

UNIVERSIDADE FEDERAL DE PELOTAS
Instituto de Física e Matemática
Programa de Pós-Graduação em Modelagem Matemática



Dissertação

Desenvolvimento de uma infraestrutura automatizada para aquisição e integração de dados hidrológicos ao sistema de armazenamento HEC-DSSVue, aplicada à previsão do nível máximo diário por meio de redes neurais Long Short-Term Memory (LSTM)

Carlos Henrique Moraes Praia

Pelotas, 2026

Carlos Henrique Moraes Praia

Desenvolvimento de uma infraestrutura automatizada para aquisição e integração de dados hidrológicos ao sistema de armazenamento HEC-DSSVue, aplicada à previsão do nível máximo diário por meio de redes neurais Long Short-Term Memory (LSTM)

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Matemática da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Modelagem Matemática

Orientador: Prof. Dra. Daniela Buske
Coorientadores: Prof. Dr. Samuel Beskow
Prof. Dr. Claus Haetinger

Pelotas, 2026

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação da Publicação

P895d Praia, Carlos Henrique Moraes

Desenvolvimento de uma infraestrutura automatizada para aquisição e integração de dados hidrológicos ao sistema de armazenamento HEC-DSSVue, aplicada à previsão do nível máximo diário por meio de redes neurais Long Short-Term Memory (LSTM) [recurso eletrônico] / Carlos Henrique Moraes Praia ; Daniela Buske, orientadora ; Samuel Beskow, Claus Haetinger, coorientadores. — Pelotas, 2026.

74 f.

Dissertação (Mestrado) — Programa de Pós-Graduação em Modelagem Matemática, Instituto de Física e Matemática, Universidade Federal de Pelotas, 2026.

1. Automatização de dados. 2. Séries temporais. 3. Redes neurais LSTM. I. Buske, Daniela, orient. II. Beskow, Samuel, coorient. III. Haetinger, Claus, coorient. IV. Título.

CDD 519.55

**DESENVOLVIMENTO DE INFRAESTRUTURA AUTOMATIZADA PARA
AQUISIÇÃO E INTEGRAÇÃO DE DADOS HIDROMETEOROLÓGICOS AO HEC-
DSSVUE, VISANDO À APLICAÇÃO DO MODELO LONG SHORT-TERM MEMORY
(LSTM) NA PREVISÃO DO NÍVEL MÁXIMO DIÁRIO**

por

Carlos Henrique Moraes Praia

Dissertação submetida ao Programa de Pós-Graduação em Modelagem Matemática, PPGMMat, do Instituto de Física e Matemática da Universidade Federal de Pelotas, como parte dos requisitos necessários para a obtenção do Título de

Mestre em Modelagem Matemática

Banca Examinadora:

Prof. Dr. Daniela Buske- Orientadora (UFPel)

Prof. Dr. Samuel Beskow- Co-Orientador (UFPel)

Prof. Dr. Claus Haetinger- Co-Orientador (UFPel)

Prof. Dr. Guilherme Jahnecke Weymar (FURG)

Prof. Dr. Diuliana Leandro (UFPel)

Profa. Dra. Douglas da Silva Lindemann (UFPel)

Profa. Dra. Aryane Araujo Rodrigues (UFPel)

Pelotas, 25 de fevereiro de 2026

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por me guiar e fortalecer espiritualmente ao longo do mestrado, pois, sem a sua presença em minha vida, não teria sido possível concluir essa etapa; nos momentos difíceis, Ele sempre me amparou. Além disso, agradeço, de forma especial, às minhas duas mães, Miracele de Oliveira Moraes e Miecylene Oliveira Moraes, pelo apoio incondicional, tanto emocional quanto financeiro, e, sobretudo, pelos ensinamentos que me motivaram a nunca desistir dos meus sonhos. São duas mulheres que me prepararam para os obstáculos da vida e me ensinaram a enfrentá-los. Por fim, deixo meus sinceros agradecimentos ao Programa de Pós-Graduação em Modelagem Matemática, por ter me proporcionado a oportunidade de realizar o mestrado, e à CAPES, pelo apoio financeiro concedido, fundamental em vários momentos até a sua conclusão.

RESUMO

MORAES PRAIA, Carlos Henrique. **Desenvolvimento de uma infraestrutura automatizada para aquisição e integração de dados hidrológicos ao sistema de armazenamento HEC-DSSVue, aplicada à previsão do nível máximo diário por meio de redes neurais Long Short-Term Memory (LSTM)**. 2026. 71 f. Dissertação (Mestrado em Modelagem Matemática) – Programa de Pós-Graduação em Modelagem Matemática, Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas, 2026.

