 de 2011. - -  
Piracicaba, 2024.

67 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura "Luiz  
de Queiroz".

1. Hidrologia 2. Bacias hidrográficas 3. Inteligência artificial 4. Redes  
neurais artificiais 5. Modelagem I. Título

*Para a Sylvia e Tiana, as mulheres da minha vida,*

*Dedico*

## AGRADECIMENTOS

À Deus, por me dar sabedoria e força para enfrentar uma jornada totalmente nova na minha vida.

À minha família, principalmente para minha mãe Sylvia e minha avó Maria Sebastiana, por compreenderem meus períodos de ausência e me guiarem para minha evolução pessoal.

À professora Dr<sup>a</sup> Patricia Angélica Alves Marques, que acreditou nas minhas capacidades e me orientou, tanto na pesquisa quanto na minha vida pessoal.

Ao professor Sérgio Nascimento Duarte, por sempre estar presente e compartilhar seus conhecimentos, me orientando durante a condução da minha pesquisa.

Ao professor Carlos Roberto Padovani, que esteve a todo momento presente ao meu lado contribuindo com novas ideias e me ensinando sobre ciência.

À professora Dr<sup>a</sup> Roseli Aparecida, que me ensinou sobre novas tecnologias e contribuiu grandemente com ideias para a melhoria do trabalho.

Aos meus amigos, especialmente Vinícius Villa e Carlos Quiloango que além de pesquisadores incríveis, foram meus grandes amigos durante o período do mestrado e os levarei para sempre em minha vida.

À minha namorada Mariane Buglioli Sousa, por todo o amor e por me fazer infinitamente mais forte em todos os momentos. Sou imensamente grato por ter você na minha vida.

À minha amiga Julia Sconfienza, por todos os momentos de risadas e “expulsões” que proporcionou.

À Laís e Martiliana, que me introduziram um mundo novo durante minha entrada no mestrado.

À Thaís Charles, por todos os momentos de vida pessoal, profissional e acadêmica que compartilhamos juntos.

Aos meus demais parceiros e amigos da ESALQ: Bruna, Junior, Ana, Silas, Larissa, Fernanda, Iedo, Francisco, André, Alan, Bruno e Marcos. Sou grato por todos vocês fazerem parte dessa jornada junto comigo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

À toda a equipe do Serviço de Pós-Graduação, especialmente Davilmar, pelas suas orientações durante o processo de depósito da dissertação.

Finalmente, gostaria de agradecer a todos que me ajudaram de forma direta ou indireta para a conclusão do meu mestrado.

Obrigado.

*“A inteligência não é um privilégio, é um dom que deve ser usado para o bem da humanidade.”*

***Otto Octavious***

## SUMÁRIO

RESUMO.....	8
ABSTRACT.....	10
LISTA DE FIGURAS.....	11
LISTA DE TABELAS.....	13
1. INTRODUÇÃO .....	15
2. REVISÃO DE LITERATURA.....	19
2.1. Importância da previsão de vazões em uma bacia hidrográfica.....	19
2.2. Bioma Pantanal.....	19
2.3. Bacia Hidrográfica do Rio Miranda.....	20
2.4. Variáveis influentes na vazão de uma bacia hidrográfica.....	22
2.5. Uso de coleções de imagens satélite para monitoramento climatológico no Pantanal.....	22
2.6. Modelos de IA como alternativa aos modelos hidrológicos.....	23
2.7. Redes Neurais Artificiais do tipo <i>Multi Layer Perceptron</i> (MLP) e seu uso como ferramenta de previsão de vazões em bacias hidrográficas.....	24
2.8. Métricas de avaliação das redes neurais artificiais do tipo MLP .....	28
3. MATERIAL E MÉTODOS.....	31
3.1. Caracterização da área de estudo.....	31
3.2. Dados.....	34
3.2.1. Vazão.....	34
3.2.2. Pluviometria.....	35
3.2.3. Evapotranspiração.....	36
3.3. Pré-processamento dos dados.....	37
3.4. Construção das redes neurais artificiais.....	38
3.5. Análises das métricas dos modelos .....	39
4. RESULTADOS E DISCUSSÃO.....	41
4.1. Análise exploratória dos dados.....	41
4.1.1. Precipitação.....	42
4.1.2. Evapotranspiração.....	43
4.1.3. Vazão.....	44

4.2. Redes neurais com dados em escala diária.....	46
4.3. Redes neurais com dados suavizados por médias móveis.....	49
4.4. Redes neurais com dados em escala mensal.....	54
4.5. Redes neurais com retirada de variáveis dos modelos.....	56
5. CONSIDERAÇÕES FINAIS.....	59
6. CONCLUSÕES.....	61
REFERÊNCIAS.....	63



## RESUMO

### **Modelagem de redes neurais artificiais MLP para previsão de vazões na bacia do rio Miranda afluente do Pantanal**

A crescente necessidade global de prever eventos extremos e desastres naturais, intensificada pelas instabilidades ambientais, é evidente no Brasil, destacando-se o impacto no bioma do Pantanal. Considerado como uma das maiores superfícies alagadas do planeta, esse bioma enfrenta desafios socioeconômicos devido a sua dinâmica de cheias e vazantes, promovendo períodos frequentes de inundações e secas. Uma das formas de quantificar o volume de água que abastece o Pantanal é a partir do monitoramento das bacias hidrográficas do planalto adjacente que circundam o bioma. O monitoramento hidrológico é crucial para compreender e mitigar danos, sendo a modelagem hidrológica convencional uma ferramenta valiosa, apesar de sua complexidade de aplicação e sua dependência por associar-se a outras ferramentas, como softwares de SIG. Somado a isso, a região pantaneira enfrenta deficiência em estações pluviométricas, impulsionando a busca por alternativas na obtenção de dados, como coleções de imagens de satélite, para suprir lacunas de informações. A integração de Aprendizado de Máquina (AM) e Inteligência Artificial (IA) emerge como uma solução, possibilitando a análise de padrões nos dados climatológicos que podem ser aplicados em estudos de previsões de vazão. Dentre as técnicas de IA, as redes neurais artificiais (RNA) se destacam, simulando o aprendizado do cérebro humano e apresentando potencial para prever o comportamento hidrológico de bacias hidrográficas. Assim, este trabalho propõe a hipótese de que algoritmos de IA, utilizando a técnica de RNA, podem compreender e prever o fluxo de água de uma bacia de planalto componente do Pantanal. Diante disso, o objetivo geral do trabalho proposto consistiu em realizar a modelagem de diferentes redes neurais artificiais do tipo Multilayer Perceptron (MLP) para a previsão de vazões na bacia hidrográfica do rio Miranda, afluente do bioma Pantanal. Foram utilizadas séries históricas de vazão, precipitação e evapotranspiração como entradas dos modelos, por estas serem componentes principais do ciclo hidrológico. Os dados de vazão foram coletados a partir de postos fluviométricos presentes nos municípios de Miranda/MS e Bonito/MS e os dados de evapotranspiração e precipitação foram extraídos a partir de coleções de imagens de satélite. Em sequência, foi feito o pré-processamento dos dados e verificado a existência de dados faltantes nas séries para o período de interesse e todas as informações foram sincronizadas na escala diária. Foi usada a técnica de Médias Móveis para a suavização de dados, visando reduzir a variabilidade presente nas informações coletadas. Foi realizado o estudo de cada variável, independentemente, para as porções de montante e total da sub-bacia do rio Miranda. Em sequência foram testadas seis arquiteturas de redes neurais artificiais do tipo MLP em quatro blocos de análise: a) Redes neurais com dados em escala diária; b) Redes neurais com dados suavizados por Médias Móveis; c) Redes neurais com dados em escala mensal e d) Redes neurais com retirada de variáveis dos modelos. A avaliação de performance dos modelos se deu através das métricas estatísticas de Erro médio absoluto (MAE), Raiz quadrada do erro médio (RMSE) e coeficiente de determinação ( $R^2$ ), após as etapas de teste e validação dos modelos. Os resultados obtidos comprovam que o desempenho de uma rede neural do tipo MLP é bastante variável de acordo com os hiper parâmetros da rede escolhidos, assim como o número de neurônios e camadas intermediárias utilizadas em suas arquiteturas. A técnica de suavização por Médias Móveis e a utilização de dados mensais melhorou a performance e taxas de previsão dos modelos propostos em comparação com os dados em escala diária. Dessa forma, conclui-se que a utilização de redes neurais artificiais do tipo MLP é uma técnica promissora para a previsão de vazões em áreas de bacias hidrográficas pouco monitoradas e que carecem de informações importantes sobre seu comportamento hidrológico.

Palavras-chave: Hidrologia, Bacias hidrográficas, Inteligência artificial, Multilayer perceptron, Aprendizado de máquina

## ABSTRACT

### **Modeling of MLP artificial neural networks for flow forecasting in the Miranda river basin, a tributary of the Pantanal**

The growing global need to predict extreme events and natural disasters, heightened by environmental instabilities, is evident in Brazil, particularly impacting the Pantanal biome. As one of the largest wetland areas globally, this biome faces socio-economic challenges due to its dynamic flooding and drying cycles, resulting in frequent periods of floods and droughts. Quantifying the water volume that sustains the Pantanal involves monitoring the adjacent plateau watersheds surrounding the biome. Hydrological monitoring is crucial for understanding and mitigating damages, with conventional hydrological modeling being a valuable tool, despite its complex application and dependence on other tools like GIS software. Additionally, the Pantanal region lacks sufficient rainfall stations, driving the search for alternative data sources such as satellite image collections to fill information gaps. The integration of Machine Learning (ML) and Artificial Intelligence (AI) emerges as a solution, enabling the analysis of patterns in climatological data applicable to flow prediction studies. Among AI techniques, Artificial Neural Networks (ANNs) stand out, simulating human brain learning and demonstrating potential in predicting the hydrological behavior of watersheds. Therefore, this work hypothesizes that AI algorithms, employing ANN techniques, can understand and predict water flow in a plateau basin component of the Pantanal. The overall objective of the proposed work was to model different Multilayer Perceptron (MLP) artificial neural networks for flow prediction in the Miranda River watershed, a Pantanal biome tributary. Historical flow, precipitation, and evapotranspiration data were used as model inputs, as they are key components of the hydrological cycle. Flow data were collected from river gauging stations in Miranda/MS and Bonito/MS, while evapotranspiration and precipitation data were extracted from satellite image collections. Data preprocessing was conducted, and missing data for the period of interest were addressed, synchronizing all information on a daily scale. Moving Averages were applied for data smoothing to reduce variability in the collected information. Each variable was independently studied for the upstream and total portions of the Miranda River sub-basin. Subsequently, six MLP artificial neural network architectures were tested in four analysis blocks: a) Neural networks with daily scale data; b) Neural networks with smoothed data by Moving Averages; c) Neural networks with monthly scale data; d) Neural networks with variable removal from models. Model performance evaluation utilized statistical metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and coefficient of determination ( $R^2$ ), after model testing and validation stages. The results confirm that the performance of an MLP neural network is highly variable depending on the chosen network hyperparameters, including the number of neurons and intermediate layers used in their architectures. The Moving Averages smoothing technique and the use of monthly data improved the performance and prediction rates of the proposed models compared to daily scale data. Therefore, it is concluded that the use of MLP artificial neural networks is a promising technique for flow prediction in poorly monitored watershed areas lacking essential information about their hydrological behavior.

Keywords: Hydrology, Watersheds, Artificial intelligence, Multilayer perceptron, Machine learning

## LISTA DE FIGURAS

Figura 1 – Esquema representativo de um neurônio artificial.....	25
Figura 2 – Gráfico de intervalo da normalização da função ReLU.....	26
Figura 3 – Representação esquemática de uma rede neural artificial.....	27
Figura 4 – Mapa de localização da Bacia Hidrográfica do Rio Miranda e Sub-bacias delimitadas.....	32
Figura 5 - Diferença de altitudes das estações fluviométricas.....	35
Figura 6 – Arquitetura geral das redes neurais artificiais testadas.....	39
Figura 7 – Matriz de correlação contendo dados de precipitação, evapotranspiração e vazão das estações de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) para as porções montante e jusante da sub-bacia do rio Miranda.....	41
Figura 8 – Precipitação diária de 2000 a 2020 das porções à montante e à jusante da sub-bacia do rio Miranda.....	42
Figura 9 – Box plot dos valores diários médios de precipitação para os meses de janeiro a dezembro dos anos de 2000 a 2020 da porção montante e jusante da sub-bacia do rio Miranda.....	42
Figura 10 – Evapotranspiração real diária de 2000 a 2020 das porções à montante e à jusante da sub-bacia do rio Miranda.....	43
Figura 11 – Box plot dos valores diários médios de evapotranspiração para os meses de janeiro a dezembro dos anos de 2000 a 2020 da porção montante e jusante da sub-bacia do rio Miranda.....	44
Figura 12 – Hidrograma de vazões diária de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) da sub-bacia do rio Miranda.....	44
Figura 13 – Box plot dos valores diários médios de vazão para os meses de janeiro a dezembro dos anos de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e Miranda (cód. 66910000), da sub-bacia do rio Miranda.....	45
Figura 14 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 5 a partir dos dados diários.....	48
Figura 15 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 6 a partir dos dados diários.....	49
Figura 16 – Box Plot e Hidrograma de vazões diária de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) da sub-bacia do rio Miranda suavizados por médias móveis em escala decencial.....	51

Figura 17 – Box Plot de vazões mensais médias e hidrograma de vazões diária de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) da sub-bacia do rio Miranda suavizados por médias móveis em escala quinzenal.....	52
Figura 18 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 5 a partir dos dados suavizados em escala decenal.....	53
Figura 19 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 3 a partir dos dados suavizados em escala quinzenal.....	54
Figura 20 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 3 a partir dos dados suavizados em escala mensal.....	56
Figura 21 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 2 com retirada de variáveis a partir dos dados suavizados em escala mensal.....	58

**LISTA DE TABELAS**

Tabela 1 – Caracterização das porções da sub-bacia do rio Miranda.....	33
Tabela 2 – Estações fluviométricas e pluviométricas dos municípios de Miranda e Bonito.....	33
Tabela 3 – Resumo das variáveis utilizadas no estudo.....	37
Tabela 4 – Resultado da análise das métricas dos modelos testados com dados diários durante o treinamento, validação e teste.....	46
Tabela 5 – Tabela resumo das métricas dos modelos testados com dados diários durante o treinamento, validação e teste.....	47
Tabela 6 – Tabela resumo das métricas dos modelos testados com dados suavizados em escala decencial durante o treinamento, validação e teste.....	50
Tabela 7 – Tabela resumo das métricas dos modelos testados com dados suavizados em escala quinzenal durante o treinamento, validação e teste.....	50
Tabela 8 – Tabela resumo das métricas dos modelos testados com dados suavizados em escala mensal durante o treinamento, validação e teste.....	55
Tabela 9 – Tabela resumo das métricas dos modelos testados com retirada de variáveis em escala diária durante o treinamento, validação e teste.....	56
Tabela 10 – Tabela resumo das métricas dos modelos testados com retirada de variáveis em escala mensal durante o treinamento, validação e teste.....	57



## 1. INTRODUÇÃO

A necessidade de prever eventos extremos e desastres naturais tem se tornado necessária no Brasil e no mundo devido às instabilidades ambientais cada vez mais severas. Cenários como enchentes, inundações e deslizamentos ocasionadas por períodos de chuva intensas têm sido vistos com frequência ao redor do globo, e essas situações têm como um dos seus principais causador as mudanças climáticas, sobretudo os períodos de eventos extremos.

Essa é uma problemática que afeta o bioma do Pantanal brasileiro, uma das maiores superfícies alagadas do planeta. Visto que as alterações na hidrologia desse bioma mudam a vista da paisagem natural e influenciam as atividades econômicas locais como a pecuária e a piscicultura, a sua dinâmica de cheias e vazantes tem consequências diretas sobre a vida da população pantaneira. Embora seja, intrinsecamente, um bioma de superfície alagada, um dos principais problemas ambientais e socioeconômicos que afetam o Pantanal são os períodos frequentes de inundação quase permanente que essa região enfrenta (GALDINO et al., 2002). O Pantanal está situado na porção central da Bacia Hidrográfica do Alto Paraguai (BAP), cujas sub-bacias adjacentes do planalto geram vazões que abastecem o bioma e promovem a alternância entre períodos de seca e enchentes (COLLISCHON et al., 2007). Uma forma de quantificar o montante de água que entra na planície pantaneira é o monitoramento das sub-bacias localizada no seu planalto circundante.

Na hidrologia, as áreas de monitoramento e previsão apresentam grande importância, pois uma vez caracterizado o ciclo hidrológico de uma região, consegue-se compreender os padrões de comportamento do mesmo, tornando-se possível pensar e elaborar estratégias visando a redução de danos socioeconômicos e ambientais gerados por esses impactos. Na área de recursos hídricos, uma das maneiras de monitoramento do comportamento de uma bacia hidrográfica é feita através da previsão de vazões utilizando modelagem hidrológica.

Embora esses modelos apresentem grande aplicabilidade como ferramenta de modelagem hidrológica, uma das suas desvantagens reside na grande quantidade de variáveis envolvidas e na extensa sequência de etapas necessárias para sua aplicação. Além disso, alguns desses modelos funcionam em associação a algum *software* de SIG, o que significa que eles não são completamente independentes para sua utilização. Além disso, a complexidade do monitoramento hidrológico é afetada em função de aspectos como custo e disponibilidade de recursos financeiros, acesso restrito a pontos de interesse, distância entre as estações nos grandes centros, necessidades momentâneas (ABREU, 2021; COSTI, 2012) e ainda à baixa disponibilidade de dados de estações fluviométricas e pluviométricas em algumas regiões do país.

De forma geral, a Bacia do Alto Paraguai apresenta deficiência em sua densidade de estações pluviométricas (COLLISHONN et al., 2007; PADOVANI, 2010), sendo necessário buscar fontes alternativas ou utilizar outras ferramentas e tecnologias para a obtenção dessas informações. Visto que



a falta no detalhamento satisfatório de dados e informações sobre fenômenos caráter ambiental tem sido um obstáculo nas pesquisas de eventos extremos no Brasil (LOHMANN, 2013), uma alternativa para o monitoramento de dados climatológicos é o uso de informações provenientes de produtos de satélite (GUIMARÃES et al., 2018; PEREIRA et al., 2021; MARENGO, et al., 2022). Assim, dados como precipitação e evapotranspiração podem ser inseridos em modelos de Aprendizado de Máquina (AM) e Inteligência Artificial (IA) para mapear padrões existentes nessas informações e compreender o ciclo hidrológico de uma bacia hidrográfica.

A área da programação e modelagem computacional tem se inserido cada vez mais na área da hidrologia, sendo uma ferramenta utilizada em condução de estudos sobre regionalização hidrológica (DINIZ, 2008); preenchimento de falhas em dados de precipitação (RUEZZENE et al., 2021) e previsões de vazão em bacias hidrográficas (ULIANA et al., 2020; SANTOS NETO et al., 2020; MENDONÇA et al., 2021). Diversas são as metodologias dentro da inteligência artificial que são utilizadas para estudos hidrológicos e uma das técnicas utilizada nos trabalhos sobre previsão de vazões em bacias hidrográficas é o desenvolvimento de modelos empíricos guiados por dados ou *Data Driven Models* (SANTOS; CELESTE, 2014), nos quais se destacam as redes neurais artificiais.

As redes neurais artificiais ou Artificial Neural Networks (RANs ou ANNs) são um sistema de processamento de informações que simula a habilidade de um cérebro humano de aprender e reconhecer padrões a partir da tentativa e erro (PATIL; DEKA, 2016). Esse sistema possui uma estrutura disposta em camadas nas quais a resposta de saída da primeira funciona como combustível para alimentação da camada seguinte. Em seu processo de codificação, as RNAs utilizam algoritmos capazes de se ajustarem automaticamente e aprenderem com sua própria experiência, a partir da identificação de padrões nos dados fornecidos. Trabalhos com IA vem sendo realizados visando automatizar tarefas de forma rápida e precisa, além de auxiliar na tomada de decisões em processos complexos (MAGETO et al., 2020), tornando-se uma ferramenta promissora para obter bons resultados em previsões hidrológicas.

