

## UTILIZING LONG SHORT-TERM MEMORY (LSTM) NETWORKS FOR RIVER FLOW PREDICTION IN THE BRAZILIAN PANTANAL BASIN

C. S. DESCOVI\*, A. C. ZUFFO, S.M. MOHAMMADIZADEH, L. F. MURILLO-BERMÚDEZ, D. A. SIERRA

State University of Campinas (UNICAMP)

ORCID ID: <https://orcid.org/0000-0002-2847-2064>\*

[cassianodescovi@fec.unicamp.br](mailto:cassianodescovi@fec.unicamp.br)\*

Submitted June 5, 2023 – Accepted September 11, 2023

DOI: 10.15628/holos.2023.16315

### ABSTRACT

This article demonstrates the successful application of Long Short-Term Memory (LSTM) recurrent neural networks to simulate streamflow in the Aquidauana River basin, located in the Brazilian Pantanal. The LSTM network used daily precipitation data as input to predict future streamflow in the region. The results obtained from this research show a coefficient of determination ( $R^2$ ) of 0.82, indicating a strong fit of the model to the observed data. Additionally, the Root Mean Squared Error (RMSE) was found to be 0.53, indicating the model's

accuracy in predicting streamflow compared to actual data. These findings highlight the effectiveness of LSTM networks in hydrological modeling for the Pantanal region, which is crucial for water resource planning and sustainable management in this ecologically significant area. This study is expected to serve as a catalyst for further research and make a substantial contribution to the advancement of streamflow prediction techniques in complex watersheds such as the Aquidauana River basin.

**KEYWORDS:** LSTM, River flows, Pantanal.

## REDES DE MEMÓRIA DE LONGO E CURTO PRAZO (LSTM) PARA PREDIÇÃO DE FLUXO DE RIO NA BACIA DO PANTANAL BRASILEIRO

### RESUMO

Este artigo mostra uma aplicação bem-sucedida de rede neural recorrente - Long Short-Term Memory (LSTM), para simular a vazão na bacia do rio Aquidauana, dentro dos limites do Pantanal brasileiro. Os dados diários de precipitação serviram como variáveis de entrada para permitir que a rede LSTM previsse o fluxo futuro na região. Os resultados obtidos demonstram um coeficiente de determinação ( $R^2$ ) de 0,82, indicando um ajuste favorável do modelo aos dados observados, juntamente com um erro quadrático médio (RMSE) de

0,53, demonstrando precisão na previsão do modelo em comparação com a vazão real. Tais métricas ressaltam a eficiência das redes LSTM para modelagem hidrológica na região do Pantanal, um aspecto crucial para o planejamento e gestão sustentável dos recursos hídricos na área. Espera-se que este estudo inspire novas pesquisas e contribua significativamente para o avanço das técnicas de previsão de vazões em bacias hidrográficas complexas e com deficiência de dados, como a bacia do Rio Aquidauana.

**Palavras chave:** LSTM, simulação de vazão, Pantanal.

## 1 INTRODUÇÃO

O Pantanal, localizado no coração da América do Sul, é a maior planície inundável do mundo. Esta região é conhecida não só pela sua excepcional diversidade ecológica, mas também pela sua contribuição crítica para a regulação da hidrologia regional. (Assine et al., 2016; Couto & Oliveira, 2010; Macedo et al., 2014). Entre os numerosos rios que serpenteiam por esta vasta área úmida, o Rio Aquidauana desempenha um papel significativo na preservação do delicado equilíbrio do ecossistema pantaneiro. É uma intrincada rede de rios, pântanos e diversos habitats, proporcionando um refúgio para inúmeras espécies de vida selvagem. A compreensão e a modelagem precisa do fluxo dos rios nesta região são de suma importância para a gestão dos recursos hídricos, controle de enchentes e conservação ambiental (Joia et al., 2018).

Contudo, apesar da importância ecológica do Pantanal, existe um desafio notável – a escassez de dados hidrológicos. Esta escassez de dados representa um obstáculo formidável ao desenvolvimento de modelos fiáveis de fluxo fluvial. Em particular, as áreas remotas e menos acessíveis do Pantanal carecem de observações hidrológicas abrangentes. Esta escassez impede abordagens convencionais de modelagem, enfatizando a necessidade de técnicas inovadoras capazes de capturar a intrincada dinâmica espaço-temporal do fluxo do rio. Para resolver essas limitações, este estudo abrange as capacidades das redes Long Short-Term Memory (LSTM), uma classe de redes neurais recorrentes reconhecidas por sua proficiência na captura de padrões sequenciais e dependências de longo alcance. (J. Fan et al., 2018; Kratzert et al., 2018; T. Liu et al., 2019).

O objetivo central desta pesquisa é empregar o LSTM para modelar a vazão fluvial na Bacia do Rio Aquidauana, com dados de precipitação servindo como uma variável de entrada crítica. O modelo LSTM será treinado para aprender as correlações históricas entre os padrões de precipitação e os fluxos dos rios, facilitando a geração de previsões precisas de fluxo, mesmo durante períodos com dados disponíveis limitados. Ao utilizar o LSTM, o estudo não apenas alcança uma modelagem precisa do fluxo fluvial, mas também oferece um meio de preencher lacunas de dados ausentes, uma consideração vital dada a escassez de registros hidrológicos na região do Pantanal.

Este manuscrito descreve a metodologia empregada para pré-processamento e integração dos dados hidrológicos e de precipitação disponíveis. O processo envolve a coleta e organização de dados históricos relativos à vazão e precipitação dos rios na Bacia do Rio Aquidauana. Os dados passam por normalização, garantindo que sejam padronizados em uma escala comum, o que é essencial para o funcionamento eficaz do modelo LSTM. Posteriormente, a arquitetura LSTM e o procedimento de treinamento são detalhados de forma abrangente, fornecendo insights sobre o funcionamento interno da rede neural e como ela processa sequências de dados.

A pesquisa vai além ao avaliar a proficiência do modelo LSTM em capturar as flutuações espaço-temporais da vazão dos rios na Bacia do Rio Aquidauana. Esta avaliação visa demonstrar quão bem a rede LSTM pode se adaptar às condições hidroclimáticas únicas da região do Pantanal e fornecer previsões precisas. Além disso, será realizada uma análise comparativa, contrastando as previsões do LSTM com aquelas geradas pelos modelos hidrológicos tradicionais. Esta avaliação

comparativa determinará a eficácia do LSTM no gerenciamento de limitações de dados e na captura de dinâmicas hidrológicas complexas.