O aumento da frequência e da intensidade de eventos hidrometeorológicos extremos, associado às mudanças climáticas, tem ampliado a vulnerabilidade socioambiental e evidenciado a necessidade de sistemas eficientes de monitoramento e previsão hidrológica. Nesse contexto, esta dissertação tem como objetivo desenvolver uma infraestrutura automatizada para aquisição, integração e gerenciamento de dados hidrometeorológicos, visando à aplicação de um modelo de *Deep Learning* do tipo *Long Short-Term Memory* (LSTM) na previsão do nível máximo diário com um dia de antecedência. O trabalho foi organizado em duas etapas. Na primeira etapa, foram desenvolvidos três algoritmos responsáveis pela aquisição automatizada de dados e pela integração dessas informações ao sistema de armazenamento de séries temporais HEC-DSSVue. Esses algoritmos geram três arquivos binários distintos: o primeiro contendo dados provenientes da Agência Nacional de Águas e Saneamento Básico (ANA), o segundo contendo dados do Centro Nacional de Monitoramento e Alerta de Desastres Naturais (CEMADEN) e o terceiro contendo dados do Instituto Nacional de Meteorologia (INMET). Na segunda etapa, utilizou-se o arquivo gerado a partir dos dados da ANA para a consolidação da base de dados destinada à aplicação do modelo LSTM. O modelo foi treinado, validado e testado com dados históricos da estação selecionada. No conjunto de teste, obteve-se RMSE de 0,37 m, MAE de 0,20 m e coeficiente de eficiência de Nash–Sutcliffe (NSE) igual a 0,9067, indicando elevada capacidade preditiva e boa aderência entre os valores previstos e observados. Os resultados demonstram que o modelo foi capaz de capturar adequadamente as dependências temporais da série, apresentando desempenho consistente para previsão de curto prazo. Assim, o trabalho contribui para a avaliação do uso de modelos baseados em *Deep Learning* (DL) em previsões hidrológicas, bem como para o desenvolvimento de uma infraestrutura automatizada de integração e armazenamento de dados hidrometeorológicos no HEC-DSSVue, estabelecendo uma base consistente para aplicações futuras em sistemas de monitoramento e previsão hidrológica em tempo quase real.

Palavras-chave: Automatização de dados, Séries Temporais, Redes Neurais LSTM.

ABSTRACT

MORAES PRAIA, Carlos Henrique. **Desenvolvimento de uma infraestrutura automatizada para aquisição e integração de dados hidrológicos ao sistema de armazenamento HEC-DSSVue, aplicada à previsão do nível máximo diário por meio de redes neurais Long Short-Term Memory (LSTM)**. 2026. 71 f. Dissertação (Mestrado em Modelagem Matemática) – Programa de Pós-Graduação em Modelagem Matemática, Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas, 2026.

The increase in the frequency and intensity of extreme hydrometeorological events, associated with climate change, has intensified socio-environmental vulnerability and highlighted the need for efficient hydrological monitoring and forecasting systems. In this context, this dissertation aims to develop an automated infrastructure for the acquisition, integration, and management of hydrometeorological data, with the purpose of applying a *Deep Learning* model of the *Long Short-Term Memory* (LSTM) type to forecast the daily maximum water level one day in advance. The work was structured in two stages. In the first stage, three algorithms were developed for automated data acquisition and for integrating this information into the HEC-DSSVue time series storage system. These algorithms generate three distinct binary files: the first containing data from the Brazilian National Water and Sanitation Agency (ANA), the second containing data from the National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), and the third containing data from the National Institute of Meteorology (INMET). In the second stage, the file generated from ANA data was used to consolidate the dataset for the application of the LSTM model. The model was trained, validated, and tested using historical data from the selected station. In the test set, the model achieved an RMSE of 0.37 m, an MAE of 0.20 m, and a Nash–Sutcliffe Efficiency (NSE) coefficient of 0.9067, indicating high predictive capability and strong agreement between predicted and observed values. The results demonstrate that the model was able to adequately capture the temporal dependencies of the time series, presenting consistent performance for short-term forecasting. Thus, this work contributes to the evaluation of the use of *Deep Learning* (DL) models in hydrological forecasting, as well as to the development of an automated infrastructure for the integration and storage of hydrometeorological data in HEC-DSSVue, establishing a consistent foundation for future applications in near real-time hydrological monitoring and forecasting systems.

Keywords: Data Automation, Time Series, LSTM Neural Networks.