Dessa forma, tem-se como hipótese deste trabalho que algoritmos de machine learning utilizando a técnica de redes neurais artificiais podem compreender o comportamento hidrológico de uma bacia à montante do Pantanal, de forma a ter previsões do volume de fluxo de água que adentra nesse bioma. Diante do exposto, o objetivo geral do trabalho é realizar a modelagem de diferentes redes neurais artificiais do tipo Multilayer Perceptron (MLP) para a previsão de vazões na bacia hidrográfica do rio Miranda afluente do bioma Pantanal a partir de séries históricas de vazão, precipitação e evapotranspiração, que são componentes essenciais para entendimento do ciclo hidrológico.

Os objetivos específicos do trabalho proposto foram:

- a) Entender o comportamento de cada variável meteorológica coletada que funcionará como entrada da rede neural artificial, a fim de obter uma caracterização hidrológica da bacia de montante;
- b) Avaliar a variabilidade presente nos dados coletados;
- c) Entender a dificuldade de realizar previsões em bacias hidrográficas, sobretudo em regiões onde existe uma deficiência de aquisição de informações hidrológicas e bacias hidrográficas pouco monitoradas;
- d) Modelagem de redes neurais artificiais do tipo MLP testando diferentes arquiteturas visando a previsão de vazões que alimentam o bioma Pantanal;
- e) Analisar as melhores arquiteturas de rede neural a partir de métricas estatísticas.



## **2. REVISÃO DE LITERATURA**

### **2.1 Importância da previsão de vazões em uma bacia hidrográfica**

A previsão de vazões tem uma grande importância no gerenciamento de recursos hídricos. Uma das técnicas empregadas para mitigar os efeitos das incertezas climáticas na gestão de recursos hídricos é a previsão de vazão (SOUZA; SOUZA, 2010). De acordo com Malfatti, Cardoso e Hamburguer (2018), o estudo da vazão dos cursos d'água apresenta informações sobre os padrões hidrológicos dos rios, que permitem o planejamento e tomada de decisão.

Em países como o Brasil, onde a geração majoritária de energia provém de redes hidroelétricas, a posse dessas informações desempenha um papel crucial no controle do montante de energia gerada para abastecer o país. O sistema de planejamento das hidroelétricas brasileiras é altamente influenciado pelas variações sazonais e anuais na precipitação (LUIZ SILVA et al., 2019). Portanto, a realização de previsões de variáveis do ciclo hidrológico desempenha um papel fundamental no planejamento dessas usinas, permitindo a gestão eficaz e a garantia de um suprimento estável de energia elétrica.

Estudos sobre a previsão de vazões tem papel importante na prevenção de período de seca ou inundações extremas. A necessidade de previsão de eventos extremos para promover procedimentos de mitigação antes da ocorrência de danos já foi identificada em estudos anteriores (ADIKARI et al., 2021).

### **2.2 Bioma Pantanal**

Um bioma é um conceito fisionômico (BATALHA, 2011), uma vez que leva em consideração o aspecto geral da vegetação, e que por sua vez é resultado da influência de certas formas de vida. O Pantanal é um bioma brasileiro de superfície alagada, rico em biodiversidade e é uma faixa de transição entre demais biomas, favorecendo uma troca entre diferentes espécies (WEBER; COUTO, 2008). Além de ser considerado como patrimônio nacional pela Constituição Federal de 1988, é rico em biodiversidade biológica e cultural (FERREIRA, 2013), que é favorecida pela sua dinâmica de cheias e de vazantes. Por possuir grande interação com ecossistemas terrestres e aquáticos, é considerado como um espaço de tensão ecológica com grande importância socioeconômica (PEREIRA; CHÁVEZ; SILVA, 2012). O bioma, reconhecido como uma das maiores superfícies alagadas do planeta, encontra-se situado na Bacia Alto Paraguai (BAP), uma bacia hidrográfica transfronteiriça que abrange áreas da Bolívia e Paraguai, ultrapassando as fronteiras nacionais brasileiras.

O Pantanal é dividido em uma grande planície alagada e seus planaltos adjacentes e muito da sua dinâmica hídrica e alterações em sua paisagem são influenciadas por essa estrutura. É possível correlacionar as mudanças que ocorrem na fisiografia do Pantanal ao conceito fisionômico de bioma

apresentado por Batalha (2011), que diz que um bioma é resultado da influência de certas formas de vida.

A principal atividade econômica desenvolvida no Pantanal é a agropecuária, e sobretudo na década de 70, a sua expansão contribuiu negativamente para alterações na paisagem e na dinâmica das águas desse bioma (GALDINO et al., 2002; GALDINO et al., 2006). Portanto, a busca por aumento de produtividade da pecuária no Pantanal, decorrente das pressões políticas, econômicas e sociais tem ocasionado ações de desmatamento para o cultivo de pastagens (CARDOSO et al., 2009).

Visto que a pecuária desenvolvida no Pantanal é realizada, sobretudo, nas áreas de seus planaltos circundantes, a paisagem desse bioma é altamente afetada. Segundo Silva, Abdon e Moraes (2010), para a Bacia do Alto Paraguai, essa questão é de extrema importância, pois os impactos ambientais que a afetam como assoreamento de rios, perda de habitats e inundações são decorrentes do transporte de sedimentos originados em seus planaltos carregados para a planície. De forma geral, o Pantanal apresenta um pulso de inundação por ano e esse evento retorna diferentes níveis de áreas inundadas, dependendo de fatores como as características geomorfológicas locais e a vegetação predominante (ALHO; SILVA, 2012).

Bergier (2013), corrobora que a planície pantaneira apresenta um aumento na vazão e escoamento superficial devido a influência das áreas do planalto. De acordo com a Agência Nacional de Águas (2020), os hidrogramas dos rios na saída do planalto para o Pantanal influenciam a amplitude de variação de nível e vazão em curtos períodos de tempo de dias ou de horas. Portanto, entende-se que os pulsos de inundação do Pantanal são influenciados, além da precipitação local do bioma, pelo escoamento de água que proveniente dos planaltos adjacentes, localizados à montante.

### **2.3 Bacia Hidrográfica do Rio Miranda**

A Bacia Hidrográfica do Rio Miranda (BHRM) encontra-se inserida na BAP e está situada na região centro-oeste do Estado do Mato Grosso do Sul, com área aproximada de 42.124 km<sup>2</sup> (LEITE; BEREZUK; SILVA, 2022). Essa bacia divide-se em duas sub-bacias menores denominadas Bacia do Rio Miranda e Bacia do Rio Aquidauana. Ambas são consideradas como bacias de planalto, e apresentam altitude que varia de 83 a 750 m, desde a foz do rio Miranda ao ponto mais alto, na Serra Bodoquena (MATO GROSSO DO SUL, 2012). As bacias do Miranda e Aquidauana apresentam diferentes geologias, solos e tipo predominante de vegetação. Na sub-bacia o rio Miranda existe maior ocorrência de solos argilosos, devido a presença de calcários e maiores áreas de florestas estacionais. Na sub-bacia do rio Aquidauana encontram-se solos arenosos em decorrência da presença de arenitos e vegetação característica de Cerrado (POTT; DAMASCENO-NUNIOR; SILVA, 2014).

Em relação aos usos da terra e cobertura vegetal, a BHRM apresenta majoritariamente, área destinada ao cultivo de pastagens (21.263 km<sup>2</sup>), ocupando pouco mais de 50% da área da bacia, também podendo ser encontradas áreas de formações florestais (26,22%), formações savânicas

(6,99%), formações campestres (5,31%), e área de cultivo de soja (5,86%) (LEITE; BEREZUK; SILVA, 2022).

A busca por aumento de produtividade da pecuária no Pantanal, decorrente das pressões políticas, econômicas e sociais foi um fator contribuinte para ações de desmatamento para o cultivo de pastagens. Visto que a pecuária desenvolvida no Pantanal é realizada, sobretudo, nas áreas de seus planaltos circundantes, a paisagem desse bioma é altamente afetada (CARDOSO et al., 2009). Para toda a BAP, essa questão é de extrema importância, pois os impactos ambientais que a afetam como assoreamento de rios, perda de habitats e inundações são decorrentes do transporte de sedimentos originados em seus planaltos carregados para a planície (SILVA; ABDON; MORAES, 2010)

Outro fator que merece atenção dentro da BHRM é a estabilidade e a propensão a processos erosivos dentro da bacia. Em estudo feito por Leite, Berezuk e Silva (2022), os autores quantificaram a estabilidade e vulnerabilidade ambiental da BHRM e concluíram que 66,14 % da área total da bacia classifica-se como “medianamente estável/vulnerável”. Esse total corresponde a todo o médio e baixo curso da bacia. Áreas com baixa estabilidade a erosão são um problema pois períodos de eventos extremos de chuva podem gerar um alto fluxo de vazão nos rios e favorecer deslizamentos e desastres ambientais.

O regime hidrológico do Pantanal é caracterizado pela sazonalidade dos níveis dos seus rios, onde para cada período de doze meses, o nível de água apresenta uma fase de enchente e outra de vazante, como exemplo do rio Paraguai em Ladário-MT (GALDINO et al., 2002). Dessa forma, a ocorrência de dois anos consecutivos de seca determina o início de um ciclo de seca, e a mesma lógica se aplica para os ciclos de cheia. Como visto por Hamilton (2002), as inundações do Pantanal como um todo são sazonais, e os pulsos de inundação são tipicamente unimodais em grande parte do bioma. Segundo Galdino et al. (2002), no período de 1974 a 2001 ocorreram três das maiores cheias do Pantanal, com níveis médios do rio Paraguai variando entre 3,52 e 6,64 m em seus valores extremos. O maior ciclo de cheia durou dez anos consecutivos (1964 a 1973), onde os valores da cota do rio variaram entre 61 cm e 2,74 m. A variação no nível dos rios e suas vazões também podem ser vistas para porções específicas do bioma. Os autores também destacam um comportamento errático nas planícies de inundação da porção sul do Pantanal, especialmente afetados pelos rios Miranda e Aquidauana.

Ao analisar a variabilidade das vazões dos postos fluviométricos do Aquidauana e do Miranda, no período de 1968 a 2002 e de 1969 a 1995, respectivamente, Curado (2003) identificou que as vazões médias mensais apresentam suas máximas nos meses de dezembro e janeiro, decrescendo até os meses de julho e agosto, onde apresentam suas mínimas. Para o período estudado pelo autor, observou-se uma grande variabilidade nas vazões máximas mensais para ambas as estações fluviométricas. O posto do Aquidauana apresentou variações entre valores próximos de  $150 \text{ m}^3 \cdot \text{s}^{-1}$  a

450 m<sup>3</sup>.s<sup>-1</sup> em seus pontos extremos. Para o posto do Miranda, as vazões máximas observadas variaram entre valores acima de 500 m<sup>3</sup>.s<sup>-1</sup> e abaixo de 100 m<sup>3</sup>.s<sup>-1</sup> nos seus valores extremos.

#### **2.4 Variáveis influentes na vazão de uma bacia hidrográfica**

O ciclo hidrológico é um sistema fechado que regula a distribuição de água na superfície terrestre, com entradas e saídas dependentes de várias variáveis. Tucci e Clarke (1997) consideram que o volume das bacias hidrográficas depende inicialmente dos processos verticais como precipitação e evapotranspiração. Esses fatores influenciam na vazão das bacias. Entende-se por vazão o volume de água que flui por uma seção do canal principal em um período de tempo específico (SOUZA; MOURA, 2022).

Desses processos, a precipitação desempenha um papel fundamental na regulação da vazão dos cursos d'água, como destacado por Charles (2020). No entanto, como mencionado por Souza, Souza e Cardoso (2017), a vazão dos rios não é influenciada apenas pela precipitação, mas também por outros fatores, como o tipo de uso e cobertura do solo, e dessa forma as variações na precipitação média nem sempre resultarão em influências diretamente proporcionais na vazão média dos rios. Contribuição de fatores como o escoamento subterrâneo também é relevante para a manutenção dos rios perenes, e é resultado da capacidade de infiltração de água no solo (TUCCI; CLARKE, 1997). Além disso, diferentes bacias e redes de drenagem apresentam diferentes características quanto à sua forma, declividade, densidade de cursos d'água, entre outros valores.

A fisiografia de uma bacia hidrográfica compreende essa descrição das características físicas e geográficas de uma rede de drenagem, e segundo Wenzel et al. (2017) essas informações são utilizadas para entender o comportamento hidrológico e tem aplicações em estudos de regionalização de vazões mínimas, médias e máximas. Essa relação complexa entre diferentes variáveis demonstra a necessidade de uma compreensão abrangente e integrada do ciclo hidrológico e seus determinantes para uma gestão eficaz dos recursos hídricos em uma bacia hidrográfica.

#### **2.5 Uso de coleções de imagens satélite para monitoramento climatológico no Pantanal**

De acordo com Geritana et al. (2009), extensas regiões do mundo ainda são pouco monitoradas devido às dificuldades na acessibilidade, especialmente para grandes bacias hidrográficas de regiões tropicais. Para o Pantanal Matogrossense, onde existe deficiência em densidade de estações pluviométricas (COLLISHONN et al., 2007; PADOVANI, 2010) estudos de caráter de monitoramento climático foram realizados utilizando coleções de imagens de satélite.

Pereira et al. (2021) realizaram o estudo de mapeamento da variabilidade e frequência de áreas inundadas no Pantanal em uma escala regional usando dados orbitais de resolução espacial média disponibilizadas pela coleção CHIRPS (Climate Hazard Group InfraRed Precipitation Satellite's) v. 2.0. Os autores concluíram que embora existam dificuldades devido à alta cobertura de nuvens das

imagens utilizadas, o método aplicado representa de forma satisfatória a variabilidade espacial e temporal das inundações do bioma.

O uso da coleção CHIRPS é documentado em outros estudos para o mapeamento de chuvas no Pantanal. Guimarães et al. (2018) utilizaram dados diários de 1981 a 2018 da base de dados CHIRPS, comparando-os com dados registrados em estações meteorológicas convencionais do Instituto Nacional de Meteorologia (INMET) em 23 municípios localizados no Pantanal. Os autores concluíram que a base pluviométrica CHIRPS apresentou resultados semelhantes às séries históricas das estações do INMET para os municípios de Cáceres/MT e Corumbá/MS, sendo uma importante base de dados de precipitação para os estudos na dinâmica dos ciclos de águas do Pantanal. Beck et al. (2017) avaliaram 22 quasi-global conjuntos de dados de precipitação provenientes de satélites e observaram que para regiões tropicais, a coleção CHIRPS v 2.0 é uma escolha viável se uma resolução temporal diária for suficiente para a análise hidrológica que se deseja fazer.

O uso de séries temporais de evapotranspiração a partir de produtos de satélite também é encontrado na literatura. Coleções como a Global Land Data Assimilation System (GLDAS) foram utilizadas por Ruhoff et al. (2022) para diferentes biomas presentes na América do Sul e os autores evidenciaram concordância satisfatória em relação às estimativas do balanço hídrico da evapotranspiração para o bioma Pantanal. As coleções como o CHIRPS e o GLDAS apresentam vantagem devido às possibilidades de uso de informações em escala temporal diária e com resolução espacial de  $0.05^\circ$  (aproximadamente 5,4 km) e  $0.25^\circ$  (aproximadamente 25 km), respectivamente.

## **2.6 Modelos de IA como alternativa aos modelos hidrológicos**

Compreender os fenômenos hidrológicos e realizar previsões baseando-se em dados não é uma tarefa trivial. A complexidade dos regimes hidroclimáticos, resultante de variações nos padrões climáticos, uso do solo e fatores fisiográficos, como a densidade de drenagem de uma bacia e a capacidade de retenção de água no solo, torna desafiadora a modelagem do ciclo hidrológico em grandes bacias (BELVEDERESI et al., 2020). Diversos são os modelos matemáticos clássicos utilizados na área da hidrologia, como os Autorregressivos (AR), Média Movel (MA), Média Móvel Autorregressiva (ARMA), Regressão Linear (LR) e Regressão Linear Múltipla (MLR). Esses modelos têm sido aplicados para previsão de vazões desde a década de 70, entretanto, esses modelos lineares não são capazes de traduzir integralmente a não estacionariedade e não linearidade dos processos hidrológicos (YASEEN et al, 2015). De acordo com Khan et al. (2023), esses modelos podem nem sempre ser precisos, pois não conseguem captar a não linearidade que ocorre em uma bacia hidrográfica durante a transformação da precipitação em escoamento.



Para a compreensão dos fenômenos de caráter hidrológico são utilizados os modelos hidrológicos convencionais. Esses modelos são divididos em: modelos conceituais, modelos empíricos e modelos de base física.

Segundo Jaiswal, Ali e Bharti (2020) e Uliana et al. (2020), os modelos empíricos são baseados em dados de observações e atuam sem a caracterização dos diferentes processos do sistema hidrológico, e utilizam equações que estabelecem relações entre os dados de entrada e saída do sistema hídrico; os modelos de base física utilizam conceitos matemáticos simples de um sistema ou se baseiam na solução de equações do sistema físico, com o auxílio de armazenamentos interconectados que representam diferentes aspectos de processos hidrológicos; por sua vez, os modelos conceituais realizam uma interpretação agrupada da natureza da bacia, usando os mesmos parâmetros de avaliação para toda a extensão da área de interesse, ignorando a variação espacial existente. Os três grupos de modelos variam no número de parâmetros de avaliação, na qualidade dos dados de entrada e na escala temporal de análise. Portanto, a escolha para sua utilização está função da necessidade da resolução da variável em estudo.

A complexidade intrínseca dos processos hidrológicos, assim como a dificuldade na adequação dos dados de entrada dos modelos tradicionais, são obstáculos que limitam a eficiência na condução de estudos sobre bacias hidrográficas. Dentro desse contexto, a área da Inteligência Artificial (IA) e Aprendizado de Máquina (AM) tem ganhado destaque nos últimos anos por automatizar processos e tarefas complexas, além de conseguir reconhecer padrões em bases de dados para realização de previsões. De acordo com dados levantados por Yaseen et al. (2015) e Ibrahim et al. (2022), a IA tem sido usada em diversos métodos dentro da hidrologia para a predição de parâmetros hidrológicos e tem obtido progresso em simulações, capturando a complexidade de bases de dados de processos não lineares.

O incremento contínuo do poder computacional das máquinas permitiu o desenvolvimento de redes complexas baseadas em ML, podendo ser aplicadas dentro da variedade das aplicações hidrológicas, sobretudo em tarefas computacionalmente demandantes, como previsões de vazão, previsões de máximas de inundação, manejo de recursos hídricos, entre outros (ROZOS; DIMITRIADIS & BELLOS, 2022). Com relação a previsões de inundações, Kim et al. (2021) identificaram que modelos de ML obtiveram melhores performances para regimes de altas e médias vazões quando comparados aos modelos hidrológicos padrões, indicando uma ferramenta prática para redução de danos econômicos ocasionados por inundações.

## **2.7 Redes Neurais Artificiais do tipo Multi Layer Perceptron (MLP) e seu uso como ferramenta de previsão de vazões em bacias hidrográficas**

Dentre as abordagens utilizadas na modelagem hidrológica com IA para a previsão de vazões, o uso das Redes Neurais Artificiais (RNAs) é bastante estudado. De acordo com Haykin (1994), uma RNA é um sistema de processamento de informações paralelamente distribuído que se assemelha às

redes neurais biológicas do cérebro humano. O primeiro conceito de neurônio artificial foi proposto por McCulloch e Pitts (1943), onde seu modelo matemático discutido conseguia reproduzir operações lógicas como “AND”, “OR” e “NOT”, espelhando o comportamento dos neurônios biológicos. Esse neurônio artificial, quando alimentado com dados de entrada, conseguia produzir saídas através de funções de ativação.