Concluindo, este estudo representa um avanço significativo no domínio da modelagem hidrológica na região do Pantanal. Ele mostra as capacidades das redes LSTM na utilização eficaz de dados hidrológicos limitados para previsão de fluxo de rios e preenchimento de lacunas de dados. O fornecimento de simulações precisas do fluxo fluvial através desta pesquisa pode contribuir significativamente para processos de tomada de decisão bem informados relativos à gestão sustentável dos recursos hídricos, preservação ambiental e iniciativas de controle de enchentes no Pantanal. Estes são aspectos críticos para manter o equilíbrio ecológico e salvaguardar a biodiversidade desta região única.

Além disso, os conhecimentos obtidos com a utilização do LSTM para enfrentar os desafios da escassez de dados podem oferecer implicações valiosas para circunstâncias análogas noutras regiões do mundo. A abordagem adoptada neste estudo pode ser um modelo para abordar as limitações de dados na modelação hidrológica em vários ecossistemas, fornecendo uma solução inovadora para regiões onde a recolha de dados tradicional pode ser um desafio. As redes LSTM, com a sua capacidade de capturar dependências de longo alcance, têm o potencial de revolucionar o campo da hidrologia, permitindo previsões mais precisas e uma melhor gestão dos recursos hídricos em áreas ecologicamente sensíveis como o Pantanal. Esta pesquisa não só amplia a nossa compreensão da hidrologia do Pantanal, mas também oferece um caminho promissor para a gestão sustentável dos recursos hídricos em outras regiões vulneráveis do mundo. A contribuição das redes LSTM para a modelagem hidrológica é um passo em direção a uma abordagem mais resiliente e ecologicamente consciente para a gestão dos recursos hídricos, com implicações muito além do Pantanal.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Modelagem Hidrológica em Ecossistemas Alagados (Zonas úmidas)

O Pantanal, situado no centro da América do Sul, é um exemplo notável de ecossistema de zonas úmidas caracterizado por seus processos hidrológicos dinâmicos. As zonas úmidas, como o Pantanal, são ecossistemas únicos que servem como centros ecológicos vitais e são frequentemente chamados de “rins” da Terra devido ao seu papel na purificação e regulação da água. (Couto & Oliveira, 2010). Estas regiões são particularmente suscetíveis aos impactos das alterações climáticas, uma vez que alterações nos padrões de precipitação e temperatura podem influenciar significativamente a hidrologia das zonas húmidas. (Cui et al., 2021; S. M. Mohammadzadeh, Filho, Descovi, Murillo-Bermúdez, & Sierra, 2023; Murillo Bermudez et al., 2023; Sierra et al., 2023). Consequentemente, uma modelização hidrológica precisa é indispensável para compreender a resposta dos ecossistemas de zonas húmidas às mudanças ambientais, garantindo a sua conservação e apoiando a subsistência das comunidades que dependem destas regiões.

As zonas húmidas são caracterizadas pelos seus níveis de água flutuantes, que são influenciados por vários factores, incluindo a precipitação, a evapotranspiração, as interações das águas subterrâneas e a intrincada rede de rios e canais. O fluxo do rio dentro das zonas húmidas não é apenas crítico para o equilíbrio ecológico destas regiões, mas também desempenha um papel

crucial na ciclagem de nutrientes e no apoio à diversidade da vida selvagem. (Back et al., 2023; Xi et al., 2021). O Pantanal, com sua vasta planície aluvial e extensa rede de rios, exemplifica a intrincada interação entre a hidrologia das zonas úmidas e a saúde dos ecossistemas. Portanto, obter insights sobre a dinâmica hidrológica do Pantanal, e das zonas úmidas em geral, é essencial para a gestão ambiental sustentável, especialmente diante das mudanças nas condições climáticas.

Além disso, os desafios associados à modelação hidrológica em ecossistemas de zonas húmidas são multifacetados. Estas regiões carecem frequentemente de dados hidrológicos abrangentes, especialmente em áreas remotas. As abordagens tradicionais de modelização, que dependem fortemente de observações históricas e de métodos estatísticos bem estabelecidos, podem ser insuficientes para captar a complexidade da hidrologia das zonas húmidas. (Kratzert et al., 2018). A incorporação de técnicas inovadoras, tais como algoritmos avançados de aprendizagem automática, como redes Long Short-Term Memory (LSTM), torna-se imperativa para colmatar as lacunas na nossa compreensão da hidrologia das zonas húmidas. As redes LSTM, com a sua capacidade de capturar padrões sequenciais e adaptar-se às condições em mudança, oferecem soluções promissoras para modelar a natureza complexa e dinâmica do fluxo do rio dentro dos ecossistemas de zonas húmidas.

## 2.2 Redes de memória longa e de curto prazo (LSTM)

Nos últimos anos, as redes Long Short-Term Memory (LSTM) surgiram como uma ferramenta poderosa no domínio da inteligência artificial e da aprendizagem automática, particularmente para modelar dados dependentes do tempo. LSTMs pertencem à classe de redes neurais recorrentes (RNNs), mas possuem uma arquitetura única que aborda um dos principais desafios na análise sequencial de dados: o problema do gradiente de fuga (Hochreiter & Schmidhuber, 1997). Esse problema ocorre quando as RNNs tradicionais lutam para capturar dependências de longo prazo em dados sequenciais, tornando-as menos adequadas para tarefas que exigem a lembrança de informações por longos períodos de tempo. Os LSTMs foram projetados para superar essa limitação, tornando-os adequados para aplicações em previsão de séries temporais, processamento de linguagem natural e, principalmente, hidrologia.

No centro de uma rede LSTM está a célula de memória, que pode armazenar informações em intervalos de tempo estendidos e reter ou esquecer seletivamente as informações à medida que processa dados sequenciais. Esta célula de memória é complementada por três portas: a porta de entrada, a porta de esquecimento e a porta de saída. Essas portas controlam o fluxo de informações que entram e saem da célula de memória, permitindo que o LSTM capture e retenha padrões e contexto relevantes enquanto descarta informações menos importantes. (Mohammadzadeh et al., 2023; Yu et al., 2019). Essa arquitetura torna os LSTMs altamente adeptos da modelagem de sequências com intervalos de tempo variados, tornando-os particularmente valiosos para aplicações onde observações passadas influenciam significativamente os resultados futuros, como modelagem de fluxo de rios.