LISTA DE FIGURAS

Figura 1	Estrutura de uma requisição HTTP.	20
Figura 2	Esquema de comunicação via <i>web service</i> entre sistemas e banco de dados.	21
Figura 3	Interação entre o usuário, o SGAA e o SWS para autenticação e acesso aos dados.	22
Figura 4	Arquitetura básica de uma RNN	28
Figura 5	Estrutura geral da arquitetura LSTM ao longo dos passos temporais	30
Figura 6	Estrutura interna de um bloco de memória da LSTM	30
Figura 7	Distribuição das estações hidrometeorológicas utilizadas da BHRJ.	34
Figura 8	Bloco principal do algoritmo	38
Figura 9	Processo de aquisição e tratamento dos dados	40
Figura 10	Processo de armazenamento dos dados no HEC-DSSVue	41
Figura 11	Estrutura lógica de controle do algoritmo	42
Figura 12	BHRJ com a estação selecionada	43
Figura 13	Catálogo do conjunto de dados da estação 88260000	52
Figura 14	Plotagem da variável <i>PRECIP-INC</i>	52
Figura 15	Plotagem da variável <i>STAGE</i>	53
Figura 16	Catálogo da estação 430130501A	54
Figura 17	Catálogo do conjunto de dados das estações A827 e A836	54
Figura 18	Histórico de perda no treino e validação	57
Figura 19	Comparação entre os valores observados e previstos do nível máximo diário (conjunto de treinamento)	58
Figura 20	Comparação entre os valores observados e previstos do nível máximo diário (conjunto de validação)	58
Figura 21	Comparação entre os valores observados e previstos do nível máximo diário (conjunto de teste)	59

LISTA DE TABELAS

Tabela 1	Resumo das séries temporais diárias utilizadas no estudo	45
Tabela 2	Arquitetura LSTM e configuração	48
Tabela 3	Estações utilizadas para teste de aplicação dos algoritmos	50
Tabela 4	Síntese comparativa das métricas de erro nos conjuntos de treino, validação e teste	56

LISTA DE ABREVIATURAS E SIGLAS

ANA	Agência Nacional de Águas e Saneamento Básico
APIs	<i>Application Programming Interfaces</i>
BDMEP	Banco de Dados Meteorológicos para Ensino e Pesquisa
CEMADEN	Centro Nacional de Monitoramento e Alerta de Desastres Naturais
DL	<i>Deep Learning</i>
DLCM	<i>Discrete Linear Cascade Model</i>
GFS	<i>Global Forecast System</i>
GRU	<i>Gated Recurrent Unit</i>
HEC-HMS	<i>Hydrologic Modeling System</i>
HEC-RAS	<i>River Analysis System</i>
HTTP	<i>Hypertext Transfer Protocol</i>
INMET	Instituto Nacional de Meteorologia
IPCC	Painel Intergovernamental sobre Mudanças Climáticas
LSTM	<i>Long Short-Term Memory</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
ML	<i>Machine Learning</i>
MLP	<i>Multi-Layer Perceptron</i>
NSE	<i>Nash–Sutcliffe Efficiency</i>
PMF	<i>Probable Maximum Flood</i>
PMP	<i>Probable Maximum Precipitation</i>
R ²	Coeficiente de Determinação
RF	<i>Random Forest</i>
RMSE	<i>Root Mean Squared Error</i>
RNN	<i>Recurrent Neural Network</i>
ReqSim	<i>Requisitely Simple Flood Forecast System</i>

SARIMA	<i>Seasonal Autoregressive Integrated Moving Average</i>
SDM	<i>Scaled Distribution Mapping</i>
SGAA	Sistema de Gerenciamento de Autenticação e Autorização
SNIRH	Sistema Nacional de Informações sobre Recursos Hídricos
SWS	Sistema de Web Services
TRMM	<i>Tropical Rainfall Measuring Mission</i>
URL	<i>Uniform Resource Locator</i>
WI	<i>Willmott Index</i>
WRF	<i>Weather Research and Forecasting Model</i>
WSFS	<i>Watershed Simulation and Forecasting System</i>
XML	<i>Extensible Markup Language</i>
XGBoost	<i>Extreme Gradient Boosting</i>