Esses estudos foram fundamentais para o desenvolvimento do modelo *Perceptron*, cujo algoritmo de aprendizado permite realizar o ajuste de pesos de cada valor de entrada de acordo com os erros cometidos durante a classificação de amostras no processo de treinamento. Dessa maneira, o *Perceptron* consegue aprender a reconhecer padrões. Esse modelo, proposto por Rosenblatt (1958) é a base para a construção de diversos modelos de IA guiados por dados (*Data Driven Models*), e é utilizado dentro das RNAs. Uma RNA é composta por uma estrutura de camadas, onde cada unidade de processamento é denominada de neurônios e é compreendida por suas entradas associadas a pesos, uma função de ativação e uma saída (Figura 1).

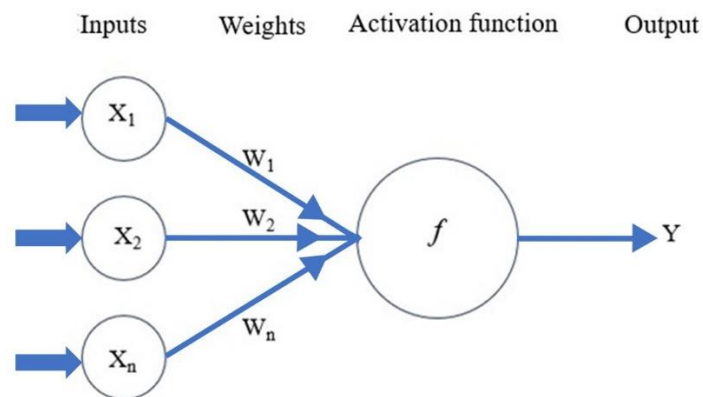


Figura. 1 – Esquema representativo de um neurônio artificial.  
Fonte: próprio autor (2023).

A camada de entrada consiste em um conjunto de neurônios ( $X_1$  a  $X_n$ ), cada um associado a seu peso respectivo ( $W_1$  a  $W_n$ ), uma função de ativação  $f$  que determina se o neurônio deve ou não ser ativado e transmitir o sinal para a próxima camada da rede. Cada neurônio é multiplicado pelo seu peso associado e somado ao seu viés. Em seguida, é feito o somatório dos produtos. Essa resultante é propagada para as demais camadas intermediárias da rede e essa transferência sequencial da informação é chamada de *feedforward*.

Cada neurônio da camada intermediária recebe o somatório do produto ocorrido anteriormente e esses valores são transferidos para a camada de saída após sua passagem pela função de ativação (RASHID; AHMAN, 2016). A relação matemática da função de ativação pode ser vista na equação (1).

$$f(a) = \sum_{i=1}^n (x_i w_{ij}) \quad (01)$$

Em que:

$X_i$  = Valor do neurônio  $i$ ;

$W_{ij}$  = Peso associado do neurônio  $i$

A função de ativação irá determinar o intervalo numérico da informação de saída, decidindo o nível de ativação do neurônio e as diferentes funções de ativação que podem ser utilizadas de acordo com os problemas de classificação ou regressão. Entretanto, a função Rectified Linear Unit (ReLU) é uma das funções mais usadas nas arquiteturas de redes neurais artificiais devido a seu baixo custo computacional em comparação com outras funções como a Tangente Hiperbólica (Tanh) e Sigmóide (ALI; SHAHBAZ, 2020). A função ReLU retorna zero para valores negativos e o próprio valor para valores não negativos, como mostrado de forma gráfica na Figura 2. Esquemáticamente, a função ReLU pode ser vista na equação (2):

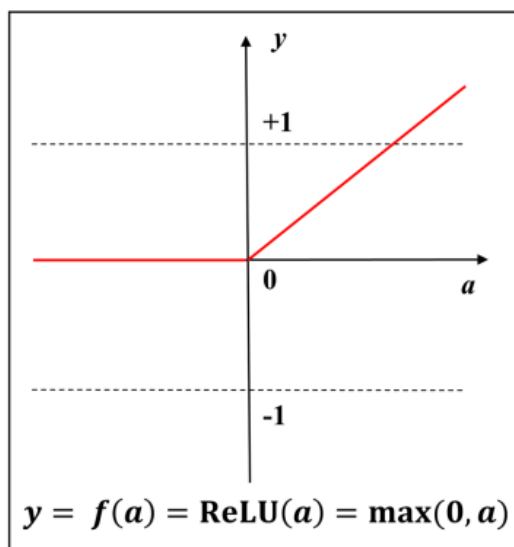


Figura 2 – Gráfico de intervalo da normalização da função ReLU.

Fonte: Ali e Shahbaz (2020).

$$f(a) = \{0, \text{for } a < 0 \ a, \text{for } a \geq 0 \quad (02)$$

Após a passagem pela primeira camada intermediária da rede, o processo é repetido até a sua última camada, que computa a saída desejada. Dessa forma, o neurônio da camada de saída (que contém o valor da camada de saída) computa o erro em relação a saída desejada do modelo e se o erro apresenta valor elevado e as previsões estiverem aquém das observadas, os pesos da inicialização do

processo de aprendizado e seus vieses são atualizados até que haja a minimização dos erros. Essas relações podem ser vistas conforme as equações (3) e (4) apresentadas por Ali e Shahbaz (2020):

$$E_k = (d_k - y_k)f'(\sum_{i=1}^m y_j W_{jk}) \quad (03)$$

$$\Delta W_{jk} = \alpha E_k y_k \quad (04)$$

Em que:

$\alpha$  = taxa de aprendizado do modelo;

$d_k$  = valor de saída observado;

$y_k$  = valor de saída predito pelo modelo;

$f'$  = derivada da função de ativação

A retropropagação do erro é realizada por algoritmo denominado *backpropagation* e este é amplamente utilizado em redes neurais do tipo MLP (ALOTAIBI et al., 2018). Dessa forma, uma RNA (Figura 3) consiste em uma rede interconectada com camadas, neurônios e pesos chamada de redes de múltiplas camadas ou redes neurais totalmente conectadas (KIM et al., 2021).

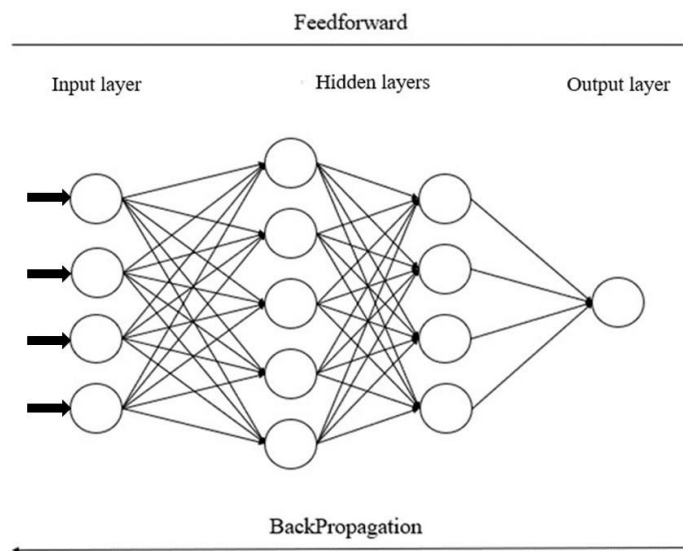


Figura 3 – Representação esquemática de uma rede neural artificial.  
Fonte: próprio autor (2023).

Por ser uma ferramenta capaz de trabalhar em diversos contextos, o uso de IA na hidrologia tem se difundido cada vez mais à nível global. Essa área de estudo tem sido alvo de aplicações de diversos modelos de IA objetivando a compreensão dos campos da modelagem de sedimentos, qualidade da

água, precipitação, previsão de nível, evapotranspiração, inundações, mapeamento de lençóis freáticos, entre outros (YASEEN et al., 2015; IBRAHIM et al., 2022).

A técnica de RNAs tem aplicações no desenvolvimento de algoritmos para previsão de vazões e cotas mensais em bacias brasileiras e tem obtido bons coeficientes de determinação nas suas previsões. Santos Neto et al. (2020) desenvolveram um modelo de previsão de cotas máximas mensais para uma bacia do rio Amazonas utilizando RNAs de Múltiplas Camadas, e todas suas arquiteturas de redes testada apresentaram coeficiente de determinação ( $R^2$ ) acima de 0.81, com base nos valores observados. Ainda para o bioma da Amazônia, Mendonça et al. (2021) testaram diferentes arquiteturas de modelos de RNA de Múltiplas Camadas e RNA Recorrentes do tipo NARX e identificaram que as redes recorrentes apresentaram um coeficiente de determinação superior (99,08% para sua fase de teste) do que as redes de Múltiplas Camadas (28,78%, também para fase de teste) ao prever vazões diárias médias de uma bacia.

Para o bioma Pantanal, estudos de caráter semelhante foram conduzidos. Santos Antoniassi (2022) realizou a previsão de níveis de rios utilizando dados de cota de estações à montante para prever o nível de estações à jusante da Região Hidrográfica do rio Paraguai. O autor testou modelos com diferentes algoritmos de AM, baseados na técnica de RNAs e obteve de 18,09 a 23,84% mais precisão do que modelos de regressão usualmente utilizados para as predições. Visando prever as cotas diárias na bacia do Rio Aquidauana, Cristaldo et al. (2018) testaram diferentes redes neurais de Múltiplas Camadas com o algoritmo de aprendizado *backpropagation* (BP) e Algoritmo Genético (AG) para a definição dos parâmetros das RNAs. Os autores concluíram que os índices de erro aumentaram em função do horizonte de previsão. Entretanto, para um horizonte de previsão de um dia, as arquiteturas testadas alcançaram um coeficiente de determinação de 95%, indicando boa precisão de previsão.

## 2.8 Métricas de avaliação das redes neurais artificiais do tipo MLP

O sucesso da modelagem de redes neurais artificiais vai variar de acordo com as respostas acertadas pela rede sobre a caracterização de um determinado problema. Dessa forma, cada rede neural é modelada visando mapear um tipo determinado de informação para um conjunto específico de dados. Na literatura, as métricas estatísticas mais utilizadas para a avaliação do desempenho de uma rede neural são dentro dos estudos hidrológicos para a previsão de séries temporais de vazão são a Raiz quadrada do erro médio (RMSE), o Coeficiente de determinação ( $R^2$ ) e o Erro médio absoluto (MAE) (APAYDIN et al., 2020; OLIVEIRA, RAMOS, NEVES, 2023; ZAKARIA et al., 2023). As relações matemáticas para cálculo das métricas podem ser vistas nas equações de (5) a (7). Para avaliar o desempenho, um coeficiente  $R^2$  mais próximo de 1 é indicativo de melhor desempenho, enquanto coeficientes RMSE e MAE mais próximos de 0 são sinais de um desempenho superior.

$$RMSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2 \quad (05)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - X_i| \quad (06)$$

$$R^2 = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{Y})^2}} \quad (07)$$

Em que:

$X_i$  é a variável observada com sua média atribuída a  $\bar{X}$ ;

$Y_i$  é a variável predita, com sua média atribuída a  $\bar{Y}$

$N$  é o número de instâncias.



### **3. MATERIAL E MÉTODOS**

Para a condução do trabalho, foi escolhida uma bacia hidrográfica componente do planalto circundante ao bioma Pantanal, pois bacias presentes nesta região servem como porta de entrada para quantificação do montante de precipitação e vazão que entram na planície pantaneira. Foi realizada a coleta e tratamento de dados de vazão, precipitação e evapotranspiração real, todos em escala diária, que serviram como dados de entrada das redes neurais testadas. Em seguida, foram montadas diferentes arquiteturas de redes neurais, variando seus hiperparâmetros e foi realizada a análise de comparação de eficiência dos modelos testados, avaliando as seguintes métricas: Erro médio absoluto (MAE), Raiz quadrada do erro médio (RMSE), o coeficiente de eficiência de Nash Sttucliffe e coeficiente de determinação ( $R^2$ ).

#### **3.1 Caracterização da área de estudo**

A área de estudo do presente trabalho é a Bacia Hidrográfica do Rio Miranda (BHRM), localizado no Estado do Mato Grosso do Sul (Figura 4). A área total compreendida pela BHRM é de aproximadamente 42.023 km<sup>2</sup>, onde encontram-se o rio Miranda e o rio Aquidauana como os principais rios das sub-bacias. Parte da BHRM encontra-se dentro do bioma Pantanal, sendo as sub-bacias do Miranda e Aquidauana duas portas de entrada da vazão que chega ao bioma. Para este estudo foi considerada a Bacia do Rio Miranda (BRM).



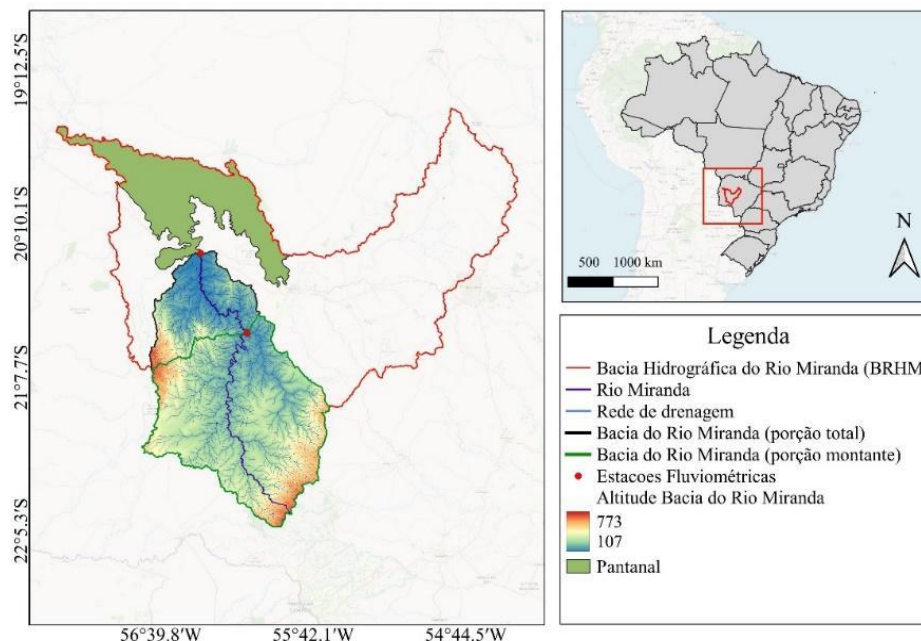


Figura 4 – Mapa de localização da Bacia Hidrográfica do Rio Miranda e Sub-bacias delimitadas. Fonte: próprio autor (2023).

A BRM foi dividida em duas porções, delimitadas a partir do QGIS (ver. 3.28.0). A porção total considera a estação fluviométrica do município de Miranda (cód. 66910000) como ponto de exultório e essa delimitação compreende a área completa da sub-bacia até o final da extensão do rio Miranda. A porção montante tem como ponto de exultório a estação fluviométrica de Estrada MT-738 (cód. 66900000), localizada no município de Bonito e a delimitação compreende aproximadamente da porção média da sub-bacia até o final da extensão do rio Miranda.

A delimitação da sub-bacia em duas porções foi feita visando estabelecer a relação de causa e efeito entre as estações fluviométricas de montante-jusante. As vazões medidas pela estação da porção montante funcionará como um indicativo das vazões da estação à jusante da bacia, sendo possível durante a fase de treinamento das redes neurais, compreender esses padrões nos dados e estabelecer essa relação de causa e efeito entre as variáveis de entrada e saída da rede. A caracterização das porções total e montante da sub-bacia estão detalhadas na Tabela 1. Dentro da área da bacia do rio Miranda existem algumas estações pluviométricas e fluviométricas responsáveis pela Agência Nacional de Águas (ANA) e pelo Instituto de Meio Ambiente de Mato Grosso do Sul (IMASUL-MS), detalhadas na Tabela 2.

Tabela 1 – Caracterização das porções da sub-bacia do rio Miranda.

Porção da bacia	Área (km <sup>2</sup> )	Perímetro (km)	Comp. Trecho Rio (km)	Alt. Mín. (m)	Alt. Máx. (m)
BH Miranda (porção total)	15.190,99	686,70	468,46	107,00	773,00
BH Miranda (porção montante)	11.552,14	589,92	292,03	135,00	754,00

Fonte: próprio autor (2023).

Tabela 2 – Estações fluviométricas e pluviométricas dos municípios de Miranda e Bonito.

Nome da estação	Código	Município	Localização	Área de drenagem (km <sup>2</sup> )	Tipo	Responsável
<b>Miranda</b>	<b>66910000</b>	<b>Miranda</b>	<b>20°24' S, 56°39' O</b>	<b>15.000,00</b>	<b>Fluviométrica</b>	<b>ANA</b>
<b>Estrada MT-738</b>	<b>66900000</b>	<b>Bonito</b>	<b>20°76' S, 56°09' O</b>	<b>11.600,00</b>	<b>Fluviométrica</b>	<b>ANA</b>
PONTE RODOVIA BR-262 - Miranda / Bodoquena	66911000	Miranda	20°24' S, 56°39' O	15.000,00	Fluviométrica	IMASUL-MS
Miranda	66910500	Miranda	20°24' S, 56°39' O	Não informado	Fluviométrica	IMASUL-MS
Bonito	66896000	Bonito	20°76' S, 56°09' O	11.600,00	Fluviométrica	IMASUL-MS
Miranda	2056001	Miranda	20°24' S, 56°39' O	n.a*	Pluviométrica	ANA
Miranda	2056006	Miranda	20°23' S, 56°38' O	n.a*	Pluviométrica	ANA
Estrada MT-738	2056003	Bonito	20°76' S, 56°09' O	n.a*	Pluviométrica	ANA

\*não aplicável. Valores em negrito foram utilizados como ponto de exultório para a delimitação das porções das sub-bacias em estudo. Fonte: Agência Nacional de Águas, 2023.

As informações de precipitação e vazão dessas estações estão disponíveis no Portal Hidroweb do Sistema Nacional de Informações de Recursos Hídricos (SNIRH) da ANA, disponível através do endereço eletrônico: <https://www.snirh.gov.br/hidroweb/serieshistoricas>. Entretanto, as estações existentes concentram-se somente nos municípios de Miranda/MS e Bonito/MS, não abrangendo a totalidade da área da bacia, confirmando a baixa densidade de estações dessa região já apontada por Padovani (2010) e Collishonn et al. (2007). A Organização Meteorológica Mundial (WMO, 2008) recomenda algumas densidades de cobertura por pluviômetros mínimas para simulação hidrológica em determinadas regiões fisiográficas, por exemplo: a) Pequenas ilhas: 25 km<sup>2</sup> por estação; b) Regiões montanhosas: 250 km<sup>2</sup> por estação; c) Planícies interiores e montanhoso/ondulado: 575 km<sup>2</sup> por estação; d) Áreas costeiras: 900 km<sup>2</sup> por estação. Esses valores também foram utilizados em estudo feito por Cristaldo et al., (2017), que encontraram baixa densidade de estações para a bacia do rio Aquidauana, também considerada uma bacia de planalto adjacente ao Pantanal.

Considerando a área total de 15.190,99 km<sup>2</sup> da sub-bacia do rio Miranda, a densidade de pluviômetros dessa região encontra-se aquém do recomendado pela WMO. Dessa forma, visando uma abrangência completa da precipitação dentro da área da sub-bacia em estudo, foram utilizados dados extraídos de coleções de imagens de satélite que serão detalhados na seção 3.2.2.