No contexto da hidrologia, onde os dados de séries temporais desempenham um papel central, os LSTMs oferecem diversas vantagens sobre os modelos estatísticos tradicionais (Fan et al., 2021; Gavidia, Mohammadzadeh, et al., 2023; Kratzert et al., 2019; Lees et al., 2021; Nikeghbali et al., 2014; Xiang et al., 2020). Eles podem capturar relações complexas e não lineares entre variáveis de entrada, como padrões de precipitação, e a variável alvo, como o fluxo do rio. Além

disso, os LSTMs podem se adaptar às mudanças nos padrões e à sazonalidade dos dados, tornando-os versáteis para modelar processos hidrológicos influenciados pelo clima, tempo e outros fatores. Sua capacidade de lidar com intervalos de tempo irregulares e dados faltantes aumenta ainda mais sua utilidade na modelagem hidrológica, especialmente em regiões com disponibilidade limitada de dados, como o Pantanal (Assine et al., 2016).

Como resultado, a aplicação de redes LSTM à hidrologia ganhou atenção considerável em pesquisas recentes (Descovi et al., 2023; Gavidia, Chinelatto, et al., 2023; Kratzert et al., 2018; Li et al., 2023; S. Mohammadzadeh et al., 2021; S. M. Mohammadzadeh, Filho, Descovi, Murillo-Bermúdez, & Sierra, 2023; Sahoo et al., 2019). Os pesquisadores aproveitaram os modelos LSTM para melhorar a precisão da previsão do fluxo dos rios, da previsão de inundações e da compreensão de fenômenos hidrológicos complexos (Liu et al., 2020). A capacidade inerente dos LSTMs de capturar flutuações de curto prazo e dependências de longo prazo em dados hidrológicos alinha-se com os desafios colocados por ecossistemas de zonas úmidas como o Pantanal, onde o fluxo do rio é influenciado por múltiplos fatores que interagem e exibe dinâmicas espaço-temporais intrincadas. Portanto, a integração das redes LSTM no estudo do fluxo dos rios no Pantanal apresenta um caminho promissor para melhorar a nossa compreensão deste ecossistema vital.

### 3 METODOLOGIA

#### 3.1 Área de Estudo e Coleta de Dados

A área de estudo de significativa importância hidrológica pertence à Bacia do Rio Aquidauana, região localizada no vasto Pantanal, uma extensa área úmida tropical situada no coração geográfico da América do Sul. O Pantanal abrange principalmente os estados brasileiros de Mato Grosso e Mato Grosso do Sul, estendendo-se por partes da Bolívia e do Paraguai. Cobrindo uma área estimada que varia de aproximadamente 150 mil a 195 mil quilômetros quadrados, o Pantanal se destaca como a maior área úmida de água doce em escala global. Assume um papel fundamental como hotspot ecológico e um recurso natural inestimável e de grande importância.

O Rio Aquidauana, um dos numerosos rios que atravessam o Pantanal, ocupa uma posição fundamental na formação da dinâmica hidrológica da região. Servindo como afluente do Rio Paraguai, um importante rio da América do Sul, o Rio Aquidauana influencia substancialmente o fluxo geral e o regime hídrico desta intrincada zona úmida. A bacia hidrográfica abrange uma paisagem diversificada e complexa, com zonas úmidas, savanas e florestas tropicais. A topografia da bacia, em conjunto com o seu clima tropical, dá origem a um sistema hidrológico complexo caracterizado por variações sazonais que incluem períodos de cheias e secas.

A bacia hidrográfica do Rio Aquidauana (conforme ilustrado na Figura 1) constitui uma sub-bacia do Rio Miranda, ele próprio um dos afluentes do Rio Paraguai. Geograficamente, está situado entre as latitudes de 19° 19' 01" e 21° 13' 49" sul, e as longitudes de 56° 49' 11" e 54° 16' 44" oeste. Esta bacia está localizada no centro-noroeste do estado de Mato Grosso do Sul, estendendo-se desde a Serra de Maracaju, situada no município de São Gabriel do Oeste, até a extensa planície pantaneira, onde converge com o Rio Miranda. Rio dentro dos limites do município de Aquidauana.

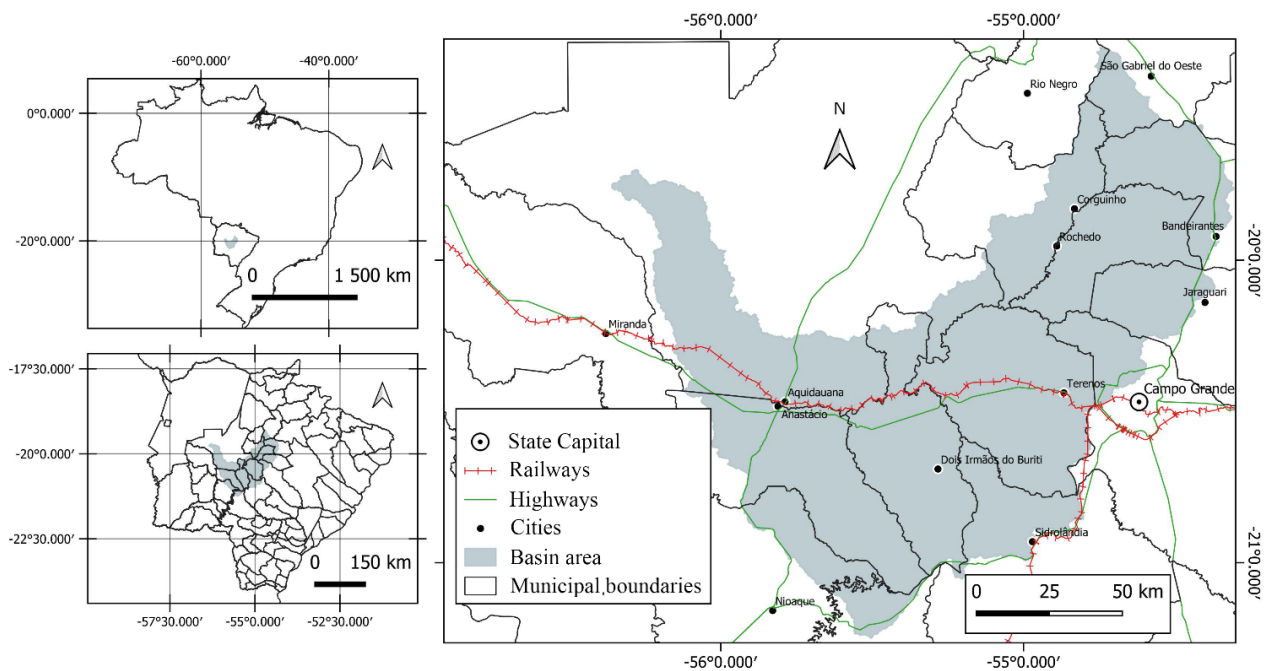


Figura 1: Mapa de localização da bacia hidrográfica do Rio Aquidauana.