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Objetivo geral	12
1.2	Objetivos específicos	12
2	REFERENCIAL TEÓRICO	14
2.1	Mudanças climáticas	14
2.2	Utilização de Sistemas de Alerta Antecipado	17
2.3	Dados Hidrometeorológicos: Aplicações e Disponibilização no Brasil	18
2.4	Soluções de Acesso aos Dados Meteorológicos: APIs e Webservices	19
2.5	Sistema HEC-DSSVue para Gerenciamento de Séries Temporais	23
2.6	Aplicações de Deep Learning (DL) na Previsão de Níveis de Rios	24
2.7	Origem da Arquitetura LSTM	28
2.8	Estrutura Básica da arquitetura LSTM	29
3	METODOLOGIA	33
3.1	Estrutura Geral do Trabalho	33
3.2	Área de Estudo	34
3.3	Etapa de Desenvolvimento dos Algoritmos	35
3.3.1	Algoritmo de integração com ServiceANA	36
3.3.2	Algoritmo de integração com a API do CEMADEN	38
3.3.3	Algoritmo de integração com a API do INMET	41
3.4	Consolidação da Base de Dados	43
3.5	Pré-processamento das Séries	43
3.6	Desenvolvimento do modelo LSTM	45
4	RESULTADOS E DISCUSSÃO	50
5	CONCLUSÃO	61
6	PERSPECTIVAS FUTURAS	63
	REFERÊNCIAS	64

1 INTRODUÇÃO

Globalmente, diversos setores da sociedade vêm sendo prejudicados pela frequência e intensidade de eventos extremos, como ondas de calor, tempestades severas e outros, resultantes das mudanças climáticas fortemente relacionadas às atividades humanas. De acordo com IPCC (2023), essas ações estão provocando a alteração do clima da Terra, o que tem contribuído para o aparecimento mais frequente desses eventos.

Chuvvas extremas se tornam mais comuns em algumas regiões, causando prejuízos econômicos e perda de vidas. A chuva extrema tornou-se um risco ambiental severo, causando não somente a inundação urbana, mas também desastres naturais graves, como por exemplo, enchentes repentinas, deslizamentos de terra e fluxos de lama. Esse evento extremo representa uma grave ameaça à produção agrícola, à estabilidade socioeconômica e aos ecossistemas (Lesk et al., 2016; Gründemann et al., 2023; Ombadi et al., 2023).

As inundações são consideradas um dos desastres naturais mais frequentes e severos em escala global, causando prejuízos progressivamente mais expressivos. Esse agravamento está relacionado ao aumento da ocorrência de eventos extremos, às mudanças nos regimes de precipitação e temperatura e às modificações na cobertura e no uso do solo, fatores que intensificam a vulnerabilidade socioambiental e ampliam os danos associados a esses desastres (Souza et al., 2025).

No Brasil, especificamente no ano de 2024, entre o final do mês de abril e o início de maio, o estado do Rio Grande do Sul, dando destaque para a área central do estado, foi atingido por eventos chuvosos extremos. A intensidade desses eventos provocou inúmeros desastres naturais severos, como inundações e deslizamentos de terra, causando mortes de moradores e danos à infraestrutura de cidades (Ribeiro, 2025).

Nesse contexto de crescente vulnerabilidade, associado ao aumento da frequência de eventos de inundação, a previsão do nível dos rios torna-se essencial para a emissão de alertas relacionados a esse tipo de desastre. Conforme destacado por Vizi et al. (2023), a previsão confiável dos níveis fluviais desempenha papel funda-

mental no suporte às ações de controle de cheias, na segurança da navegação e no planejamento da captação de recursos hídricos.

Assim, surgem diferentes abordagens, sejam modelos baseados em processos físicos ou modelos baseados em dados. De modo geral, a estimativa do nível da água pode ser realizada por meio de abordagens fundamentadas em princípios físicos, como os modelos numéricos, os quais têm sido amplamente empregados sobretudo em simulações bidimensionais (2D) e tridimensionais (3D) para estimar nível de água (Olsen et al., 1998; Liao et al., 2014; Reinking, 2016)

Porém, esses modelos apresentam alta complexidade, de acordo com Kim et al. (2022) essas abordagens fundamentadas em princípios físicos demandam procedimentos computacionais mais elaborados e a definição de diversos parâmetros. Essa complexidade pode introduzir maiores níveis de incerteza nas estimativas, além de ampliar o tempo necessário para a previsão do nível de água.

Por outro lado, os modelos baseados em dados estão sendo cada vez mais utilizados. Segundo Hieu et al. (2023), uma estratégia distinta baseia-se na aplicação desses modelos, cujo objetivo é identificar e explorar padrões e relações estatísticas entre variáveis de entrada e saída. Essa abordagem tem potencial para superar as restrições anteriormente apontadas nos modelos fundamentados exclusivamente em representações físicas dos processos.

Além disso, segundo Adamowski; Sun (2010), esses modelos concentram-se na extração e análise de padrões a partir das informações disponíveis, permitindo examinar extensos conjuntos de dados hidrológicos sob diferentes enfoques, sem a necessidade de representar explicitamente os processos físicos envolvidos.