### **3.2 Dados**

A coleta de dados para alimentação das redes neurais, detalhada na sequência, foi realizada separadamente tanto para as áreas de contribuição das porções à montante, quanto na sua área de contribuição total.

#### **3.2.1 Vazão**

As informações de vazão dos postos fluviométricos das bacias foram coletadas a partir das séries históricas do Portal Hidroweb do Sistema Nacional de Informações de Recursos Hídricos (SNIRH) da Agência Nacional de Águas (ANA), disponível através do endereço web: <https://www.snirh.gov.br/hidroweb/serieshistoricas>. Foram extraídos dados de vazão diária de 01 janeiro de 2000 a 31 de dezembro de 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e Miranda (cód. 66910000), de forma a obter dados à montante e à jusante da sub-bacia.

Os dados coletados apresentaram uma defasagem quanto os picos de vazões das estações fluviométricas de jusante e montante dentro das bacias, portanto foi realizado o alinhamento das séries seguindo a metodologia de cálculo da máxima correlação entre os dados, utilizada por Santos Antoniassi (2022) e pela Embrapa Pantanal. Essa metodologia consiste em posicionar em planilha eletrônica as bases de dados das duas estações em estudo e avaliar sua correlação. Em sequência, os valores da estação de montante são deslizados uma célula para baixo na planilha eletrônica, sendo calculada uma nova correlação entre os dados. Esse processo é repetido até que se encontre o maior valor de R<sup>2</sup>. O número de deslizamentos da série temporal indica o intervalo de dias que a água da estação de montante leva para chegar à estação de jusante.

A defasagem de tempo corresponde ao período decorrido entre a passagem da água da estação de montante à estação de jusante (SANTOS ANTONIASSI, 2022). Esse tempo varia de acordo com as características da fisiografia da bacia hidrográfica como rugosidade, inclinação, cobertura do solo, taxa de infiltração de água no solo e até a sinuosidade do rio principal. Pela metodologia de Santos Antoniassi (2022), para a BRM foi encontrado uma defasagem de 7 dias, visto que o rio Miranda apresenta uma sinuosidade acentuada. A distância em linha reta entre as estações fluviométricas é de 65,86 km e o comprimento do trecho do rio é de 176,43 km (Figura 5).

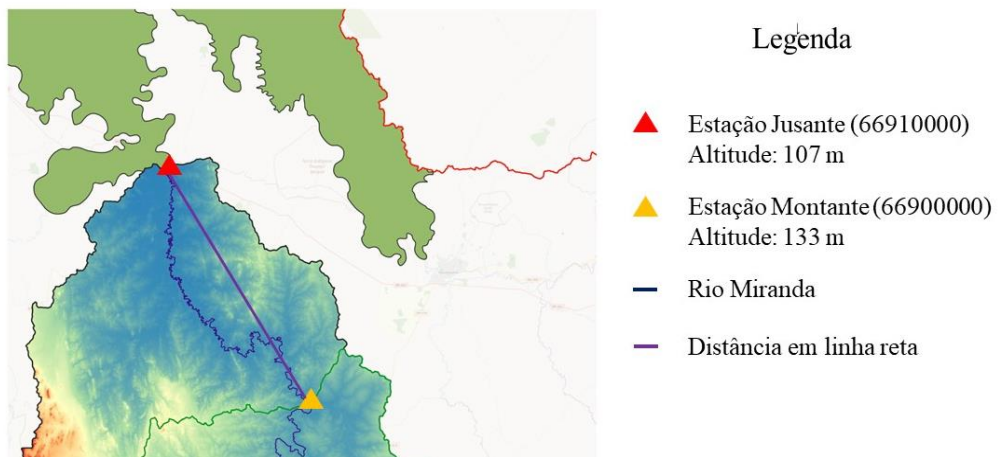


Figura 5 – Diferença de altitudes das estações fluviométricas.  
Fonte: próprio autor (2023).

Além do alto índice de sinuosidade presente no rio Miranda (Merino et al. 2013), a defasagem encontrada pode ser relacionada ao tempo de concentração da bacia. O tempo de concentração da bacia ( $t_c$ ) é definido por Mulvany (1851) como o tempo necessário para que a chuva que cai no ponto mais remoto da bacia chegue ao ponto de exsutório. Essa relação pode ser vista na equação (8) a partir do cálculo proposto por Kirpich (1940).

$$t_c = \frac{L}{\sqrt{\frac{\Delta h}{L}}} \quad (8)$$

Em que:

$T_c$  é o tempo de concentração, em horas;

$L$  é a distância do curso, em km;

$\Delta h$  é a diferença de cota entre as estações;

Para o trecho do rio em linha reta, o tempo de concentração calculado é de 104,82 horas, o que totaliza 4,36 dias. Considerando a sinuosidade natural do rio Miranda, o tempo de concentração calculado é de 459,70 horas, o que equivale a 19,15 dias. Essa diferença demonstra uma grande sinuosidade do rio Miranda, promovendo a defasagem entre as estações avaliadas.

### 3.2.2 Pluviometria

Os dados de pluviometria diários foram extraídos a partir de imagens de satélite da coleção CHIRPS versão 2.0 Final, disponibilizados pela Climate Hazards Center US Santa Barbara, a partir da plataforma Google Earth Engine. A coleção conta com uma resolução espacial de  $0.05^\circ$  e está disponível para utilização através do endereço web: <https://developers.google.com/earth->

engine/datasets/catalog/UCSB-CHG\_CHIRPS\_DAILY. A coleção CHIRPS é uma fonte de dados de precipitação global que combina observações de satélite com dados climáticos em Terra para produzir estimativas de precipitação. Foi inserido um arquivo no formato *shapefile* delimitado através do *software* QGIS v3.28.0 Firenze, contendo a área da bacia hidrográfica em estudo e foram extraídas as imagens diárias referentes à precipitação do período estudado para sua delimitação. Cada pixel das imagens extraídas corresponde a um valor numérico que representa a quantidade de precipitação diária em uma determinada área geográfica. A área representada por cada pixel corresponde à resolução da coleção (Tabela 3).

Para trabalhar com a precipitação na escala diária, foram extraídos os valores numéricos de precipitação dos pixels de cada imagem gerada e seus valores foram somados para a área da bacia delimitada, obtendo-se a precipitação diária acumulada. Deve-se trabalhar com o somatório de precipitação pois esse valor representa o total de entrada de água dentro da área da bacia proveniente das chuvas para um dia. Todo o procedimento descrito foi realizado para a porção à montante da bacia, assim como para a porção total (ou jusante).

### **3.2.3 Evapotranspiração**

A Evapotranspiração Real diária média foi extraída e calculada a partir de imagens de satélite da coleção NASA Global Land Data Assimilation System (GLDAS) versão 2, em resolução espacial de 0.25°, utilizando a metodologia do tópico 3.2.2 para os dados de pluviometria. A coleção encontra-se disponível através do endereço: <https://ldas.gsfc.nasa.gov/gldas>. Os dados foram coletados a partir do Google Earth Engine, similarmente aos dados de precipitação, descritos no tópico 3.2.2. Para os dados de evapotranspiração, foi calculada a média diária de evapotranspiração, de forma a representar a estimativa média dessa variável para a região de interesse. A evapotranspiração corresponde ao processo pelo qual a água é devolvida para a atmosfera pela evapotranspiração das plantas e evaporação de água do solo. Portanto, ela representa a saída de água do sistema hidrológico (CARVALHO et al. 2012).

A Tabela 3 apresenta um resumo das variáveis utilizadas.

Tabela 3 – Resumo das variáveis utilizadas no estudo.

Variáveis	Fonte	Resolução	Escala	Período	Unidade
Precipitação	Coleção CHIRPS v. 2.0	0.05°	Diária	01/01/2000 - 31/12/2020	mm.d <sup>-1</sup>
Evapotranspiração	Coleção GLDAS v. 2.0	0.25°	Diária	01/01/2000 - 31/12/2020	mm.d <sup>-1</sup>
Vazão	Hidroweb (SNIRH)	n.a*	Diária	01/01/2000 - 31/12/2020	m <sup>3</sup> .s <sup>-1</sup>

\* não aplicável Fonte: próprio autor (2023).

Foi montada uma planilha eletrônica Excel, que contou com as variáveis coletadas, onde cada coluna representou uma variável e cada linha representou um dia, na escala temporal. Dessa forma, para a primeira linha da planilha, estavam dispostos os dados de Vazão, Evapotranspiração e Precipitação (para jusante e montante nas três variáveis, totalizando 6 colunas de dados) para o dia 01 de janeiro de 2000.

Se observado algum dado faltante ou NaN (*Not a Number*) de alguma célula de determinada variável para determinada data, toda a linha referente a essa data foi excluída a fim de deixar a planilha sem nenhuma falha. Após esse tratamento e sincronização de todas as variáveis analisadas em planilha eletrônica, a base de dados somou um total de 6118 para a BRM.

### 3.3 Pré-processamento dos dados

Para a análise dos dados diários coletados, foi verificado primeiramente se houve dados faltantes na leitura dos produtos de satélite ou no banco de dados da ANA para o período de interesse. Os dados coletados foram visualizados em formato de planilha eletrônica excel, para todas as variáveis. Os dados dos postos fluviométricos da Hidroweb (SNIRH) apresentaram falhas nas leituras de vazões, para os quatro postos fluviométricos estudados, enquanto nos dados provenientes dos produtos de satélite não foram identificados períodos de falhas.

Bertoni e Tucci (2007) analisaram as metodologias de regressão linear e ponderação regional para realizar o preenchimento de falhas de vazões e concluíram que essas metodologias não são adequadas para preencher as escalas diárias e horárias. Dessa forma, visando dar ênfase na análise das variáveis e em testar as arquiteturas das redes neurais, optou-se por, ao invés de realizar o preenchimento de falhas nos dados, excluir as linhas diárias de todas as variáveis, visando a sincronização de todas as informações. Todo o pré-processamento dos dados foi realizado utilizando a linguagem Python®, através do *software* de código aberto, utilizando a plataforma Jupyter Notebook®.

Posteriormente, os dados processados foram subdivididos em subconjuntos de treinamento, validação e teste. Essa divisão representou uma divisão de 75% para os dados de treinamento e

validação e 25% destinados a teste. São encontradas na literatura diferentes tamanhos dos conjuntos de subdivisão dos dados para treinamento de uma rede neural MLP em estudos hidrológicos. São encontradas as divisões entre 70% para treinamento e validação e 30% para teste (KADAM et al., 2019), 80% para treino e validação e 20% para teste (SNEIDER, ABOGADIL, KHAN, 2021) ou 85% para treino e validação e 15 % para teste (ABDALRAHMAN et al., 2022).

A escolha dos subconjuntos utilizados no estudo foi feita respeitando esse intervalo de divisão. Durante a etapa de treinamento e validação, foram utilizados os dados disponíveis de 01 de janeiro de 2000 a 31 de dezembro de 2014. Os dados de teste utilizados foram de 01 de janeiro de 2015 a 31 de dezembro de 2020. Foi utilizada uma planilha eletrônica independente com os dados de teste para a facilitação dos códigos.

### **3.4 Construção das redes neurais artificiais**

As redes neurais propostas foram alimentadas de forma a analisar a sub-bacia do rio Miranda. O número de neurônios das camadas de entrada e de saída da rede neural dependem do modelo que se deseja obter de acordo com o problema estudado. Os neurônios da camada escondida são parâmetros que devem ser determinados, visando a melhor performance do modelo proposto. As redes neurais artificiais do tipo *Multilayer Perceptron* (MLP) foram usadas no estudo e testadas diferentes combinações de arquiteturas, até serem encontradas os melhores coeficientes.

Para isso, foram definidos os dados das camadas de entrada e saída e foi alterado o número de camadas intermediárias e seus neurônios, através de tentativa e erro. O dado de saída (neurônio da camada de saída) foi sempre o mesmo para todas as arquiteturas testadas. Utilizou-se a técnica de k-fold cross validation para séries temporais, seguindo a metodologia de Jung et al. (2020) e Rohani, Taki e Abdollahpour (2018) e a divisão dos dados foi fixada em 5 folds para todos os modelos testados, visando aumentar a rapidez de treinamento dos modelos. A técnica de K-fold cross validation é utilizada para avaliar o desempenho dos modelos de aprendizado de máquina. Ela subdivide a base de dados em diversos subconjuntos chamados de k-folds, de tamanho igual ou aproximado, no qual o treinamento do modelo é utilizado um k-fold de dados e para teste utilizam-se os subconjuntos restantes (JUNG, 2017). Na sua utilização padrão, a escolha dos k-folds usados para treinamento e teste é realizada de forma aleatória pelo algoritmo. Entretanto, quando se trata de séries temporais, é necessário que os k-folds respeitem a temporalidade dos dados, mantendo a base de dados em ordem cronológica. Dessa forma, os k-folds utilizados para o treinamento encontram-se sempre no tempo passado em relação aos k-folds de teste.

Para otimizar o processo de treinamento das redes neurais foi utilizado o otimizador Adam (*Adaptive Moment Estimation*). O objetivo desse algoritmo é obter os parâmetros ótimos da rede neural MLP (Salem et al., 2022). Foi utilizada a função de ativação Rectified Linear Unit (ReLU) nas camadas intermediárias e a função de ativação linear na camada de saída, pois o problema trata de prever valores contínuos de vazão. Os pesos das conexões sinápticas entre camadas da rede foram

escolhidos randomicamente de forma a excluir tendências na etapa de aprendizado dos modelos. A arquitetura geral das redes neurais testadas pode ser vista na Figura 6.

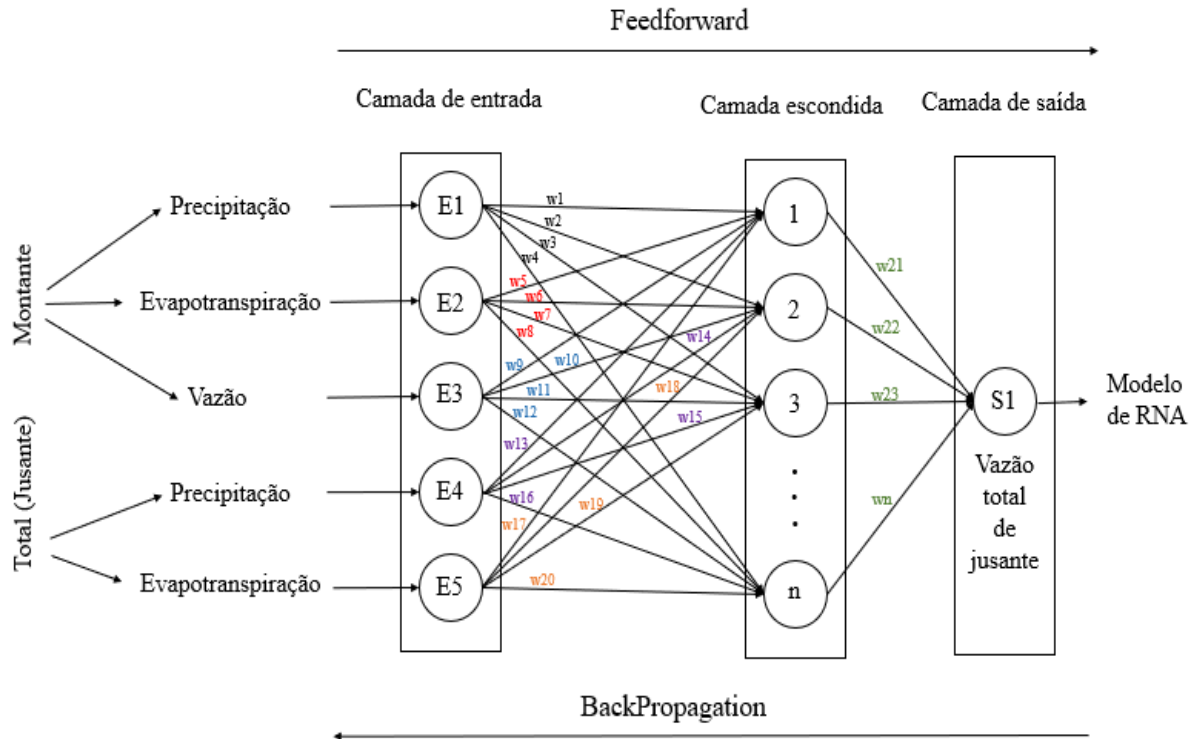


Figura 6 – Arquitetura geral das redes neurais artificiais testadas.  
Fonte: próprio autor (2023).

### 3.5 Análise de performance dos modelos

Inicialmente foi feita uma análise exploratória dos dados coletados. Durante a condução da pesquisa, foi analisada a correlação entre os dados para entender as taxas de acerto das arquiteturas de rede neural testadas. Para minimizar os efeitos da variabilidade presente nos dados, foi aplicado a metodologia de Médias Móveis em uma janela de 10 e 15 dias, assim como a transformação dos dados em escala diária para escala mensal. O método das médias móveis é utilizado para a suavização de séries temporais de vazão (YILMAZ; ONOZ, 2020) e precipitação (WU; CHAU, 2013) visando a redução de ruídos nos dados observados durante o pré-processamento. Adicionalmente, foi feita uma redução das variáveis nos modelos e avaliação da performance.

Visto que a sub-bacia do rio Miranda foi dividida em porção montante e total (seção 3.1), a retirada de variáveis dos modelos foi feita visando compreender se os dados da porção montante apresentavam redundância ou sobreposição de valores nos dados de entrada da porção total (jusante) e influenciavam negativamente na previsão dos modelos. O dado de vazão montante ( $Q$  montante), referente a estação de Estrada MT-738 (cód. 66900000) permaneceu como entrada desses modelos por



apresentar uma alta correlação com os dados de vazão total (Q total) da estação de Miranda (cód. 66910000), como será discutido na Seção 4.1 deste trabalho.

Dessa forma, as análises a seguir seguirão dispostas em blocos:

- a) Análise exploratória dos dados
- b) Teste de redes neurais com dados em escala diária
- c) Teste de redes neurais com dados provenientes da suavização por Médias Móveis (10 e 15 dias)
- d) Teste de redes neurais com dados em escala mensal
- e) Teste de redes neurais com retirada de variáveis

Para avaliação do desempenho dos modelos, foram usadas as métricas do Erro médio absoluto (MAE), Raiz quadrada do erro médio (RMSE) e coeficiente de determinação ( $R^2$ ), como discutido na seção 2.7.

## 4. RESULTADOS E DISCUSSÃO

### 4.1 Análise exploratória dos dados

A matriz de correlação apresenta as correlações entre variáveis de um conjunto de dados, em uma amplitude de -1,0 a 1,0. Na matriz obtida (Figura 7), é possível observar o grau de associação das variáveis de precipitação e evapotranspiração das porções montante e jusante da sub-bacia e vazão das estações Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000). A correlação entre a evapotranspiração de montante e jusante apresentou um coeficiente  $R^2$  de 1,0, sendo essa uma alta correlação. O mesmo foi observado para os dados de precipitação, no qual a correlação entre os dois conjuntos de dados apresentou coeficiente de  $R^2 = 0,99$ . O grau de correlação entre as estações fluviométricas também apresentou um alto coeficiente de  $R^2 = 0,78$ .