A avaliação das estações pluviométricas que abrangem a área especificada foi realizada usando o banco de dados proveniente da AGÊNCIA NACIONAL DE ÁGUAS (ANA, 2022), e os resultados estão descritos na Tabela 1. A pesquisa identificou um total de 6 estações de monitoramento pluviométrico e 1 estação de monitoramento fluviométrico dentro da região definida.

Tabela 1 - Estações de Monitoramento Pluviométrico dentro da Área Delimitada.

ID	Nome	Localização (MS)	Estação de monitoramento
1 1956003	Entre Rios	Aquidauana	Precipitação
2 02055002	Palmeiras	Dois Irmãos do Buriti	Precipitação
3 02054009	Santa Elisa	Terenos	Precipitação
4 02054019	Jaraguari	Jaraguari	Precipitação
5 01954002	Rochedo	Rochedo	Precipitação
6 2155001	Nioaque	Nioaque	Precipitação
7 66950000	Porto Ciriaco	Aquidauana	VAZÃO

### 3.2 Rede de memória longa e de curto prazo (LSTM)

A rede LSTM pertence a uma classe distinta de redes neurais recorrentes (RNNs), superando as restrições das RNNs convencionais no aprendizado eficaz de dependências de longo prazo. Inicialmente proposto por Hochreiter e Schmidhuber (1997) e posteriormente refinado e popularizado por Kawakami (2008), o LSTM aproveita sua arquitetura de aprendizagem profunda para determinar o tempo de retenção e esquecimento de informações, realizado por meio de portas e células de memória projetadas propositalmente (Figura 2).



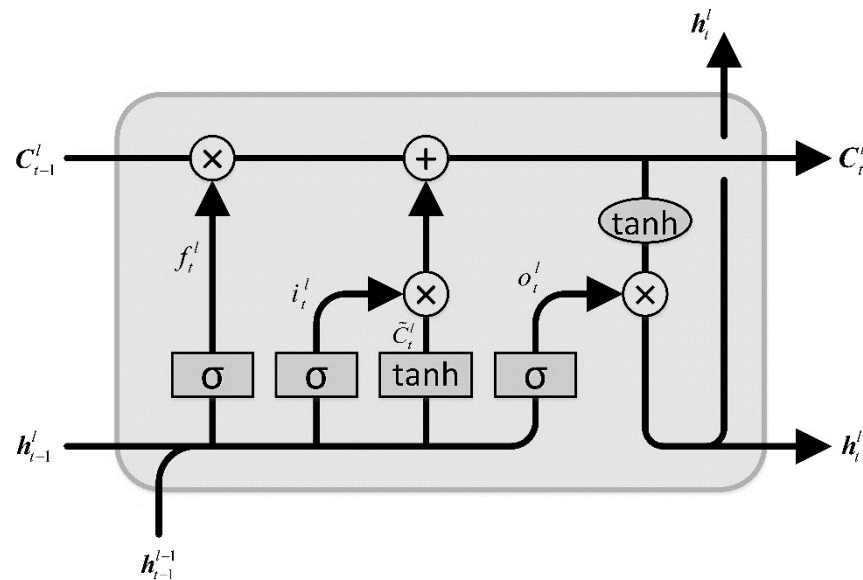


Figura 2 – Arquitetura da célula LSTM. Fonte: FAN et al., (2018).

A chave para o LSTM é o estado da célula ( $C_t$ ), que permite que as informações fluam inalteradas. A célula de memória LSTM é regulada por três portas que permitem opcionalmente a passagem de informações. A primeira porta é chamada de porta de esquecimento, que controla quais elementos do estado anterior da célula  $C_{t-1}$  serão esquecidos.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Onde  $f_t$  é um vetor de saída da camada sigmóide com valores variando de 0 a 1, indicando o grau de esquecimento.  $W_f$  e  $b_f$  definem o conjunto de parâmetros treináveis para a porta de esquecimento.

A seguir, a porta de entrada decide qual valor será atualizado:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

Onde é uma variável de saída com valor que varia de 0 a 1.  $W_i$  e  $b_i$  são parâmetros treináveis. A seguir, um vetor candidato para o estado da célula é calculado usando a entrada atual ( $x_t$ ) e o último estado oculto ( $h_{t-1}$ ):

$$\hat{C} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

Onde  $\hat{C}$  é um vetor com valores variando de 0 a 1,  $\tanh$  é a função tangente hiperbólica e  $W_c$  e  $b_c$  são parâmetros treináveis. Depois disso, você pode atualizar o antigo estado da célula  $C_{t-1}$  para o novo estado da célula  $C_t$  por multiplicação elemento a elemento:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \quad (4)$$

Finalmente, a porta de saída decide qual será a saída através de uma camada sigmóide:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

Desta forma,  $o_t$  é um vetor com valores que variam de 0 a 1.  $W_o$  e  $b_o$  são parâmetros treináveis definidos para a porta de saída. O novo estado oculto  $h_t$  é então calculado combinando as Equações 5 e 6:

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

### 3.3 Adjustments to the size of the LSTM window

The quantity of preceding time steps is denoted as the window size, and this parameter exerts significant influence on prediction accuracy, necessitating careful selection to optimize the model's performance. In the present investigation, the window size is set to 5.

### 3.4 Ajustes no tamanho da janela LSTM

As redes neurais abrangem convencionalmente vários hiperparâmetros, que são predefinidos antes do início do processo de aprendizagem. A otimização ou ajuste desses hiperparâmetros envolve a busca por um conjunto específico de valores que produza um modelo que minimize a função de perda nos dados fornecidos (GOODFELLOW; BENGIO; COURVILLE, 2016). Neste estudo, o erro quadrático médio (MSE) é empregado como função de perda para otimização de hiperparâmetros, seguindo o trabalho de Kratzert et al. (2018) e Fan et al. (2020).