Corroborando as vantagens desses modelos em relação às técnicas tradicionais, no estudo de Dai et al. (2023) foi realizada a comparação de diversos algoritmos baseados em dados para a previsão do nível de cheia em bacias hidrográficas. Nesse estudo, demonstrou-se que, especificamente, uma rede neural do tipo *Long Short-Term Memory* (LSTM) obteve resultados superiores quando comparada a abordagens como ARIMA, *Random Forest*, *Linear Regression* e redes neurais convencionais treinadas por *backpropagation*.

Nesse contexto, esta dissertação propõe e avalia a aplicação de um modelo de *Deep Learning* (DL), especificamente a rede neural LSTM, para a previsão, com um dia de antecedência, do nível máximo diário. Além disso, este trabalho vai além da simples aplicação do modelo, buscou-se desenvolver três algoritmos, cada um capaz de realizar a aquisição automatizada de dados e sua integração a um sistema de banco de dados.

Assim, a dissertação está dividida em duas etapas. A primeira concentra-se no desenvolvimento e na aplicação desses algoritmos, com o propósito de integrar bancos de dados públicos, visando à criação de um sistema que opere em tempo real e con-

solide as informações provenientes das diferentes bases. Essa etapa justifica-se pela necessidade de disponibilização de dados em tempo real em determinadas situações. A segunda etapa consiste na utilização de um desses algoritmos para organizar os dados da estação que servirá de base para a aplicação do modelo LSTM.

Ao longo do texto serão apresentados mais detalhes sobre cada etapa. Dessa forma, o trabalho está organizado da seguinte maneira: a seção de Fundamentação Teórica revisa alguns trabalhos relacionados e conceitos referentes à mudança climática, aos sistemas de previsão, à utilização de dados meteorológicos, à forma como esses dados estão disponíveis no Brasil, às formas de acesso a esses dados e ao sistema de gerenciamento de dados.

Além disso, nessa mesma seção serão discutidos também conceitos e trabalhos referentes a aplicações de DL para previsão de níveis de rios, a origem do modelo LSTM e a estrutura básica desse modelo. A próxima seção busca explicar os passos metodológicos da fase de desenvolvimento, tanto dos três algoritmos quanto do modelo proposto.

A seção seguinte detalha a aplicação do modelo proposto a uma estação hidrometeorológica, apresentando o ambiente experimental, o conjunto de dados utilizado e os resultados obtidos. Além disso, busca discutir o desempenho do modelo por meio das métricas adotadas. Por fim, a última seção apresenta as conclusões desta dissertação e aponta algumas direções para trabalhos futuros.

1.1 Objetivo geral

Desenvolver e avaliar uma infraestrutura automatizada para aquisição, integração e gerenciamento de dados hidrometeorológicos provenientes de diferentes instituições públicas brasileiras, com armazenamento estruturado no HEC-DSSVue, visando à aplicação de um modelo de DL do tipo LSTM para a previsão do nível máximo diário com um dia de antecedência.

1.2 Objetivos específicos

- Desenvolver algoritmos automatizados para a aquisição de dados hidrometeorológicos por meio de APIs e *web services* da ANA, CEMADEN e INMET.
- Estruturar e armazenar os dados consolidados em arquivos no formato DSS, utilizando o HEC-DSSVue como sistema de gerenciamento.
- Desenvolver e configurar um modelo LSTM para a previsão do nível máximo diário, com horizonte de um dia à frente.
- Realizar o pré-processamento das séries temporais, incluindo a organização dos

dados em conjuntos de treinamento, validação e teste.

- Comparar sistematicamente os valores previstos pelo modelo LSTM com os valores observados nos conjuntos de treinamento, validação e teste.
- Avaliar o desempenho do modelo por meio das métricas RMSE, MAE e NSE.

2 REFERENCIAL TEÓRICO

2.1 Mudanças climáticas

As transformações observadas no sistema climático da Terra têm sido amplamente discutidas pela sociedade, sobretudo em função de seus impactos sociais e econômicos. A mudança climática pode ser compreendida como alterações nos padrões climáticos, as quais representam um dos principais desafios ambientais e sociais da atualidade, uma vez que seus impactos afetam de forma significativa tanto os sistemas naturais quanto as comunidades humanas em escala global (Teng, 2024).

A compreensão das mudanças climáticas exige não apenas sua definição, mas também a análise dos fatores responsáveis pelo seu desequilíbrio. De forma análoga, pode-se imaginar esse desequilíbrio como um motor em funcionamento, o qual necessita de um combustível para ficar operando. Nesse sentido, conforme apontado por Alves et al. (2016), a emissão de Gases de Efeito Estufa (GEE) constitui um dos principais fatores responsáveis pela mudança climática.