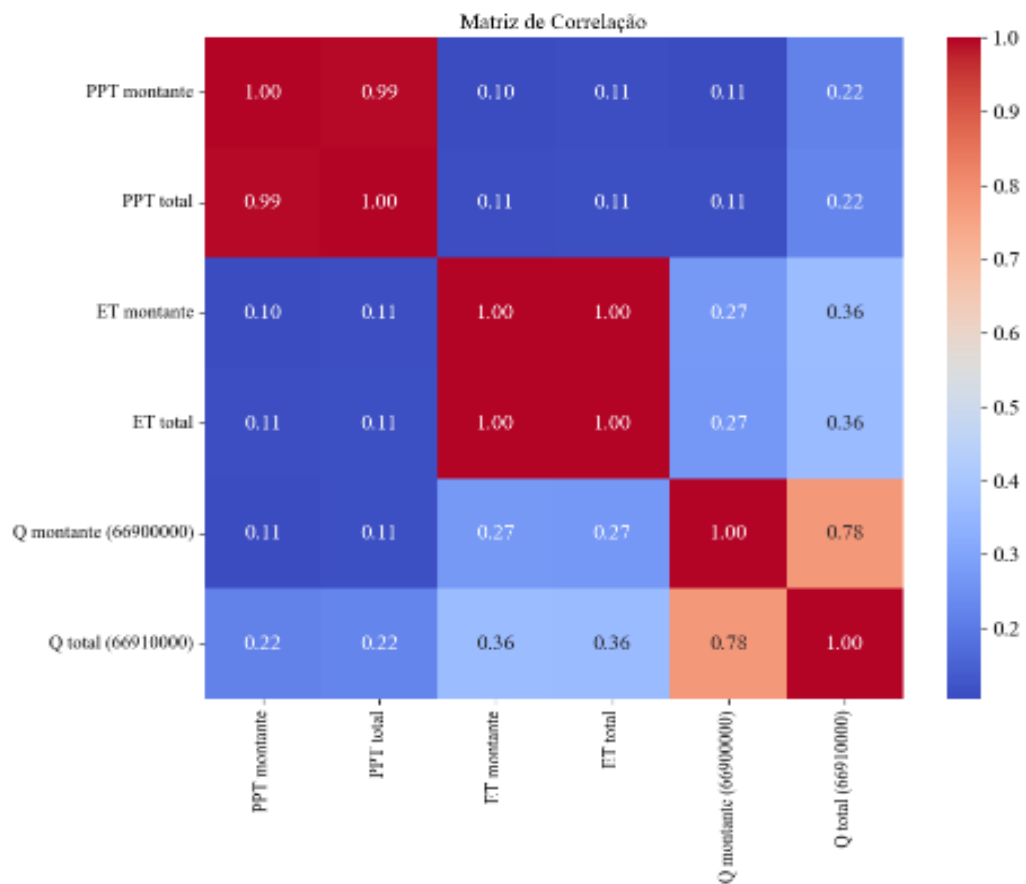


Figura 7 – Matriz de correlação contendo dados de precipitação (PPTP, evapotranspiração (ET) e vazão (Q) das estações de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) para as porções montante e jusante da sub-bacia do rio Miranda.

Fonte: próprio autor (2023).

### 4.1.1 Precipitação

Pode-se observar a precipitação diária das porções jusante (PPT Total) e montante (PPT montante) da sub-bacia do rio Miranda (Figura 8). Os valores diários máximos aproximaram-se dos 80 mm, como em períodos entre o ano de 2003 e 2004 e alguns pontos observados entre 60 e 70 mm, como visto nos anos de 2013 a 2016. Visando compreender a distribuição do conjunto de dados, foi plotado um gráfico Box Plot com os valores das médias diárias dos meses de janeiro a dezembro para os anos de 2000 a 2020 (Figura 9).

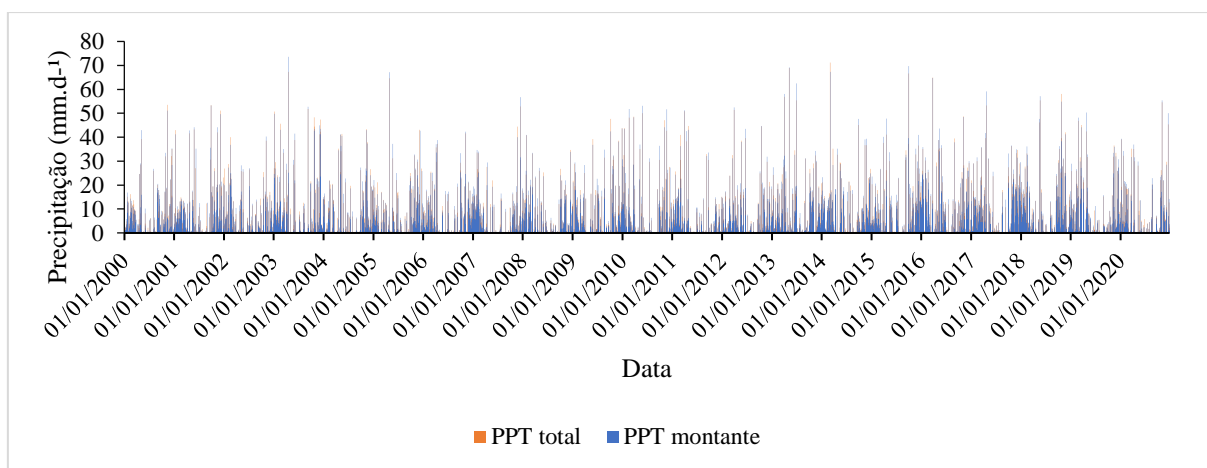


Figura 8 – Precipitação diária de 2000 a 2020 das porções à montante e à jusante da sub-bacia do rio Miranda. Fonte: próprio autor (2023).

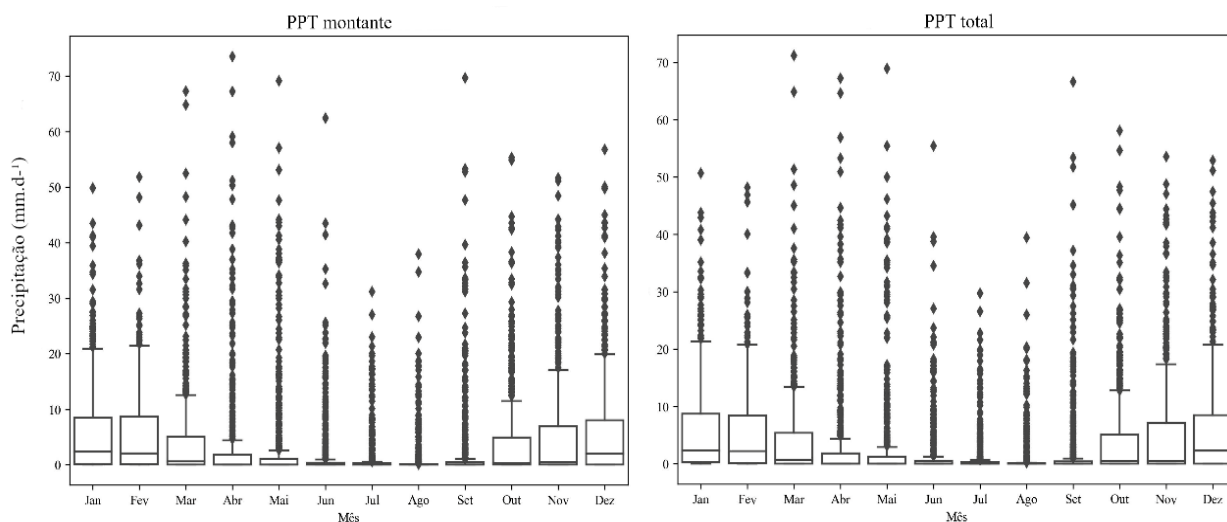


Figura 9 – Box plot dos valores diários médios de precipitação para os meses de janeiro a dezembro dos anos de 2000 a 2020 da porção montante e jusante da sub-bacia do rio Miranda. Fonte: próprio autor (2023).

Nota-se na Figura 9, um comportamento similar da precipitação para as duas porções da sub-bacia. Os meses de abril a setembro apresentam os menores valores de chuva, ficando abaixo dos 5 mm diários. Durante os meses de outubro a março, existe uma maior concentração da precipitação, onde os

valores de chuva aproximam-se de 10 mm diários. Guimarães et al. (2018), ao estudarem a precipitação média mensal para municípios do Pantanal Matogrossense, encontraram para a cidade de Miranda o mesmo comportamento, onde a maior incidência de precipitação concentrou-se entre os meses de outubro a março. Os autores usaram a mesma coleção CHIRPS para a espacialização das chuvas do bioma, entretanto, os valores mensais de chuvas apresentaram amplitude de 26 a 202 mm.d<sup>-1</sup>. Isso pode ser explicado pela maior base de dados utilizada pelos autores, do ano de 1981 a 2018. É possível observar vários pontos *outliers* para os meses estudados, os quais se distanciam da maioria dos dados considerados pelos quartis do Box Plot. Tanto para a porção montante quanto jusante, são apresentados valores acima de 70 mm para os meses de março, abril, maio e setembro, ficando bastante acima da média observada.

#### 4.1.2 Evapotranspiração

Os dados de evapotranspiração apresentaram ciclos bem definidos no decorrer dos anos (Figura 10). Os dados provenientes da coleção GLDAS v. 2.0 mostraram que a evapotranspiração se aproxima dos 7,0 mm nos períodos iniciais do ano de janeiro a maio, tendo uma redução para valores próximos a 1,0 mm nos meses de junho a setembro. Como visto na matriz de correlação (Figura 7), o coeficiente de correlação da evapotranspiração das porções montante e jusante apresentaram valor de  $R^2 = 1,00$ .

A partir do gráfico de Box Plot (Figura 11) observa-se a distribuição dos dados no decorrer dos meses. A observação do comportamento similar da evapotranspiração nas porções analisadas está de acordo com o alto valor de correlação calculada dos dados.

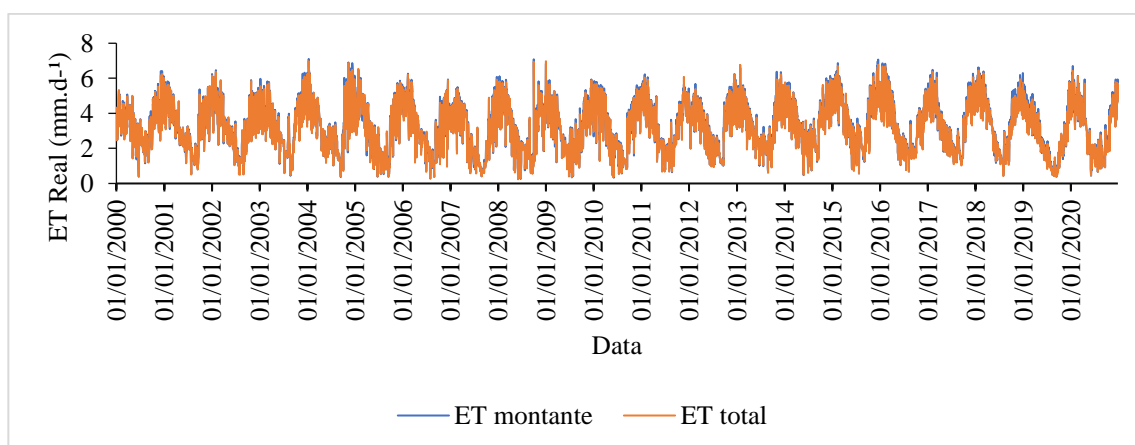


Figura 10 – Evapotranspiração real diária de 2000 a 2020 das porções à montante e à jusante da sub-bacia do rio Miranda.

Fonte: próprio autor (2023).

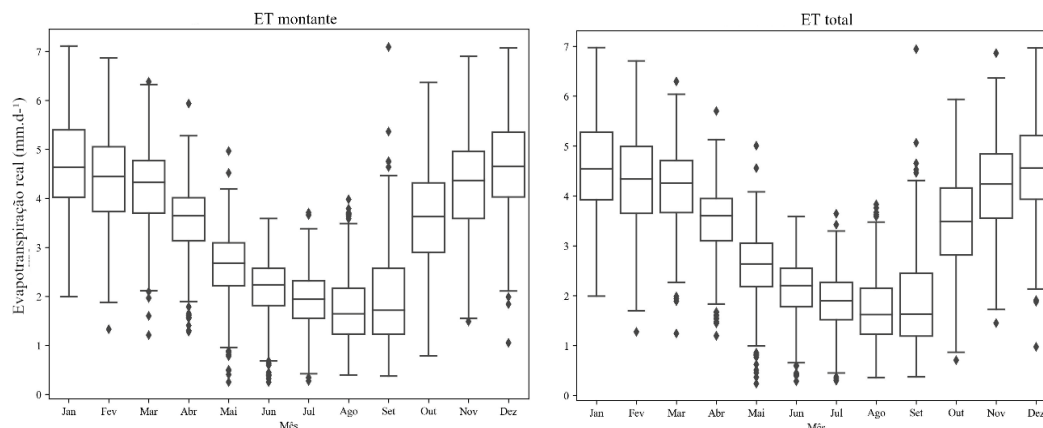


Figura 11 – Box plot dos valores diários médios de evapotranspiração para os meses de janeiro a dezembro dos anos de 2000 a 2020 da porção montante e jusante da sub-bacia do rio Miranda.  
Fonte: próprio autor (2023).

### 4.1.3 Vazão

Em relação aos dados de vazão (Figura 12), as estações fluviométricas apresentaram uma alta correlação, com coeficiente  $R^2 = 0,78$ . Esse valor foi encontrado após utilizar a metodologia de sincronização das estações, excluindo sua defasagem no tempo, aplicada por Santos Antoniassi (2022), no qual foi encontrada uma defasagem de sete dias entre as estações de montante e jusante. Essa alta correlação pode estar relacionada com a relação de causa e efeito da estação à montante em relação a jusante. Uma correlação com valor maior poderia ser obtida caso outras formas de entrada e saída de água da bacia fossem quantificadas como taxa de infiltração, contribuição de outros afluentes do rio principal, uso e cobertura da terra, escoamento subterrâneo, capacidade de infiltração de água no solo (TUCCI, CLARKE, 1997; SOUZA, SOUZA, CARDOSO, 2017).

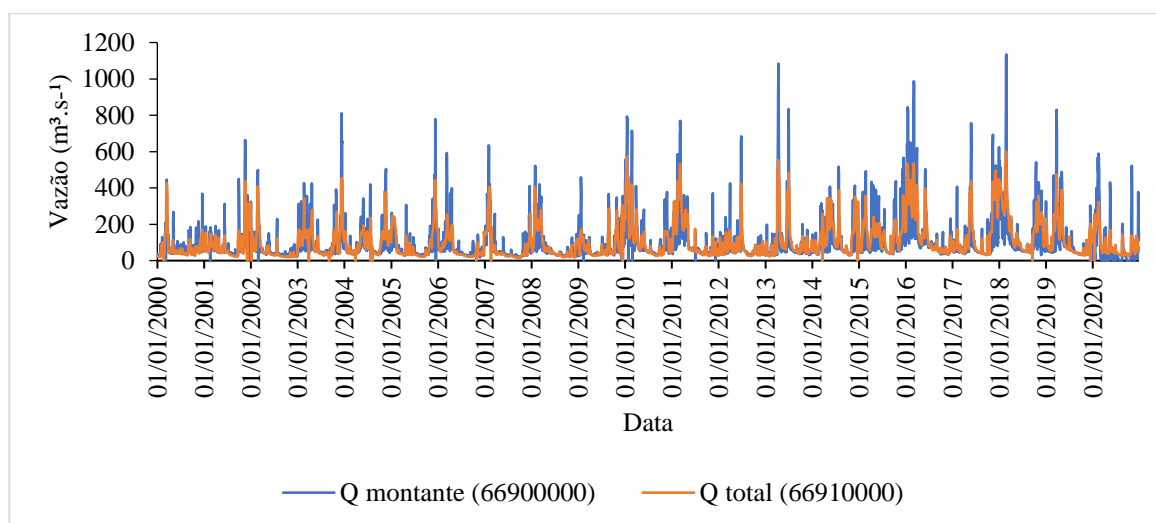


Figura 12 – Hidrograma de vazões diárias de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) da sub-bacia do rio Miranda.  
Fonte: próprio autor (2023).

De acordo com a disposição dos dados, é possível observar ciclos com máximas e mínimas de vazão, para ambas as estações. Nos meses centrais do ano, de julho a setembro, observa-se um decréscimo das vazões observadas, enquanto nos meses iniciais (janeiro a abril) e finais (outubro a dezembro) do ano, os maiores valores de vazão predominam. A estação fluviométrica de Estrada MT-738 (cód. 66900000), de forma geral, apresenta um comportamento com maiores valores de vazão em relação a estação de Miranda (cód. 66910000). Visto que a área de contribuição da bacia estudada apresenta uma grande extensão, assim como o comprimento do seu rio principal, pode-se inferir que parte da vazão observada na estação à montante é perdida através da evapotranspiração e possivelmente por fatores como infiltração de água no solo, uso e cobertura da terra, entre outros fatores não medidos neste estudo.

Nos últimos cinco anos analisados, de 2015 a 2020, observa-se uma tendência de aumento das vazões para ambas as estações fluviométricas, apresentando picos acima de  $400 \text{ m}^3 \cdot \text{s}^{-1}$  para estação do Miranda e acima de  $600 \text{ m}^3 \cdot \text{s}^{-1}$  para a estação de Estrada MT-738. O gráfico Box Plot (Figura 13) mostra o comportamento dos valores médios diários para os meses de janeiro a dezembro da vazão das estações de montante de jusante da sub-bacia.

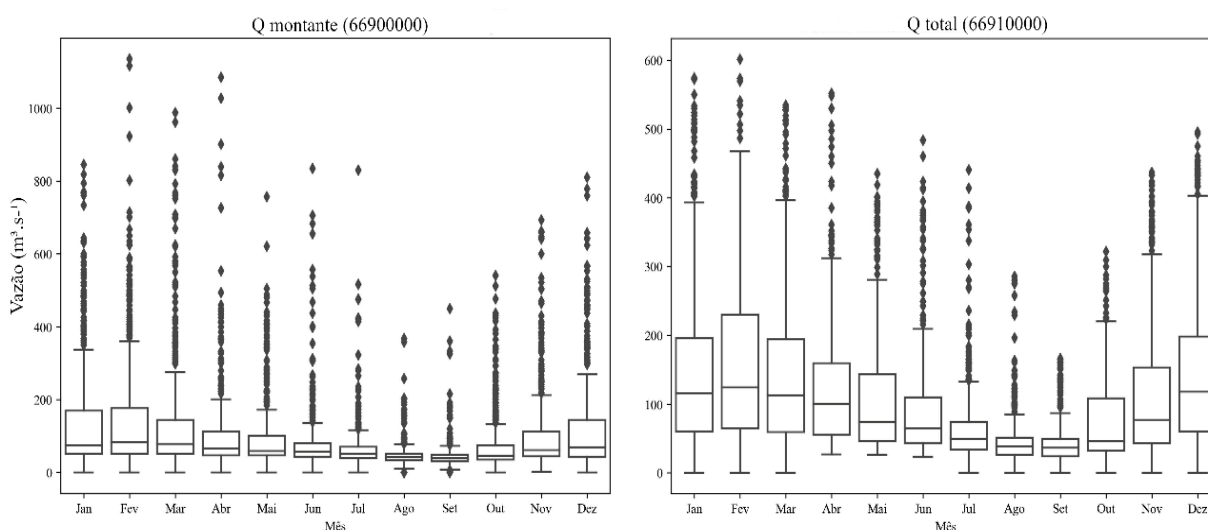


Figura 13 – Box plot dos valores diários médios de vazão para os meses de janeiro a dezembro dos anos de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e Miranda (cód. 66910000), da sub-bacia do rio Miranda.

Fonte: próprio autor (2023).

De acordo com as figuras, as medianas dos conjuntos de dados encontram-se próximas a  $100 \text{ m}^3 \cdot \text{s}^{-1}$  nos meses iniciais e finais do ano, apresentando um decréscimo durante os meses de maio a setembro, ficando próximas a  $50 \text{ m}^3 \cdot \text{s}^{-1}$ . As medianas, assim como os valores máximos de vazão são maiores para a estação de Miranda em comparação com a estação de Estrada MT-738. Em relação aos pontos outlier, a estação de Estrada MT-738 (cód. 66900000) apresenta valores superiores a estação de

Miranda (cód. 66910000), chegando a atingir vazões acima dos 1000 m<sup>3</sup>.s<sup>-1</sup>. Na estação de Miranda, os pontos *outlier* mais altos mostram valores de vazão próximos aos 600 m<sup>3</sup>.s<sup>-1</sup>.