Os hiperparâmetros frequentemente encontrados incluem a taxa de aprendizagem, o número de épocas de treinamento, a dimensionalidade do espaço de saída, entre outros (GOODFELLOW; BENGIO; COURVILLE, 2016). A taxa de aprendizagem é um hiperparâmetro que significa o tamanho do passo em um método gradiente descendente (ZEILER, 2012). Neste estudo, o otimizador Adam foi empregado como método de otimização estocástica (KINGMA; BA, 2017), com taxa de aprendizado inicial definida em 0,2. Além disso, uma taxa de decaimento baseada no tempo foi aplicada para atualizar a taxa de aprendizagem durante o processo de treinamento.

Além disso, o número de épocas, normalmente denotando uma passagem completa por todo o conjunto de dados dentro da rede neural, é empregado para particionar o treinamento em fases discretas. O treinamento prolongado pode resultar em overfitting, onde o modelo aprende padrões exclusivos do conjunto de dados de treinamento (FAN et al., 2020). Por outro lado, treinamento insuficiente pode levar ao underfitting, significando que o modelo não consegue capturar padrões relevantes nos dados de treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016).

Neste estudo, o número de épocas de treinamento será definido em 30, seguindo recomendação de Kratzert et al. (2018).

### 3.5 Validação do modelo LSTM

As métricas utilizadas para avaliar o desempenho do modelo são o Coeficiente de Determinação ( $R^2$ ) e o Root Mean Squared Error (RMSE). O cálculo de  $R^2$  é feito usando a equação:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (7)$$

O cálculo do RMSE (Root Mean Squared Error) é feito usando o seguinte procedimento:



$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n}} \quad (8)$$

Onde:

- $y_i$  é o vazão observado no tempo  $i$ .
- $\hat{y}$  é o vazão simulado no tempo  $i$ .
- $\bar{y}$  é o valor médio dos dados de vazão observados (média de  $y_i$ ).
- $n$  é o número de pontos de dados no conjunto de dados.

## 4 RESULTADOS E DISCUSSÃO

Nesta seção, apresentamos os resultados de nosso estudo, que se concentrou em aproveitar as capacidades da rede Long Short-Term Memory (LSTM) para desvendar os padrões subjacentes em dados de séries temporais históricas, particularmente no contexto da bacia do rio Aquidauana. Nossa metodologia envolveu a utilização de dados diários de precipitação, denotados por  $y(t)$ , como variável de entrada, e dados de vazão, representados por  $x(t)$ , como variável de saída alvo. O escopo temporal do nosso estudo abrangeu o extenso período entre 1º de janeiro de 1999 e 27 de dezembro de 2019, produzindo um conjunto de dados substancial compreendendo 7.670 amostras de dados.

### 4.1 Processamento dos dados

A normalização surgiu como uma etapa fundamental de pré-processamento destinada a padronizar os dados em uma escala comum. Os dados de vazão passaram por normalização, o que envolveu redimensioná-los para um intervalo entre 0 e 1 ou transformá-los para ter média zero e desvio padrão um. Este processo de normalização desempenhou um papel fundamental na estabilização do treinamento de nossa rede LSTM, especialmente quando se trata de variáveis que apresentam diferenças significativas em magnitudes. Além disso, dividimos o conjunto de dados em subconjuntos de treinamento e validação, adotando uma proporção de divisão 75/25, onde 75% dos dados normalizados foram reservados para treinamento e os 25% restantes para validação (a Tabela 2 fornece uma visão geral desse particionamento).

**Tabela 2 – Divisão de dados para treinamento e teste.**

Dados	Amostras	Porcentagem (%)
Treno	5.752	75
Teste	1.918	25
Total	7.670	100

Seguindo as etapas de pré-processamento de dados, nossa rede LSTM passou pelo processo de treinamento, que foi avaliado monitorando o erro quadrático médio (MSE) para os conjuntos de dados de treinamento e validação (Figura 3). Os dados plotados indicaram que o modelo convergiu para um valor MSE aproximadamente constante durante o treinamento, sem aumento significativo à medida que as épocas de treinamento progrediam. Esse comportamento consistente para os dados de treinamento e validação aponta para um processo de treinamento tranquilo, sem quaisquer anomalias notáveis. A estabilidade e os valores mínimos de MSE afirmam

fortemente a proficiência do modelo e sua capacidade de capturar efetivamente padrões de dados.

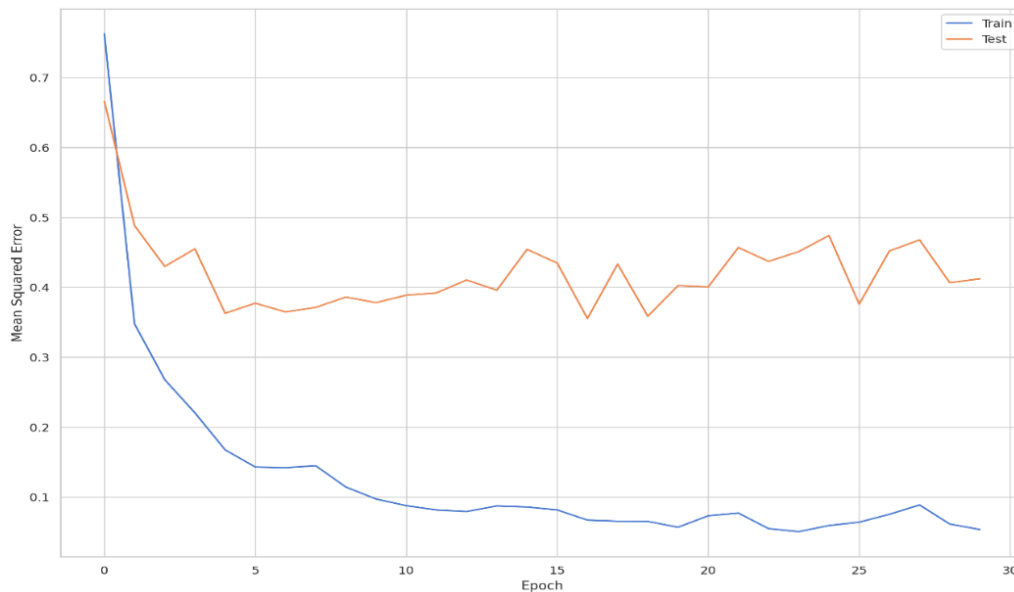


Figura 3 – Gráfico do número de épocas de treinamento.