Continuando a análise, o GEE tem origem tanto em processos naturais quanto em ações humanas. Referentes às fontes naturais destacam-se fenômenos como erupções vulcânicas, vulcões de lama, terremotos, incêndios naturais, o degelo do permafrost, áreas úmidas e os oceanos. Já as fontes não naturais associadas às atividades humanas abrangem os processos industriais, a geração de energia, a exploração florestal sem controle, bem como o uso da terra e as mudanças em seu uso, conforme sistematizado pelo (IPCC, 2014).

Historicamente, a produção do GEE era um processo equilibrado. De acordo com Adamo et al. (2021), durante a fase inicial da Terra, o clima passou por intensas variações provocadas por fatores como deslocamentos tectônicos, alterações orbitais, grandes erupções vulcânicas e a dinâmica interna do planeta. Esses elementos influenciaram diretamente a temperatura global, a circulação oceânica e atmosférica, além da formação de geleiras e desertos. A evolução biológica, iniciada com organismos fotossintetizantes há cerca de 3,5 bilhões de anos, também teve papel crucial ao modificar a composição da atmosfera. Ao longo dessas eras, as mudanças climáticas

moldaram o ambiente terrestre, favorecendo o surgimento e a extinção de diversas espécies.

No primeiro relatório publicado pelo painel intergovernamental sobre mudanças climáticas (IPCC), há uma preocupação de que as ações antrópicas possam estar indiretamente intensificando o efeito estufa, com isso o clima da terra passar por alterações, essa intensificação do efeito estufa ocorre devido às grandes emissões passadas e contínuas de gases, como dióxido de carbono e outros gases, causando o aumento da temperatura da superfície da Terra, popularmente chamado de ‘aquecimento global’. Caso isso ocorra, as mudanças consequentes poderão ter um impacto severo sobre a sociedade (IPCC, 1990).

Naquele período, as projeções do IPCC já alertavam para as consequências das ações humanas sobre o processo natural do efeito estufa. No sexto relatório, o IPCC afirma que as atividades humanas, em específico a emissão de gases de efeito estufa, são a causa inequívoca do aumento da temperatura global observado nas últimas décadas, com elevação média de 1,1 °C entre 2011 e 2020 em relação ao período de 1850 a 1900 (IPCC, 2023).

Como consequência, têm ocorrido mudanças aceleradas nos principais componentes do sistema natural da Terra: componentes atmosférico, oceânico, criosférico e biosférico. Além disso, essas mudanças não ocorrem de maneira isolada, afetam extremos climáticos e meteorológicos em várias regiões do planeta, causando consequências tanto para a natureza quanto para as pessoas e atingindo, de modo desproporcional, comunidades vulneráveis que contribuíram menos, historicamente, para a mudança climática (IPCC, 2023).

No contexto do Sul Global, essas mudanças se traduzem em um grave problema de crise hídrica, marcada por vulnerabilidades socioeconômicas acentuadas. Para Iyiola et al. (2024) a crise hídrica configura-se como uma das manifestações mais alarmantes das mudanças climáticas, especialmente no Sul Global, onde as vulnerabilidades socioeconômicas acentuam os impactos ambientais. A elevação das temperaturas médias e a alteração nos regimes de precipitação comprometem a disponibilidade e a distribuição da água, afetando diretamente a oferta para consumo humano, produção agrícola, geração de energia e manutenção dos ecossistemas aquáticos.

Em regiões onde as estratégias de adaptação são limitadas e os investimentos em infraestrutura hídrica são escassos, os efeitos tornam-se ainda mais severos, resultando em escassez, conflitos pelo uso da água e comprometimento da segurança hídrica. Além disso, práticas insustentáveis como o desmatamento, o uso intensivo do solo e a poluição de corpos d’água intensificam a degradação dos recursos hídricos. Diante desse cenário, torna-se urgente a implementação de políticas públicas integradas e estratégias de gestão sustentável da água, em consonância com os Objetivos de Desenvolvimento Sustentável, como forma de mitigar os efeitos da crise hídrica e

garantir a resiliência das comunidades mais afetadas (Iyiola et al., 2024).

O Brasil, apesar de ser reconhecido como um país abundante em recursos hídricos, teve essa percepção como fator que levou a sociedade a não se preocupar com o seu mau uso. Isso acarretou uma cultura de desperdício e desvalorização desse recurso. No entanto, no cenário recente, é mais evidente que essa abundância não é mais absoluta. Observa-se o surgimento de uma nova realidade, marcada por eventos extremos mais frequentes e intensos, como longos períodos de estiagem, enchentes severas e alterações nos padrões de precipitação (Barros; Silva, 2023).