#### 4.2 Redes neurais com dados em escala diária

Foram testadas seis arquiteturas variando o número de camadas ocultas e seus neurônios (Tabela 4). Os dados de entrada de todas as arquiteturas para os modelos propostos foram a vazão de montante (Q montante), precipitação total (PPT Total), precipitação montante (PPT montante), evapotranspiração total (ET Total) e evapotranspiração montante (ET montante). Como dado de saída para os modelos, foi definida a vazão total (Q Total), referente a vazão do posto fluviométrico do município de Miranda (cód. 66910000).

Tabela 4 – Resultado da análise das métricas dos modelos testados com dados diários durante o treinamento, validação e teste.

Arquitetura	Dados de entrada	Dado de saída	Neurônios nas camadas ocultas*	Epochs	Treinamento e validação			Teste		
					RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
1	Q		4,8	300	55,4444	35,6353	0,6282	85,8385	58,4901	0,4258
2	montante, PPT		3,5	500	64,8135	43,9596	0,4533	86,9597	58,2048	0,4107
3	total, PPT	Q total	3,6	700	55,7929	35,4686	0,6259	92,1991	59,3372	0,3376
4	montante, ET		3,7	900	55,3331	36,0964	0,627	88,1867	58,6609	0,394
5	total, ET		4,5,4	1000	56,6885	36,7229	0,609	87,1867	57,5353	0,4015
6	montante		3,5,2	1000	72,1956	49,1515	0,3205	95,788	60,4452	0,285

\*dois números correspondem a duas camadas ocultas, três números correspondem a três camadas.

Fonte: próprio autor (2023).

Nas primeiras avaliações foi realizado um aumento progressivo das épocas de treinamento de cada modelo testado. Pelos dados apresentados na Tabela 4, é possível observar que o aumento do número de épocas e camadas intermediárias não resultou em melhores métricas de avaliação. O mesmo ocorreu com Apaydin et al. (2020) ao testarem diferentes tipos de redes neurais MLP para a previsão de vazões. Os autores encontraram que não houve diferenças significativas nas performances dos modelos testados ao aumentar de 300 para 500 épocas de treinamento. Entretanto, os autores encontraram que o tipo de rede LSTM (Long Short-Term Memory), com o número de 500 épocas resultou em valores de R<sup>2</sup> = 0,87, RMSE = 17,92 e MAE = 9,98 após a etapa de validação dos modelos.

O número de camadas ocultas variou de duas a três para os modelos propostos. Oliveira, Ramos e Neves (2023) testaram de zero a quatro camadas ocultas em seus modelos para estimar a vazão diária em bacia hidrográfica na região sudoeste de Portugal. Para os modelos propostos neste estudo, o número de duas camadas intermediárias apresentou melhores valores de Coeficiente de Determinação ( $R^2$ ).

As arquiteturas 1 e 4 apresentaram os melhores coeficientes de determinação após a etapa de treinamento e validação, obtendo-se 0,6282 e 0,627, respectivamente. Entretanto, pode-se perceber um decréscimo das métricas em todas as arquiteturas após a etapa de teste com os últimos cinco anos de dados coletados. Após a etapa de teste, a arquitetura 1, contendo 4 e 8 neurônios nas camadas intermediárias e 300 épocas de treinamento apresentou os melhores coeficientes de RMSE (85,8385), MAE (58,4901) e  $R^2$  (0,4258), sendo o modelo mais eficiente em previsão dos testes iniciais. Em contraste com esses valores, Kassem, Raheem e Khidir (2020) obtiveram valores de coeficiente de determinação  $R^2 = 0,83$  e  $0,85$ , utilizando um número de 6 e 8 neurônios na camada intermediária das suas redes MLP para a previsão de vazão diária em bacia de área de 727 km<sup>2</sup> na região do Iraque.

Visto que o aumento de épocas de treinamento não foi um fator exclusivo para melhoria das taxas de acerto, optou-se por fixar as épocas para os modelos nas análises sequenciais para 300 interações, valor também usado por Apaydin et al., (2020). Também foram alterados os números de camadas intermediárias e neurônios, por tentativa e erro. A Tabela 5 apresenta os resultados com as novas configurações de arquiteturas. Os dados de entrada e de saída permaneceram inalterados.

Tabela 5 – Tabela resumo das métricas dos modelos testados com dados diários durante o treinamento, validação

Arquitetura	Dados de entrada	Dado de saída	Neurônios nas camadas ocultas*	Treinamento e validação			Teste		
				RMSE	MAE	$R^2$	RMSE	MAE	$R^2$
1	Q (montante), PPT (total), PPT (montante), ET (total), ET (montante)	Q (total)	2	56,8217	35,2663	0,6112	89,5942	58,5914	0,3745
7			55,4564	35,8326	0,628	112,41	66,6568	0,0153	
4,8			56,0505	35,6383	0,6209	104,23	63,7176	0,1534	
4			6,2	78,0322	51,5844	0,226	178,428	137,855	0
5			2,4,5	55,6268	36,3631	0,6265	88,6879	58,6414	0,3871
6			6,3,2	75,7398	51,1219	0,2358	178,445	137,877	0

e teste.

\*um número corresponde a uma camada oculta, dois números correspondem a duas camadas e assim por diante.

Fonte: próprio autor (2023).



A arquitetura 5 apresentou-se como a melhor configuração em comparação as demais de acordo com suas métricas. O aumento progressivo de 2 a 5 neurônios em suas três camadas intermediárias proporcionou um  $R^2$  de 0,3871, sendo o maior dos modelos propostos com o mesmo número de interações. Entretanto, os valores previstos pela arquitetura 5 (Figura 14) superestimam os dados de valores reais.

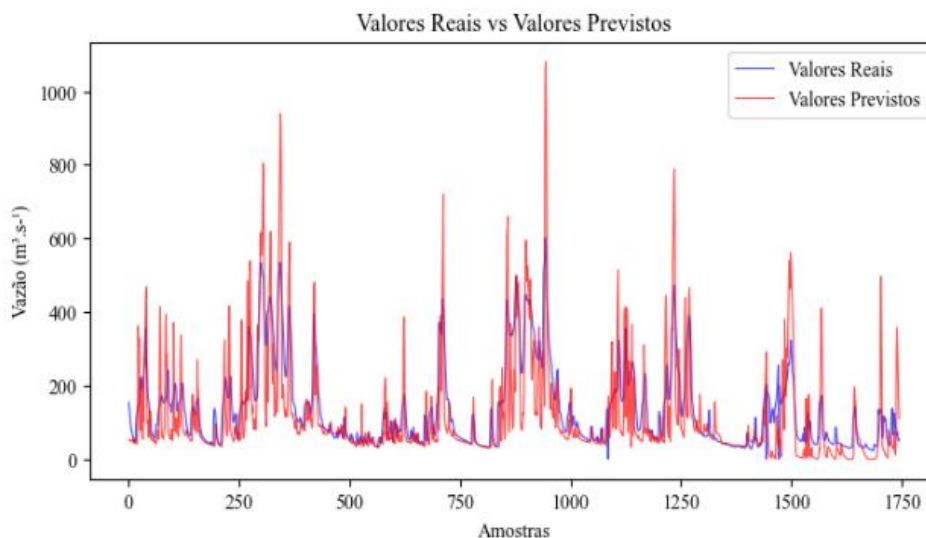


Figura 14 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 5 a partir dos dados diários. Fonte: próprio autor (2023).

É possível observar que em momentos de picos de vazão, as previsões do modelo geram valores bastante elevados, apresentando uma diferença de até aproximadamente  $400 \text{ m}^3.\text{s}^{-1}$  como no pico próximo a milésima amostra da validação. Para outros momentos de picos, como no início e no final da validação, essa diferença é diminuída, entretanto, o modelo ainda superestima os valores reais em torno de  $200 \text{ m}^3.\text{s}^{-1}$ . O inverso ocorreu no estudo feito por Apayadin et al. (2020), em que os valores preditos pelos modelos de RNA-MLP subestimaram os dados observados de vazão, apresentando uma diferença de até aproximadamente  $60 \text{ m}^3.\text{s}^{-1}$  em momentos de pico de vazão.

A redução progressiva no número de neurônios das camadas ocultas, como proposto na Arquitetura 6 não gerou previsões pelo modelo (Figura 15). O número de seis neurônios na primeira camada oculta mostrou-se como uma opção não viável para previsões com os dados em estudo durante a validação. Resultados similares foram gerados pela Arquitetura 4, a qual também contou com seis neurônios em sua primeira camada intermediária.

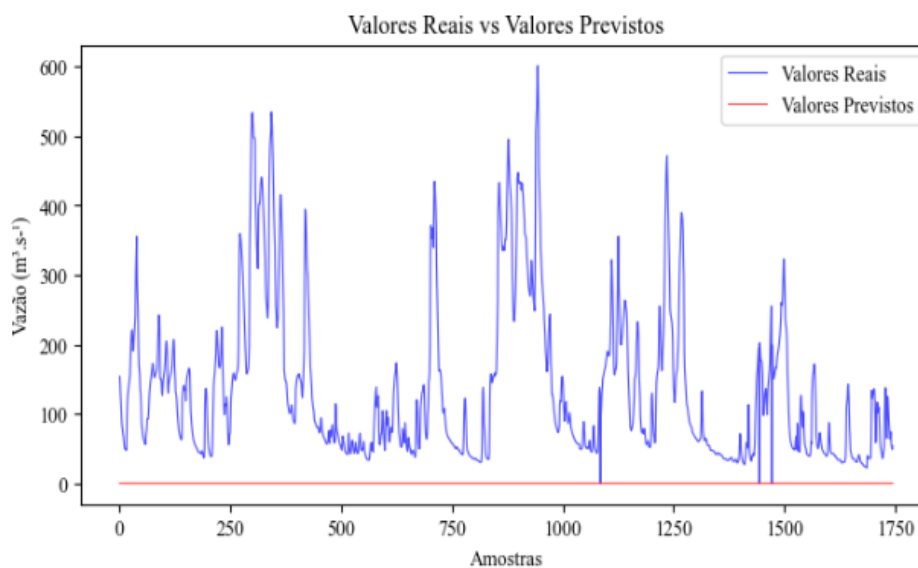


Figura 15 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 6 a partir dos dados diários. Fonte: próprio autor (2023).

#### 4.3 Redes neurais com dados suavizados por Médias Móveis

As redes neurais alimentadas por dados suavizados a partir de médias móveis obtiveram indicadores de performance melhores do que as redes que utilizaram dados diários sem suavização. Entretanto, ainda se observa que as métricas estatísticas apresentam redução durante a etapa de teste, mesmo após de passar pelo treinamento. As arquiteturas presentes na Tabela 6, apresentam os coeficientes de determinação  $R^2$  acima de 0,62 após a etapa de Teste, com exceção da Arquitetura 2 ( $R^2 = 0,2084$ ) e Arquitetura 6 ( $R^2 = 0$ ). A melhor performance apresentada para os dados suavizados em escala decendial (10 dias) foi vista na Arquitetura 1, que embora contasse com menos neurônios na camada oculta em comparação as demais, apresentou menores valores de RMSE e MAE e o maior valor de  $R^2$ .

Tabela 6 – Tabela resumo das métricas dos modelos testados com dados suavizados em escala decencial durante o treinamento, validação e teste.

Arquitetura	Dados de entrada	Dado de saída	Neurônios nas camadas ocultas*	Treinamento e validação			Teste		
				RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
1	Q		2	44,6342	31,7170	0,6414	54,8524	37,3136	0,7641
2	(montante),		7	34,2845	22,7413	0,8266	100,4891	57,3762	0,2084
3	PPT (total),		4,8	36,5297	23,9142	0,7923	69,2843	41,4977	0,6237
4	PPT (montante),	Q (total)	6,2	41,3419	28,2370	0,6673	55,8252	37,6254	0,7557
5	ET (total), ET (montante)		2,4,5	33,9854	22,0341	0,8253	57,2073	36,6804	0,7434
6			6,3,2	43,7955	28,2683	0,6488	179,8763	139,9971	0

\*um número corresponde a uma camada oculta, dois números correspondem a duas camadas e assim por diante.  
Fonte: próprio autor (2023).

Após a suavização dos dados em escala de janela quinzenal, o comportamento dos modelos foi superior àqueles que utilizaram dados sem suavização e suavizados em janela decencial. Nota-se, pelos dados das Tabelas 6 e 7 que quanto maior a janela de suavização dos dados, melhor a performance dos modelos durante o treinamento e validação, possivelmente pela redução da variação dos dados de vazão observados nos postos fluviométricos. As reduções da amplitude dos dados de vazão obtidas pela suavização dos dados podem ser vistas nas Figuras 16 e 17, para as janelas de dez e quinze dias, respectivamente.

Tabela 7 – Tabela resumo das métricas dos modelos testados com dados suavizados em escala quinzenal durante o treinamento, validação e teste.

Arquitetura	Dados de entrada	Dado de saída	Neurônios nas camadas ocultas*	Treinamento e validação			Teste		
				RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
1	Q		2	49,8487	36,7634	0,5034	60,2567	41,7287	0,6978
2	(montante),		7	31,4292	21,2419	0,8192	50,9422	33,6652	0,7840
3	PPT (total),		4,8	41,3674	27,7301	0,6720	44,8524	33,9766	0,8173
4	PPT (montante),	Q (total)	6,2	36,3751	25,8634	0,6866	53,8488	40,1482	0,7587
5	ET (total), ET (montante)		2,4,5	58,8381	44,2277	0,3274	46,5103	33,4190	0,8200
6			6,3,2	41,8801	27,3504	0,6633	177,2010	139,2319	0

\*um número corresponde a uma camada oculta, dois números correspondem a duas camadas e assim por diante.  
Fonte: próprio autor (2023).

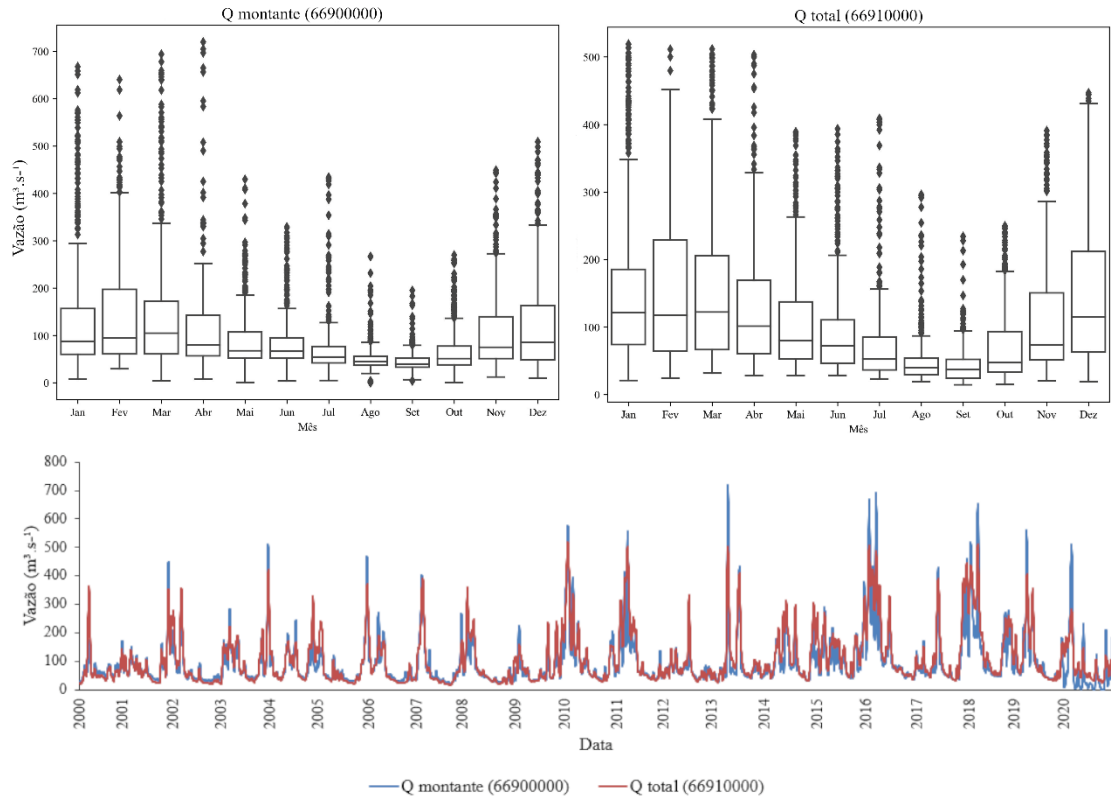


Figura 16 – Box Plot e Hidrograma de vazões diária de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) da sub-bacia do rio Miranda suavizados por médias móveis em escala decenal.  
 Fonte: próprio autor (2023).

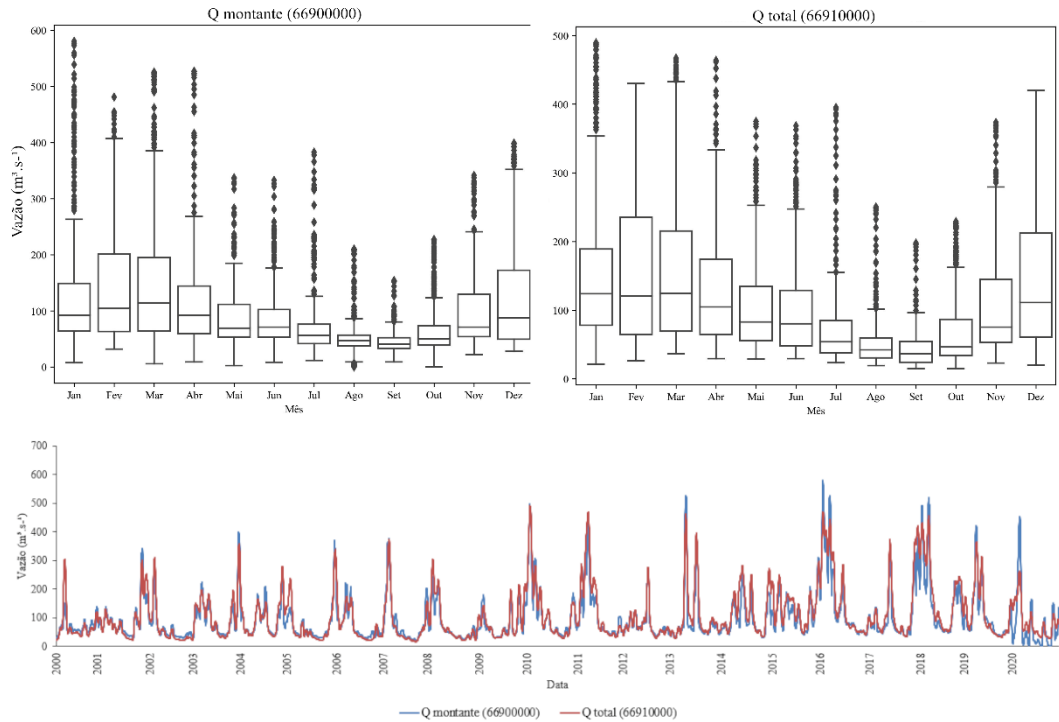


Figura 17 – Box Plot de vazões diárias médias e hidrograma de vazões diárias de 2000 a 2020 das estações fluviométricas de Estrada MT-738 (cód. 66900000) e de Miranda cód. (66910000) da sub-bacia do rio Miranda suavizados por médias móveis em escala quinzenal.

Fonte: próprio autor (2023).