## 4.2 Pré-processamento dos dados

Após a conclusão da fase de treinamento, avaliamos a eficácia do modelo na previsão de fluxo usando o conjunto de dados de validação. O processo de validação envolveu a geração de sequências de entrada-saída a partir do conjunto de dados, resultando em 1.918 amostras de validação. O mecanismo de previsão do modelo baseou-se nas sequências geradas, produzindo amostras de saída representando a variável alvo – neste caso, o fluxo do dia seguinte. Desta forma, o modelo foi alimentado de forma consistente com os dados originais, com o objetivo principal de prever a variável alvo para o dia seguinte.

A simulação representou previsões um passo à frente geradas pelo modelo treinado. As sequências de dados de entrada das estações pluviométricas foram alimentadas no modelo, possibilitando a geração da amostra de vazão subsequente, conforme ilustrado na Figura 4.

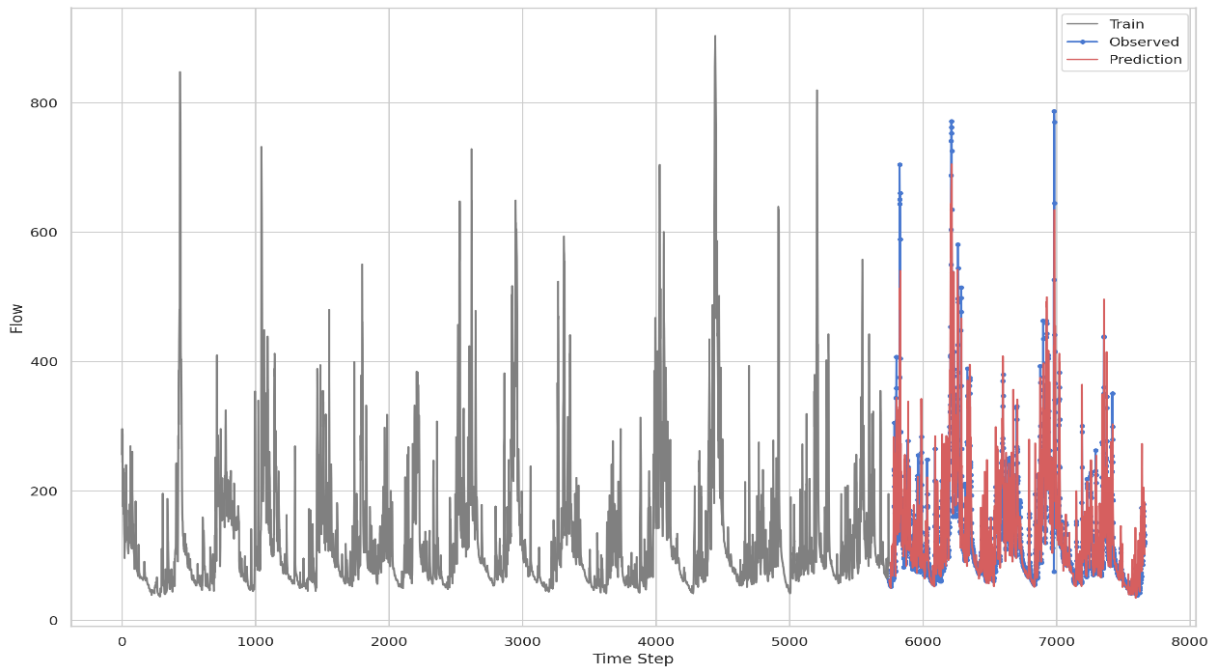


Figura 4 – Gráfico dos resultados do modelo LSTM na bacia de Aquidauana.

### 4.3 Performance do modelo

O desempenho do nosso modelo LSTM foi rigorosamente avaliado usando duas métricas principais: o coeficiente de determinação ( $R^2$ ) e a raiz do erro quadrático médio (RMSE).  $R^2$  é uma métrica comumente usada para avaliar o ajuste do modelo aos dados observados, enquanto o RMSE quantifica a precisão das previsões do modelo em comparação com os dados reais. Estas métricas capturam a proporção da variância total na variável dependente (fluxos simulados) que pode ser atribuída às variáveis independentes (dados de precipitação).

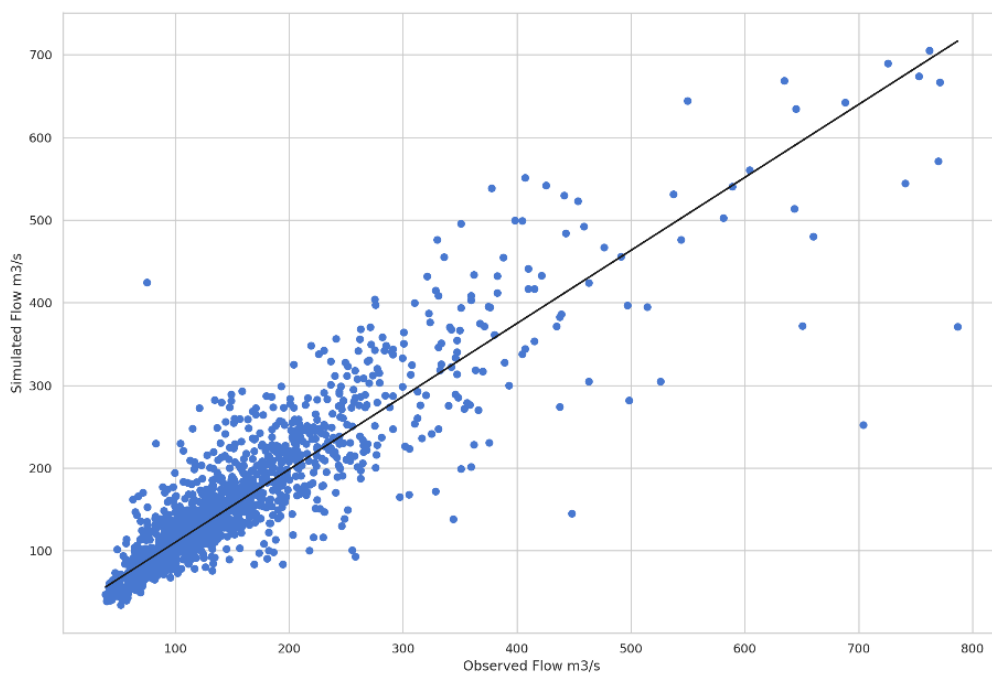


Figura 5 – Valor  $R^2$  de 0,82.