Um exemplo recente e marcante dessa realidade ocorreu no Rio Grande do Sul, em 2024. Segundo Marengo et al. (2024) desde o final de abril de 2024, o Rio Grande do Sul enfrenta temporais intensos, agravados no início de maio, que provocaram inundações generalizadas. Esses eventos afetaram grande parte do estado, incluindo Porto Alegre.

Em 1941 a região já havia testemunhado o evento, mas nesse período ele foi menos intenso. As chuvas intensas e persistentes provocaram a elevação dos níveis dos rios Taquari, Caí, dos Sinos e das Antas, inundando diversas cidades. O acúmulo das águas formou uma onda que fez o nível do Lago Guaíba subir mais de 5 metros, causando inundações em vários bairros de Porto Alegre. Essa cheia superou o recorde de 1941, quando o Guaíba atingiu 4,76 m, alcançando 5,35 m no dia 5 de maio de 2024 (Marengo et al., 2024)

Exemplos recentes, como o ocorrido no Rio Grande do Sul, e fenômenos similares observados em diversas regiões do planeta evidenciam que esses riscos são reais e tendem a se intensificar. Em setembro de 2025, uma tempestade chamada Daniel desencadeou chuvas recorde em países como Grécia, Turquia, Líbia e Bulgária, resultando em numerosas mortes e na destruição de infraestruturas, tornando-se o ciclone com maior impacto econômico já registrado fora do Atlântico Norte (Ampas et al., 2025).

Nesse sentido, investir em sistemas de previsão hidrológica é essencial para antecipar riscos futuros e orientar respostas rápidas e eficientes, fornecendo informações confiáveis para os órgãos responsáveis pela tomada de decisão. De acordo com Saito et al. (2013), dentre as ações não estruturais para a prevenção desses cenários, destacam-se os sistemas de monitoramento e alerta. Essas ferramentas são importantes, pois permitem identificar riscos antecipadamente, promovendo a redução de danos na região. Além disso, a relevância dessas ações não se limita ao aspecto econômico; elas também evitam que o impacto do desastre comprometa o desenvolvimento social e a recuperação das comunidades atingidas.

2.2 Utilização de Sistemas de Alerta Antecipado

Os sistemas de alerta antecipado constituem ferramentas estratégicas na gestão de riscos hidrológicos, pois permitem transformar informações meteorológicas e hidrológicas em ações preventivas. Segundo Golding (2022), esses sistemas possibilitam a compreensão dinâmica dos desastres naturais, fornecendo suporte à tomada de decisão em tempo hábil e reduzindo perdas humanas e materiais.

Experiências internacionais demonstram de forma concreta a efetividade desses sistemas. Na Oceania, como foi o caso do Ciclone Evan, em Samoa, foi mostrado que cada dólar investido nesses sistemas poderia gerar uma economia de até seis dólares em danos evitados. O estudo estimou que até 81,45 % das perdas poderiam ter sido reduzidas com um sistema eficiente, evidenciando a importância dessas ferramentas para a resiliência socioeconômica, ao salvar colheitas, proteger bens e reduzir impactos sociais (Fakhrudin et al., 2019).

Um exemplo relevante ocorreu na Índia, com a comparação entre os ciclones Kandla (1998) e Biparjoy (2023). Embora o primeiro tenha provocado mais de mil mortes, o segundo, mesmo sendo intenso, não registrou vítimas fatais. A diferença central esteve na evolução dos sistemas de monitoramento e previsão, que permitiram evacuações antecipadas e ações coordenadas de mitigação (Mohapatra et al., 2025). Esse caso evidencia como o avanço tecnológico e a melhoria na capacidade preditiva podem alterar drasticamente os impactos de eventos extremos.

Outro exemplo internacional refere-se à Alemanha, especificamente à enchente do Vale do Rio Ahr em 2021. No estudo de Najafi et al. (2024), foi analisada a aplicação de um sistema experimental de alerta baseado em impactos, capaz de gerar mapas de inundação em alta resolução e estimar danos potenciais à infraestrutura. O diferencial desse modelo foi a tradução de previsões hidrometeorológicas em informações operacionais diretamente utilizáveis pela proteção civil. Contudo, o estudo também apontou limitações associadas à incerteza meteorológica e à elevada demanda computacional.