O número de três camadas intermediárias com o aumento progressivo de neurônios utilizadas pela Arquitetura 5 apresentou a melhor performance para a escala decenal da suavização dos dados (Figura 18). Apesar de observar-se uma diminuição na superestimação das vazões previstas pelo modelo com a suavização dos dados, ainda é evidente uma certa resistência por parte dos modelos em realizar essa superestimação.

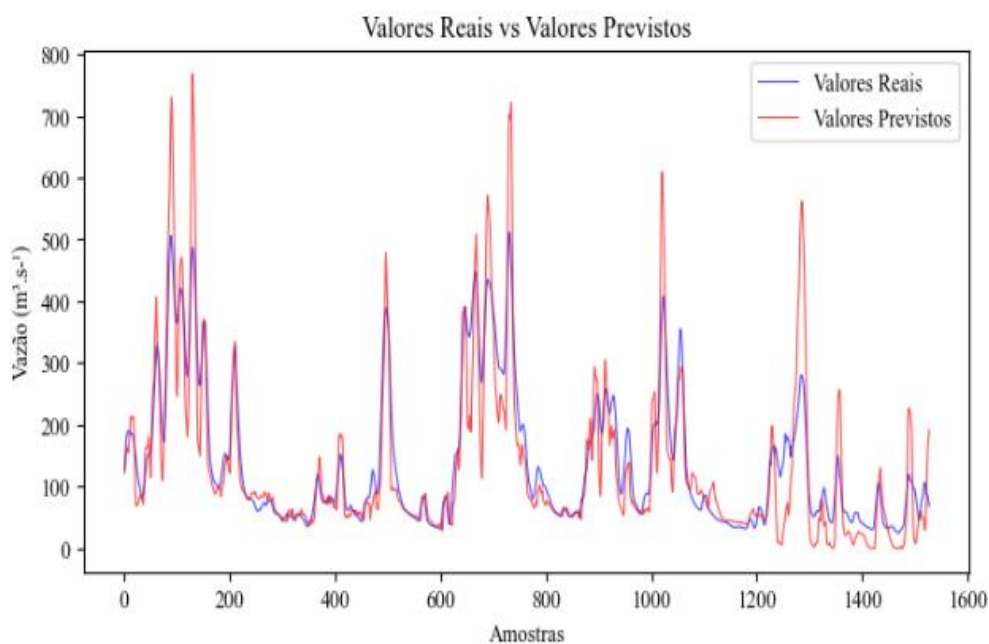


Figura 18 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 5 a partir dos dados suavizados em escala decidual.

Fonte: próprio autor (2023).

Para a escala quinzenal de suavização, a Arquitetura 3 mostrou-se a melhor em performance e na taxa de acerto das previsões (Tabela 7, Figura 19). As superestimações dos picos de vazão foram diminuídas para uma diferença de  $100 \text{ m}^3.\text{s}^{-1}$  aproximadamente, como pode ser visto na primeira metade do Hidrograma da Figura 18. Foi evidenciada apenas uma superestimação em torno dos  $200 \text{ m}^3.\text{s}^{-1}$  no final do Hidrograma, entretanto, em períodos em que a vazão real se encontrou abaixo dos  $300 \text{ m}^3.\text{s}^{-1}$ , o modelo mostrou-se eficiente nas previsões, por se aproximarem bastante dos dados reais.

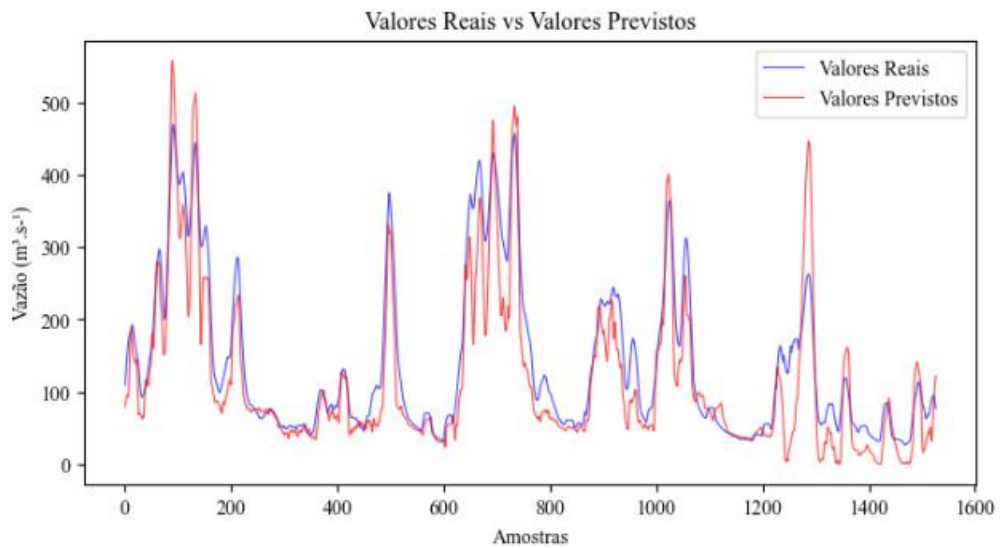


Figura 19 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 3 a partir dos dados suavizados em escala quinzenal.

Fonte: próprio autor (2023).

A melhoria dos modelos de redes neurais alimentados a partir de dados suavizados é encontrada na literatura. Wu e Chau (2013) utilizaram a metodologia de suavização por médias móveis em dados de precipitação diária e mensal na etapa de pré-processamento antes do treinamento de redes neurais MLP para a previsão de séries temporais. Os autores concluíram que o método de suavização utilizado para o pré-processamento das séries resultou em incrementos consideráveis das performances dos seus modelos para a previsão de chuvas.

Mehr e Kahya (2017) utilizaram Médias Móveis em escala de 3, 5 e 7 dias para suavizar dados diários de vazão em uma bacia na região da Turquia. Os autores encontraram uma melhoria nas performances das previsões com dados suavizados e quando comparado a dados sem a suavização. Entretanto, é destacado pelos autores que o alto valor da janela de suavização, os picos de vazão dos dados originais não são devidamente considerados. Dessa forma, se o interesse da modelagem é a previsão diária, os autores destacam que a escolha de janelas grandes na suavização diminui a eficiência de previsões diárias, sobretudo para os períodos de picos de vazão.

#### 4.4 Redes neurais com dados em escala mensal

A transformação dos dados diários para dados médios mensais das variáveis em estudo possibilitou uma melhoria na performance de três dos modelos propostos (Tabela 8). A obtenção das variáveis em escala mensal para a alimentação dos modelos também funciona como uma forma de suavização dos dados, visto que os valores extremos de cada variável são bastante diluídos a partir da obtenção da média. Isso pode ser visto a partir do Hidrograma da Figura 20. Os valores reais dos

dados de vazão observados apresentam uma variação em torno de 50 a 400 m<sup>3</sup>.s<sup>-1</sup>, diferentemente do que ocorreu ao utilizar-se dados suavizados em escalas quinzenal e decenal, ou mesmo a escala diária, como pôde ser visto nas seções 4.2 e 4.3.

Tabela 8 – Tabela resumo das métricas dos modelos testados com dados suavizados em escala mensal durante o treinamento, validação e teste.

Arquitetura	Dados de entrada	Dado de saída	Neurônios nas camadas ocultas*	Treinamento e validação			Teste		
				RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
1	Q		2	57,542	41,5725	0,3126	206,7908	164,6364	0
2	(montante),		7	24,2524	16,5407	0,8848	33,09	24,5231	0,8887
3	PPT (total),		4,8	27,7254	19,2389	0,8417	31,1731	22,3803	0,9012
4	PPT (montante),	Q (total)	6,2	42,8675	33,8715	0,511	168,0139	135,6108	0
5	ET (total),		2,4,5	32,4743	24,7655	0,6929	52,3244	40,7644	0,7217
6	ET (montante)		6,3,2	39,525	30,5494	0,5342	56,052	44,1901	0,6807

\*um número corresponde a uma camada oculta, dois números correspondem a duas camadas e assim por diante.  
Fonte: próprio autor (2023).

Dos modelos testados com dados mensais, a Arquitetura 3 obteve as melhores previsões, com o valor de R<sup>2</sup> = 0,9012, MAE = 22,3803 e RMSE = 31,1731 (Tabela 8, Figura 19) Resultados de Coeficiente de correlação similares foram encontrados por Yaghoubi, Hosseini e Nazif (2019), em simulações de vazões mensais. Os autores, ao utilizarem uma RNA com apenas duas camadas intermediárias de 14 e 7 neurônios, respectivamente, obtiveram R<sup>2</sup> = 0,90 para o treino e validação de seu modelo. Tanto em vazões abaixo dos 100 m<sup>3</sup>.s<sup>-1</sup> quanto para os picos próximos aos 400 m<sup>3</sup>.s<sup>-1</sup>, os valores previstos pela Arquitetura 3 aproximaram-se dos valores reais. Percebe-se que a diluição da variabilidade dos dados de entrada dos modelos com a obtenção da escala mensal, há um grande aumento nas taxas de acerto das previsões pelos modelos. Entretanto, para Yaghoubi, Hosseini e Nazif (2019), os valores previstos pelo modelo proposto subestimaram valores reais, principalmente em momentos de picos de vazão.



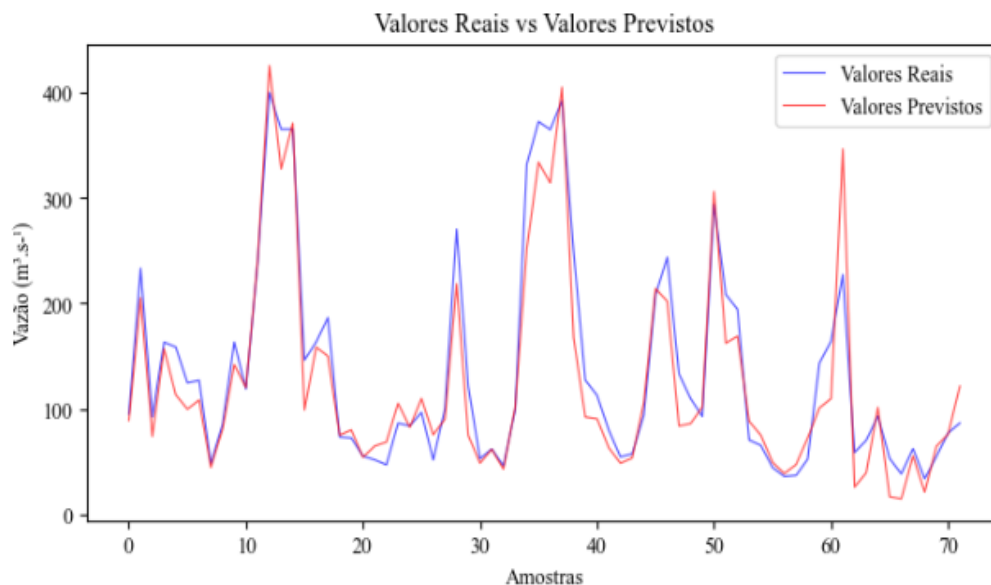


Figura 20 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 3 a partir dos dados suavizados em escala mensal.

Fonte: próprio autor (2023).

#### 4.5 Redes neurais com retirada de variáveis dos modelos

A retirada dos dados referente a porção montante da sub-bacia do rio Miranda influenciou negativamente a performance dos modelos, tanto em escala mensal quanto em escala diária (Tabelas 9 e 10).

Tabela 9 – Tabela resumo das métricas dos modelos testados com retirada de variáveis em escala diária durante o treinamento, validação e teste.

Arquitetura	Dados de entrada	Dado de saída	Neurônios nas camadas ocultas*	Treinamento validação			Teste		
				RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
1	Q (montante), PPT (total), ET (total)	Q (total)	2	64,4536	42,6174	0,4829	85,1431	58,1117	0,4351
2			56,1551	36,0448	0,6176	85,6136	60,292	0,4288	
3			55,8678	35,5414	0,6196	87,9711	60,1538	0,3969	
4			71,3916	47,8091	0,3651	178,4542	137,9017	0	
5			69,0455	45,8217	0,3798	182,4768	140,6731	0	
6			69,2713	46,2825	0,3694	178,4497	137,8828	0	

\*um número corresponde a uma camada oculta, dois números correspondem a duas camadas e assim por diante.  
Fonte: próprio autor (2023).

Tabela 10 – Tabela resumo das métricas dos modelos testados com retirada de variáveis em escala mensal durante o treinamento, validação e teste.

Arquitetura	Dados de entrada	Dado de saída	Neurônios nas camadas ocultas*	Treinamento e validação			Teste		
				RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
1			2	29,3563	18,853	0,8366	55,9106	42,0584	0,6823
2			7	23,0418	15,2984	0,8953	32,989	23,7312	0,8894
3	Q (montante), PPT (total), ET (total)	Q (total)	4,8	24,3043	16,1361	0,8894	67,4302	48,5519	0,5378
4			6,2	41,4911	30,5379	0,528	168,637	136,382	0
5			2,4,5	32,7087	24,7286	0,6902	61,4967	44,2753	0,6156
6			6,3,2	24,1843	16,9325	0,8814	91,9642	75,445	0,1404

\*um número corresponde a uma camada oculta, dois números correspondem a duas camadas e assim por diante.  
Fonte: próprio autor (2023).

Observando os três coeficientes de determinação negativos das Arquiteturas 4 a 6 durante a etapa de validação (Tabela 9), verifica-se que para os modelos alimentados com dados em escala diária a queda de performance é mais evidente do que para os modelos que utilizaram dados mensais. Embora os dados da porção montante retirados dos modelos tivessem apresentado baixa correlação com a vazão de jusante como visto na matriz de correlação (Figura 7), a inserção dessas informações permitiu aumento das performances dos modelos previamente testados nas seções de 4.2 a 4.4. Em termos de arquiteturas, a redução das variáveis provocou baixas taxas de acerto na maioria dos modelos, porém, a Arquitetura 2 alimentada por dados mensais mostrou-se eficiente nas previsões (Figura 21). O número de uma camada oculta nos modelos apresentou as melhores métricas na avaliação, tanto para as etapas de treinamento e teste quanto para a validação.

A Arquitetura 1 obteve um elevado coeficiente de determinação ( $R^2 = 0,8366$ ) para o treinamento e validação, entretanto, foi verificada uma diminuição durante a etapa de teste ( $R^2 = 0,6823$ ). Já a Arquitetura 2, com sete neurônios em sua camada intermediária apresentou uma manutenção dessa métrica, passando de 0,8953 para 0,8894, sendo o modelo com melhor performance dentre os avaliados.

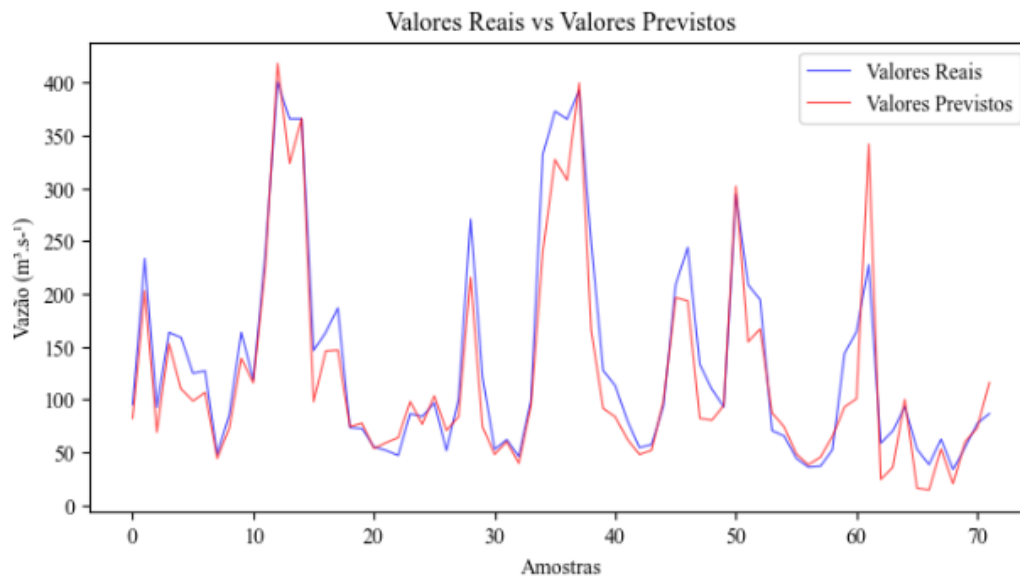


Figura 21 – Hidrograma com valores reais e previstos obtidos pela Arquitetura 2 com retirada de variáveis a partir dos dados suavizados em escala mensal.

Fonte: próprio autor (2023).

## 5. CONSIDERAÇÕES FINAIS

De acordo com Mehr e Kahya (2017), independentemente do tipo de técnica de previsão em modelos Data Driven, a acurácia dos modelos é altamente dependente das variáveis utilizadas durante o treinamento e validação. Portanto, a qualidade dos dados de entrada dos modelos é um fator importante para o sucesso das previsões de uma rede neural artificial. É possível observar de acordo com os dados apresentados na Figura 11 que os últimos anos dos dados de vazão apresentam uma tendência de aumento. Isso possivelmente pode ser um fator que contribuiu para os baixos valores das métricas RMSE,  $R^2$  e MAE dos modelos testados com os dados diários.

Além disso, os dados de Precipitação e Evapotranspiração usados para a alimentação dos modelos são provenientes de estimativas feitas por produtos de satélite, em que devido sua própria natureza, existem erros embutido não calculados, tanto pela forma que essas informações são geradas quanto pela resolução espacial dos produtos. Além disso, boa parte dos estudos que apresentam RNAs com previsões satisfatórias de vazão ou de chuvas usam dados em escala mensal ou anual. O mesmo ocorre com as análises feitas nesse estudo, em que os dados suavizados, ao servirem como entrada dos modelos, geraram previsões com métricas superiores. Entretanto, como citado por Mehr e Kahya (2017), janelas grandes na suavização dos dados diminui a eficiência de previsões diárias.

Dessa forma, o modelador deve buscar um equilíbrio entre o uso de técnicas de suavização de dados e previsões de vazões, de forma que os modelos não se afastem da realidade do ciclo hidrológico desconsiderando períodos de cheias ou secas. Além disso, o desempenho das redes neurais MLP testadas apresentaram variação de acordo com os hiperparâmetros da rede escolhidos, como o número de neurônios e camadas intermediárias utilizadas em suas arquiteturas. Portanto, deve-se ter cautela com a definição das arquiteturas das redes para a modelagem.

Ainda que as métricas calculadas dos modelos se apresentem inferiores a outros estudos da literatura, para bacias pouco monitoradas com baixa densidade de informação, essa pode ser uma técnica aprimorada e usada para prever vazões e promover sistemas de alerta.

Como sugestão para trabalhos futuros pode ser feita a utilização de outras arquiteturas e hiperparâmetros de rede neural artificial do tipo Multi Layer Perceptron ou até mesmo o uso de outros tipos de rede neural para a modelagem hidrológica. Um estudo de análise comparativa entre a rede neural do tipo LSTM (Long Short-Term Memory) e a Multi Layer Perceptron pode ser feito para verificar se qual técnica de RNA melhor se aplica aos dados de vazão estudados. Uma segunda abordagem seria o uso de diferentes coleções de imagens de satélite como dados de entrada de precipitação e evapotranspiração ou a inclusão de novas variáveis que façam parte do processo hidrológico de uma bacia hidrográfica. O estudo e teste de modelos de RNA para outras sub-bacias componentes do planalto adjacente do Pantanal Matogrossense como forma de compreender o montante de água que entra no bioma seria uma outra abordagem a ser estudada.