O valor de  $R^2$  de 0,82 significa que aproximadamente 82% da variabilidade observada nas vazões simuladas pode ser explicada pelas flutuações nos dados de precipitação utilizados como entrada para o modelo LSTM (Figura 5). Um valor de  $R^2$  mais elevado indica um ajuste mais favorável do modelo aos dados observados, indicando um alinhamento mais próximo entre as previsões do modelo e os valores reais. No entanto, é crucial considerar a aplicação e o contexto específicos ao interpretar o valor  $R^2$  com precisão.

Concluindo, nossos resultados demonstram a eficácia da rede LSTM na modelagem e previsão de vazão, oferecendo uma ferramenta valiosa para análise hidrológica na bacia do Rio Aquidauana. Esta pesquisa não só contribui para a compreensão das relações hidroclimáticas, mas também fornece uma estrutura robusta para a gestão sustentável dos recursos hídricos em bacias hidrográficas complexas. Mais pesquisas poderiam explorar arquiteturas alternativas de redes neurais e avaliar o desempenho do modelo sob várias condições climáticas para refinar e ampliar essas descobertas..

## 5 CONCLUSÃO

Este estudo aproveitou com sucesso o poder da rede neural recorrente Long Short-Term Memory (LSTM) para prever a vazão na bacia do rio Aquidauana, usando dados diários de precipitação como variáveis de entrada. Os resultados impressionantes que obtivemos, exemplificados por um valor  $R^2$  de 0,82 e um RMSE de 0,53, significam um alinhamento notável entre o nosso modelo LSTM e os dados de fluxo observados.

A principal vantagem do LSTM reside em sua capacidade de capturar dependências de longo prazo entre séries temporais de precipitação e vazões, permitindo previsões altamente precisas de vazões futuras. O desempenho robusto do modelo e sua capacidade de discernir e reter padrões temporais intrincados destacam sua adequação para uma série de tarefas de modelagem hidrológica.

O significado prático da nossa abordagem é sublinhado pelo seu papel crítico no fornecimento de previsões precisas de caudais, um componente crucial do planejamento e gestão sustentável dos recursos hídricos em bacias hidrográficas complexas. Ao integrar dados diários de precipitação com técnicas avançadas de aprendizagem automática, a nossa metodologia surge como um recurso valioso na melhoria da monitorização hidrológica e da tomada de decisões, particularmente face a condições hidroclimáticas complexas.

O significado prático da nossa abordagem é sublinhado pelo seu papel crítico no fornecimento de previsões precisas de caudais, um componente crucial do planejamento e gestão sustentável dos recursos hídricos em bacias hidrográficas complexas. Ao integrar dados diários de precipitação com técnicas avançadas de aprendizagem automática, a nossa metodologia surge como um recurso valioso na melhoria da monitorização hidrológica e da tomada de decisões, particularmente face a condições hidroclimáticas complexas. Redes LSTM para previsão de vazão usando dados diários de precipitação. Esperamos que nossas descobertas não apenas inspirem mais investigações acadêmicas, mas também encorajem a adoção prática desta metodologia para melhorar a gestão de recursos hídricos em um amplo espectro de bacias hidrográficas.

## 6 AGRADECIMENTOS

Este trabalho foi apoiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [números de bolsas 88887.701529/2022-00, 88887.497812/2020-00, 88887.701389/2022-00 e 88887.805051/2023-00].

## 7 REFERÊNCIAS

- Assine, M. L., Merino, E. R., Pupim, F. N., Warren, L. V., Guerreiro, R. L., & McGlue, M. M. (2016). Geology and Geomorphology of the Pantanal Basin. Em I. Bergier & M. L. Assine (Orgs.), *Dynamics of the Pantanal Wetland in South America* (p. 23–50). Springer International Publishing. [https://doi.org/10.1007/698\\_2015\\_349](https://doi.org/10.1007/698_2015_349)
- Back, Á. J., Souza, G. da S., Galatto, S. L., Corseuil, C. W., & Poletto, C. (2023). EROSIIVITY INDEX FOR BRASIL BASED ON CLIMATOLOGICAL NORMALS FROM 1991 TO 2020. *HOLOS*, 3(39), Artigo 39. <https://www2.ifrn.edu.br/ojs/index.php/HOLOS/article/view/16329>
- Couto, E., & Oliveira, V. (2010). The Soil Diversity of the Pantanal. Em *The Pantanal of Mato Grosso: Ecology, Biodiversity and Sustainable Management of a Large Neotropical Seasonal Wetland* (p. 71–102).
- Cui, Q., Ammar, M. E., Kariyeva, J., & Faramarzi, M. (2021). Regional wetland water storage changes: The influence of future climate on geographically isolated wetlands. *Ecological Indicators*, 120, 106941. <https://doi.org/10.1016/j.ecolind.2020.106941>
- Descovi, C. S., Zuffo, A. C., Mohammadzadeh, S. M., Murillo Bermudez, L. F., & Sierra, D. A. (2023, dezembro 29). *Modeling River Flows in the Brazilian Pantanal Basin Using Long Short-Term Memory (LSTM) Networks*. 7th SSS, Porto Alegre, Brazil.
- Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., & Sun, Z. (2021). Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy*, 220, 119708.
- Gavidia, J. C. R., Chinelatto, G. F., Basso, M., Souza, J. P., Soltanmohammadi, R., Vidal, A. C., Goldstein, R. H., & Mohammadzadeh, S. M. (2023). Utilizing integrated artificial intelligence for characterizing mineralogy and facies in a pre-salt carbonate reservoir, Santos Basin, Brazil, using cores, wireline logs, and multi-mineral petrophysical evaluation. *Geoenergy Science and Engineering*, 231, 212303. <https://doi.org/10.1016/j.geoen.2023.212303>
- Gavidia, J. C. R., Mohammadzadeh, S. M., Chinelatto, G. F., Basso, M., Souza, J. P. da P., Vidal, A. C., & Goldstein, R. H. (2023). Bridging the Gap: Integrating Static and Dynamic Data for Improved Permeability Modeling and Super K Zone Detection in Vuggy Reservoirs. *Journal of Geoenergy Science and Engineering*. In Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>



- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Benchmarking a catchment-aware long short-term memory network (LSTM) for large-scale hydrological modeling. *Hydrol. Earth Syst. Sci. Discuss*, 2019, 1–32.
- Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., Kumar Sahu, R., Greve, P., Slater, L., & Dadson, S. (2021). Hydrological concept formation inside long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences Discussions*, 2021, 1–37.
- Li, B., Li, R., Sun, T., Gong, A., Tian, F., Khan, M. Y. A., & Ni, G. (2023). Improving LSTM hydrological modeling with spatiotemporal deep learning and multi-task learning: A case study of three mountainous areas on the Tibetan Plateau. *Journal of Hydrology*, 620, 129401.
- Liu, D., Jiang, W., Mu, L., & Wang, S. (2020). Streamflow Prediction Using Deep Learning Neural Network: Case Study of Yangtze River. *IEEE Access*, 8, 90069–90086. <https://doi.org/10.1109/ACCESS.2020.2993874>
- Mohammadizadeh, S. M., Filho, J. G. D., Descovi, C. S., Murillo-Bermúdez, L. F., Sierra, D. A., & Perez, G. A. C. P. (2023, dezembro 29). *Cavitation Jet Apparatus for the Evaluation of Cavitation Erosion of Solid Surfaces*. 7th SSS, Porto Alegre, Brazil,.
- MOHAMMADIZADEH, S., Dalfré Filho, J. G., Sampaio Descovi, C., Murillo Bermúdez, L. F., Alfonso Sierra, D., & Corzo Perez, G. A. (2023). Assessment of cavitation erosion on solid surfaces using a cavitation jet apparatus. *HOLOS*, 5(39). <https://doi.org/10.15628/holos.2023.16306>
- Mohammadizadeh, S., Moghaddam, M. A., & Talebbeydokhti, N. (2021). Analysis of flow in porous media using combined pressurized-free surface network. *Journal of Porous Media*, 24(10). <http://dx.doi.org/10.1615/JPorMedia.2021025407>
- Murillo Bermudez, L. F., Abreu, A. E. S. de, Pereira, S. Y., Mohammadizadeh, S. M., Descovi, C. S., Sierra, D. A., & Martim, A. L. S. S. (2023, dezembro 29). *CONTAMINAÇÃO DAS ÁGUAS SUBTERRÂNEAS POR ESGOTO SANITÁRIO: COMO AVALIAR A FONTE DA CONTAMINAÇÃO?* 7th SSS, Porto Alegre, Brazil.
- Nikeghbali, P., Omidvar, P., Akbari, G., & Mohammadizadeh, S. M. (2014). THE STUDY ON SEDIMENT MOTION AND THE MODEL OF DAM BREAK ON THE MOVABLE BED BY SPH METHOD. *THE STUDY ON SEDIMENT MOTION AND THE MODEL OF DAM BREAK ON THE MOVABLE BED BY SPH METHOD*. INTERNATIONAL CONFERENCE ON COASTS, PORTS AND MARINE STRUCTURES (ICOPMAS). <https://sid.ir/paper/948560/en>
- Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica*, 67(5), 1471–1481.
- Sierra, D. A., Luvizotto Jr, E., Mohammadizadeh, S. M., Descovi, C. S., Murillo Bermudez, L. F., & Filho, J. G. D. (2023, dezembro 29). *TRANSITÓRIOS HIDRÁULICOS EM LINHAS DE RECALQUE DE ESGOTO - REVISÃO DA LITERATURA*. 7th SSS, Porto Alegre, Brazil.
- Xi, Y., Peng, S., Ciais, P., & Chen, Y. (2021). Future impacts of climate change on inland Ramsar wetlands. *Nature Climate Change*, 11(1), 45–51.



Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water resources research*, 56(1), e2019WR025326.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*, 31(7), 1235–1270.

#### **SOBRE OS AUTORES**

##### **C. S. DESCOVI**

Master in Civil Engineering Teaching, in the area of water, energy and environmental resources, from the State University of Campinas (Unicamp) and is currently pursuing a PhD at the same faculty. Research areas: hydrological modeling, hydrology, hydraulics.

E-mail: cassianodescovi@gmail.com

ORCID ID: <https://orcid.org/0000-0002-2847-2064>

##### **A.C. ZUFFO**

Master's degree in Civil Engineering - Concentration area in Hydraulic Engineering - University of São Paulo (EPUSP) and PhD in Hydraulic Engineering and Sanitation from EESC-USP. Professor at FEC-Unicamp. Post Doctorate from the University of Toronto - Ontario Canada.

E-mail: zuffoac@unicamp.br

ORCID ID: <https://orcid.org/0000-0002-2186-9755>

##### **S. MOHAMMADIZADEH**

Ph.D. candidate at School of Civil Engineering, Architecture and Urbanism (FECFAU), University of Campinas (UNICAMP), Brazil. He received a MS.c (2013) from University of Sistan and Baluchestan (USB), Iran, in the field of Civil Engineering – Hydraulic Structures. His main research interests are Cavitation, Porous Media, Computational Fluid Dynamics (CFD), Ultra High Performance Fibre Reinforced Concrete (UHPFRC), Petroleum Geology, as well as Hydrology.

E-mail: mohammadizadeh@fec.unicamp.br

ORCID ID: <https://orcid.org/0000-0001-6995-066X>

##### **L. F. M. BERMÚDEZ**

MSc and PhD student in Civil Engineering Education, area of water, energy and environmental resources, from the Campinas State University (Unicamp).

E-mail: luismurillo@fec.unicamp.br

ORCID ID: <https://orcid.org/0000-0001-7564-487X>

##### **D. A. SIERRA**

MSc in Civil Engineering Education, area of water, energy and environmental resources, from the Campinas State University (Unicamp) and PhD student at the same faculty. Research areas: hydraulic transients, bio-inspired algorithms, transient mitigation devices.

E-mail: d192229@dac.unicamp.br

ORCID ID: <https://orcid.org/0000-0003-0699-4512>

**Editor:** Francinaide de Lima Silva Nascimento

**Ad Hoc Reviewer:** Fernando Fan e Fernando Neves Lima



Submitted June 5, 2023  
Accepted September 11, 2023  
Published December 18, 2023