A modelagem com redes neurais artificiais pode ter aplicações práticas. Com as RNAs treinadas e validadas, poderia ser criado um sistema automatizado de previsão de vazões para o bioma do Pantanal, que utiliza dados de satélite de evapotranspiração, precipitação e vazão de estações fluviométricas. Seria necessário desenvolver um *script* que permitisse a coleta, processamento e atualização diária desses dados. Esse sistema exigiria acesso contínuo à internet para acessar as fontes de dados relevantes, como serviços de satélite do Google Earth Engine e informações de estações fluviométricas da Agência Nacional de Águas. Através de uma API (*Application Programming Interface* ou Interface de Programação de Aplicações), o sistema seria capaz de extrair os dados necessários, realizar o pré-processamento e montar uma base de dados atualizada. Em seguida, os modelos de previsão de vazões, já treinados, seriam aplicados a esses dados para gerar previsões atualizadas diariamente, fornecendo informações valiosas para a gestão e monitoramento hidrológico do bioma do Pantanal.

## 6. CONCLUSÕES

Com base nos resultados e discussão realizados previamente e de acordo com os objetivos propostos pelo trabalho, conclui-se que:

- a) A sub-bacia hidrográfica do rio Miranda apresenta variabilidade nos dados de vazão para os anos de 2000 a 2020, tanto para a estação de Miranda (cód. 66910000) quanto para a estação de Estrada-MT (cód. 66900000), sendo possível observar *outliers* bastante acima dos valores médios dessa variável;
- b) A precipitação da Coleção CHIRPS v 2.0 apresentam um comportamento semelhante para as duas porções da sub-bacia considerada, também observando-se *outliers* bastante acima das precipitações médias para os anos de 2000 a 2020;
- c) A evapotranspiração da Coleção GLDAS v 2.0 apresentam um comportamento cíclico bastante definido, para o período estudado;
- d) Para a maioria dos modelos testados, as Arquitetura 2 (uma camada oculta composta de sete neurônios) e 3 (duas camadas ocultas compostas por quatro e oito neurônios, respectivamente) obtiveram os melhores resultados na avaliação de performance de acordo com as métricas de coeficiente de determinação ( $R^2$ ), Raiz quadrada do erro médio (RMSE) e Erro médio absoluto (MAE);
- e) A mudança na escala temporal dos dados de entrada dos modelos melhorou a performance das RNAs propostas. A escala mensal proporcionou a melhor performance pela suavização dos dados, seguido da técnica de suavização por Médias Móveis de quinze e dez dias, respectivamente.
- f) A retirada dos dados da porção montante da sub-bacia considerada influenciou negativamente a performance dos modelos propostos, embora apresentassem baixa correlação com os dados de vazão de jusante (*target* dos modelos).



## REFERÊNCIAS

- ABDALRAHMAN, G. et al. Modeling the infiltration rate of wastewater infiltration basins considering water quality parameters using different artificial neural network techniques. **Engineering Applications of Computational Fluid Mechanics**, v. 16, n. 1, p. 397-421, 2022.
- ABREU, A. M. de. **Modelagem Numérica e Sensoriamento Remoto**: Um estudo para a caracterização hidrológica e estimativa de vazões na bacia hidrográfica do Alto Paraguai. Rio de Janeiro, RJ. 2021. 110 f. Dissertação (Mestrado Profissional em Rede Nacional em Gestão e Regulação de Recursos Hídricos) - Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.
- ADIKARI, K. E. et al. Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. **Environmental Modelling & Software**, v. 144, p. 105136, 2021.
- ALHO, C. J.R; SILVA, J. S.V. Effects of severe floods and droughts on wildlife of the Pantanal wetland (Brazil)—a review. **Animals**, v. 2, n. 4, p. 591-610, 2012.
- ALOTAIBI, K. et al. Future predictions of rainfall and temperature using GCM and ANN for arid regions: a case study for the Qassim Region, Saudi Arabia. **Water**, v. 10, n. 9, p. 1260, 2018.
- ANA. Agência Nacional de Águas. Relatório de Diagnóstico Final de Hidrologia – Inundações, 2020.
- APAYDIN, H. et al. Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. **Water**, v. 12, n. 5, p. 1500, 2020.
- BĄCZKIEWICZ, A. et al. An ANN Model Trained on Regional Data in the Prediction of Particular Weather Conditions. **Applied Sciences**, v. 11, n. 11, p. 4757, 2021.
- BATALHA, M. A. O cerrado não é um bioma. **Biota Neotrop.**, vol. 11, no. 1, 2011.
- BECK, H. E. et al. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. **Hydrology and Earth System Sciences**, v. 21, n. 12, p. 6201-6217, 2017.
- BELVEDERESI, C. et al. Predicting river flow using an AI-based sequential adaptive neuro-fuzzy inference system. **Water**, v. 12, n. 6, p. 1622, 2020.
- BERGIER, I. Effects of highland land-use over lowlands of the Brazilian Pantanal. **Science of the Total Environment**, v. 463, p. 1060-1066, 2013.
- BERTONI, J. C., TUCCI, C. E. M. Precipitação. In: TUCCI, C. E. M. **Hidrologia: ciência e aplicação**. Porto Alegre, UFRGS. p. 177 — 241. 2007.
- BEVEN, K. Deep learning, hydrological processes and the uniqueness of place. **Hydrological Processes**, v. 34, n. 16, p. 3608-3613, 2020.
- CARDOSO, E. L. et al. Atributos biológicos indicadores da qualidade do solo em pastagem cultivada e nativa no Pantanal. Brasília. **Pesq. Agropec. Bras**, v.44, n. 6, p. 631-637, jun, 2009.
- CARVALHO, A. P. et al. Importância do manejo da bacia hidrográfica e da determinação de processos hidrológicos. **Revista brasileira de agropecuária sustentável**, v.2, n.2, 2012.
- CHARLES, T. S. **Regionalização hidrológica para o estado de Goiás e Distrito Federal**. Dissertação de Mestrado (Engenharia de Sistemas Agrícolas) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2020.



- COLLISCHONN, B. et al. Desempenho do satélite TRMM na estimativa de precipitação sobre a bacia do Paraguai superior. **Revista Brasileira de Cartografia**, v. 59, n. 1, p. 93-99, 2007.
- COSTI, A. C. Z. **Avaliação de dados altimétricos de níveis de água adquiridos pelo satélite ENVISAT na bacia amazônica**. Brasília, DF. 2012. 141 f. Dissertação (Mestrado em tecnologia ambiental e recursos hídricos) - Universidade de Brasília, Brasília, 2012.
- CRISTALDO, M. F. et al. Analysis and distribution of the rainfall monitoring network in a Brazilian Pantanal region. **Revista Brasileira de Meteorologia**, v. 32, n. 2, p. 199-205, 2017.
- CRISTALDO, M. F. et al. Previsão probabilística de enchentes para uma pequena bacia hidrográfica do Pantanal. **Revista Ambiente & Água**, v. 13, p. e1988, 2018.
- CURADO, L. C. **Indicadores de vazões mínimas de referência em sub bacias do rio Miranda**. Dissertação de Mestrado (Saneamento Ambiental e Recursos Hídricos) – Universidade Federal do Mato Grosso do Sul, Campo Grande, 2003.
- DINIZ, L. S. **Regionalização de parâmetros de modelo chuva-vazão usando redes neurais**. 2008. Tese Doutorado (Engenharia de Recursos Hídricos e Saneamento Ambiental) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.
- FERREIRA, A. B. de B. Pantanal Mato-Grossense: considerações sobre a proteção constitucional para um desenvolvimento econômico sustentável. Campo Grande. **Interações**, v. 14, n. 1, p. 11-20, jan./jun. 2013.
- FUNK, C. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. **Scientific data**, v. 2, n. 1, p. 1-21, 2015.
- GALDINO, S. et al. **Impactos Ambientais e Socioeconômicos na Bacia do Rio Taquari – Pantanal**. Corumbá: Embrapa Pantanal, 2006.
- GALDINO, S. et al. **Mudanças no Regime Hidrológico da Bacia Hidrográfica do Rio Taquari – Pantanal**. Corumbá: Embrapa Pantanal, 2002. (Embrapa Pantanal. Boletim de pesquisa e desenvolvimento, 26).
- GETIRANA, A. C.V. et al. Hydrological monitoring of poorly gauged basins based on rainfall–runoff modeling and spatial altimetry. **Journal of hydrology**, v. 379, n. 3-4, p. 205-219, 2009.
- GUIMARÃES, D. P. et al. Caracterização das chuvas no Pantanal Mato-grossense. em: Anais do VII Simpósio de Geotecnologias no Pantanal. EMBRAPA/INPE, Jardim, Brasil, p. 555-562, 2018.
- HAMILTON, S. K. Human impacts on hydrology in the Pantanal wetland of South America. **Water Science and Technology**, v. 45, n. 11, p. 35-44, 2002.
- IBRAHIM, K. S. M. H. et al. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. **Alexandria Engineering Journal**, v. 61, n. 1, p. 279-303, 2022.
- JAISWAL, R. K.; ALI, S.; BHARTI, B. Comparative evaluation of conceptual and physical rainfall–runoff models. **Applied water science**, v. 10, p. 1-14, 2020.
- JUNG, K. et al. Evaluation of nitrate load estimations using neural networks and canonical correlation analysis with k-fold cross-validation. **Sustainability**, v. 12, n. 1, p. 400, 2020.
- JUNG, Yoonsuh. Multiple predicting K-fold cross-validation for model selection. **Journal of Nonparametric Statistics**, v. 30, n. 1, p. 197-215, 2017.

- KADAM, A. K. et al. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. **Modeling Earth Systems and Environment**, v. 5, p. 951-962, 2019.
- KASSEM, A. A.; RAHEEM, A. M.; KHIDIR, K. M. Daily streamflow prediction for khazir river basin using ARIMA and ANN models. **Zanco Journal of Pure and Applied Sciences**, v. 32, n. 3, p. 30-39, 2020.
- KHAN, M. et al. Streamflow forecasting for the Hunza river basin using ANN, RNN, and ANFIS models. **Water Practice and Technology**, v. 18, n. 5, p. 981-993, 2023.
- KIRPICH, T.P. Time of Concentration of Small Agricultural Watersheds. **J. of Civ. Eng.**, v.10, n.6, p. 362, 1940.
- LEITE, E. F.; BEREZUK, A. G.; SILVA, C. A. A vulnerabilidade ambiental da bacia hidrográfica do rio Miranda, Mato Grosso do Sul. **Revista Brasileira de Geografia Física**, v. 15, n. 05, p. 2613-2639, 2022.
- LOHMANN, M. Análise dos alagamentos no município de Curitiba entre os anos de 2005 a 2010. **Revista Ciência Geográfica**, v. 17, p. 135-149, 2013.
- LOHMANN, M. **Regressão logística e redes neurais aplicadas à previsão probabilística de alagamentos no município de Curitiba, PR**. Curitiba, PR. 2011. 231 f. Tese (Doutorado em Geografia) – Universidade Federal do Paraná, Curitiba, 2011.
- LUIZ SILVA, W. et al. Climatological and hydrological patterns and verified trends in precipitation and streamflow in the basins of Brazilian hydroelectric plants. **Theoretical and Applied Climatology**, v. 137, p. 353-371, 2019.
- MAGETO, G. A. S. et al. Artificial intelligence applications in the agriculture 4.0. **Revista Ciência Agrônômica**, v. 51, Special Agriculture 4.0, 2020.
- MAIER, H. R. et al. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. **Environmental Modelling & Software**, 25(8), 2010.
- MALFATTI, M. G. L.; CARDOSO, A. de O.; HAMBURGER, D. S. Modelo empírico linear para previsão de vazão de rios na usina hidrelétrica de Itaipu–Bacia do Rio Paraná. **Revista Brasileira de Meteorologia**, v. 33, p. 257-268, 2018.
- MARENCO, J. A. et al. Increased climate pressure on the agricultural frontier in the Eastern Amazonia–Cerrado transition zone. **Scientific reports**, v. 12, n. 1, p. 457, 2022.
- MATO GROSSO DO SUL. Instituto do Meio Ambiente de Mato Grosso do Sul/IMASUL. Diretoria de Desenvolvimento. Relatório de qualidade de águas superficiais de Mato Grosso do Sul, 2009/10. Campo Grande: IMASUL, p. 65-107, 2012.
- MEHR, A. D.; KAHYA, E. A Pareto-optimal moving average multigene genetic programming model for daily streamflow prediction. **Journal of hydrology**, v. 549, p. 603-615, 2017.
- MENDONÇA, L. M. et al. Modelagem chuva-vazão via redes neurais artificiais para simulação de vazões de uma bacia hidrográfica da Amazônia. **Revista de Gestão de Água da América Latina**, v. 18, n. 2021, 2021.
- MERINO, E. R. et al. Estilos fluviais e evidências de mudanças ambientais na planície do rio Miranda, Pantanal. **Revista Brasileira de Geomorfologia**, v. 14, n. 2, 2013.

- MULVANY, T.J. On the use of self-registering rain and flood gauges, in making observations of the relations of rain fall and of flood discharges in a given catchment. **Proc. Inst. Civ. Eng.** (Irel.), v.4, pt.2, p. 18-33, 1851.
- NG, K. W. et al. A Review of Hybrid Deep Learning Applications for Streamflow Forecasting. **Journal of Hydrology**, p. 130141, 2023.
- OLAWOYIN, A.; CHEN, Y. Predicting the future with artificial neural network. **Procedia computer science**, v. 140, p. 383-392, 2018.
- OLIVEIRA, A. R.; RAMOS, T. B.; NEVES, R. Streamflow estimation in a mediterranean watershed using neural network models: A detailed description of the implementation and optimization. **Water**, v. 15, n. 5, p. 947, 2023.
- PADOVANI, C. R. **Dinâmica espaço-temporal das inundações do Pantanal**. Piracicaba, SP. 2010. 175 f. Tese (Doutorado em Ecologia Aplicada) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2010.
- PATIL, A. P; DEKA, P. C. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. **Computers and Electronics in Agriculture**, v. 121, p. 385 – 392, 2016.
- PEREIRA, G. et al. Rainfall patterns and geomorphological controls driving inundation frequency in tropical wetlands: How does the Pantanal flood?. *Progress in Physical Geography: Earth and Environment*, v. 45, n. 5, p. 669-686, 2021.
- PEREIRA, G.; CHÁVEZ, E. S.; SILVA, M. E. S. O estudo das unidades de paisagem do bioma Pantanal. **Revista Ambiente & Água – An Interdisciplinary Journal of Applied Science**, v. 7, n. 1, 2012.
- POTT, A.; DAMASCENO-JUNIOR, G. A.; SILVA, M. P. Características da bacia hidrográfica do Rio Miranda. **Revista GeoPantanal**, v. 9, n. 16, p. 125-140, 2014.
- ROHANI, A.; TAKI, M.; ABDOLLAHPOUR, M. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). **Renewable Energy**, v. 115, p. 411-422, 2018.
- RUEZZENE, C. B. et al. Preenchimento de falhas em dados de precipitação através de métodos tradicionais e por inteligência artificial. **Revista Brasileira de Climatologia**, v. 29, p. 177-204, 2021.
- RUHOFF, Anderson et al. Global Evapotranspiration Datasets Assessment Using Water Balance in South America. **Remote Sensing**, v. 14, n. 11, p. 2526, 2022.
- SALEM, H. et al. Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization. **Desalination**, v. 522, p. 115411, 2022.
- SANTOS ANTONIASSI, R. A. **Predição de nível de rios da Região Hidrográfica do Rio Paraguai utilizando algoritmos de Aprendizado de Máquina**. Dissertação de Mestrado (Ciência da Computação) – Universidade Federal do Mato Grosso do Sul, Campo Grande, 2022.
- SANTOS NETO, L. A. et al. Modelagem hidroclimatologica utilizando redes neurais multi layer perceptron em bacia hidrográfica no sudoeste da Amazônia. **Revista Brasileira de Climatologia**, v. 26, 2020.
- SANTOS, J. R. S.; CELESTE, A. B. Avaliação de estratégias de modelagem guiada por dados para previsão de vazão em rio sergipano. Taubaté. **Revista Ambiente e Água**, v. 9, n. 3, jul/set, 2014.
- SILVA, J. S. V.; ABDON, M., M.; MORAES, J. A. **Desmatamento na bacia do Alto Paraguai no Brasil**. Anais 3º Simpósio de Geotecnologias no Pantanal, Cáceres, MT, Embrapa Informática Agropecuária/INPE, p. 458 458 -467, 2010.

- SNIEDER, E.; ABOGADIL, K.; KHAN, U. T. Resampling and ensemble techniques for improving ANN-based high-flow forecast accuracy. **Hydrology and Earth System Sciences**, v. 25, n. 5, p. 2543-2566, 2021.
- SOUSA, F. A.; MOURA, D. M. B. Evapotranspiração potencial (ETp) e sua influência na vazão de rios do Cerrado Brasileiro. Elisée, **Rev. Geo. UEG – Goiás**, v.11, n.1, e111222, jan. /jun, 2022.
- SOUSA, W. dos S.; SOUSA, F. de A. S. Rede neural artificial aplicada à previsão de vazão da Bacia Hidrográfica do Rio Piancó. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, p. 173-180, 2010.
- SOUZA, N. S.; SOUZA, W. J.; CARDOSO, J. M. S. Caracterização hidrológica e influência da cobertura do solo nos parâmetros de vazão do Rio das Fêmeas. **Engenharia Sanitaria e Ambiental**, v. 22, p. 453-462, 2017.
- TUCCI, C. E. M; CLARKE, R. T. Impacto das mudanças da cobertura vegetal no escoamento: revisão. **Rbrh: Revista Brasileira de Recursos Hídricos**. Porto Alegre, RS. vol. 2, n. 1 (jun. 1997), p. 135-152, 1997.
- ULIANA, E. M. et al. Modelo Hidrológico Híbrido para Previsão de Vazões na Bacia do Rio Piracicaba-MG. **Revista Brasileira de Meteorologia** v. 34, p. 471-480, 2020.
- VALERIANO M. M., SALVI L.L., ARAGÃO, J.R.L. Relações entre a distribuição da precipitação e o relevo da bacia do alto Paraguai. Anais 4º Simpósio de Geotecnologias no Pantanal, Bonito, MS, 20–24 de outubro 2012. Embrapa Informática Agropecuária/INPE, p. 289–98, 2012.
- WEBER, O.L.S.; COUTO, E.G. Dinâmica da matéria orgânica no complexo do Pantanal. In: SANTOS, G. de A.; SILVA, L.S. da; CANELLAS, L.P.; CARMARGO, F. de O. (Ed.). **Fundamentos da matéria orgânica do solo: ecossistemas tropicais & subtropicais**. 2.ed. rev. e atual. Porto Alegre, p. 463-482, 2008.
- WENZEL, D. A. et al. Características fisiográficas de sub-bacias do médio e alto Rio Teles Pires-MT. **Revista de Ciências Agroambientais**, v. 15, n. 2, p. 123-131, 2017.
- WMO, 2008. Guide to Hydrological Practices: Volume I Hydrology – From Measurement to Hydrological Information., Geneva, Switzerland.
- WU, C. L.; CHAU, K.W. Prediction of rainfall time series using modular soft computing methods. **Eng. Appl. Artif. Intell.** V. 26, p. 997-1007, 2013.
- YAGHOUBI, B.; HOSSEINI, S. A.; NAZIF, S. Monthly prediction of streamflow using data-driven models. **Journal of Earth System Science**, v. 128, p. 1-15, 2019.
- YASEEN, Z. M. et al. Artificial intelligence based models for stream-flow forecasting: 2000–2015. **Journal of Hydrology**, v. 530, p. 829-844, 2015.
- YILMAZ, M. U.; ONOZ, B. A comparative study of statistical methods for daily streamflow estimation at ungauged basins in Turkey. **Water**, v. 12, n. 2, p. 459, 2020.
- ZAKARIA, M. N. A. et al. Exploring machine learning algorithms for accurate water level forecasting in Muda River, Malaysia. **Heliyon**, v. 9, n. 7, 2023.