No contexto nacional, destaca-se o Sistema de Alerta Hidrológico (SAH) da bacia do rio Uruguai, em operação desde 2018. Conforme Mattiuzi et al. (2023), o sistema realiza previsões de níveis fluviais para municípios do Rio Grande do Sul, emitindo boletins preventivos de cheia. A experiência brasileira demonstra que, mesmo em bacias com resposta hidrológica rápida e ocorrência de eventos ao longo de todo o ano, é possível estruturar sistemas operacionais que forneçam suporte à gestão de risco.

No Brasil, especificamente no estado do Ceará, o estudo de Rolim et al. (2025) utilizou séries históricas de precipitação provenientes de 20 estações pluviométricas. Nesse trabalho, foi realizada a integração da teoria do caos com técnicas de *Machine Learning* (ML) para prever a variabilidade da chuva e aplicar os resultados em siste-

mas de alerta precoce contra secas. A importância desses sistemas reside na possibilidade de mitigar impactos hídricos e socioeconômicos, além de auxiliar gestores na tomada de decisão e na alocação de recursos. Entretanto, os autores destacam limitações associadas ao comportamento caótico das séries temporais, à necessidade de calibração específica para cada estação e à dificuldade de realizar previsões de longo prazo.

De forma técnica, os estudos apresentados indicam que a eficácia dos sistemas de alerta antecipado depende, fundamentalmente, da qualidade e disponibilidade tempestiva de dados hidrometeorológicos, da utilização de modelos capazes de capturar a dinâmica temporal dos eventos extremos e da adequada integração entre a previsão técnica e os mecanismos de comunicação operacional. Nesse contexto, a construção de infraestruturas automatizadas para aquisição, tratamento e organização de dados, como a proposta nesta dissertação configura-se como etapa essencial para viabilizar sistemas de previsão hidrológica mais robustos, confiáveis e passíveis de operacionalização em tempo real.

2.3 Dados Hidrometeorológicos: Aplicações e Disponibilização no Brasil

No planejamento de risco e nos sistemas de alerta antecipado, os dados meteorológicos constituem a base essencial para o funcionamento dos modelos de previsão. No caso específico da previsão de inundações, modelos computacionais são alimentados por informações como precipitação, temperatura, umidade, pressão atmosférica, radiação solar e vento, entre outras variáveis. Assim, os dados representam o principal insumo dos modelos e, por consequência, elemento estruturante dos sistemas de alerta precoce.

Estudos apresentados nesta seção evidenciam que a acurácia dos modelos está fortemente associada à qualidade e à adequação dos dados utilizados. Em aplicações de ML, como modelos LSTM, observou-se que variáveis como precipitação e vazão defasada desempenham papel determinante no desempenho preditivo (Sushanth et al., 2023). Da mesma forma, em modelos baseados em processos físicos, como o *Hydrologic Modeling System* (HEC-HMS) e o *River Analysis System* (HEC-RAS), os dados meteorológicos são indispensáveis tanto para calibração e validação quanto para a geração de cenários de simulação (Tessema et al., 2024).

De maneira semelhante, Goodarzi et al. (2024) integraram o modelo meteorológico *Weather Research and Forecasting Model* (WRF) ao HEC-HMS, evidenciando que a qualidade e a distribuição espacial da precipitação influenciam diretamente a precisão da simulação hidrológica. Já Tran et al. (2025) destacam que dados horários de precipitação são essenciais para ajustes em tempo real em sistemas de previsão

operacional.

Em escala mais ampla, Palash et al. (2024) utilizaram precipitação observada por satélites, como o *Tropical Rainfall Measuring Mission* (TRMM) e o *Global Precipitation Measurement – Integrated Multi-Satellite Retrievals for GPM* (GPM/IMERG), além de previsões do *Global Forecast System* (GFS) e séries de vazão ou nível de água como base para aprimoramento da previsão operacional de cheias. Esses exemplos reforçam que os dados meteorológicos são determinantes tanto na fase de calibração quanto na operacionalização dos sistemas de alerta.

No contexto brasileiro, a organização e disponibilização dessas informações são realizadas principalmente por instituições públicas como a Agência Nacional de Águas e Saneamento Básico (ANA), o Centro Nacional de Monitoramento e Alerta de Desastres Naturais (CEMADEN) e o Instituto Nacional de Meteorologia (INMET). Conforme Silva (2025), essas instituições desempenham papel central no monitoramento, organização e disponibilização de dados hidrológicos e meteorológicos, garantindo padronização e acesso às informações.

A ANA atua no monitoramento de variáveis como precipitação, nível e vazão, organizadas no âmbito do Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH), conforme estabelecido pela Lei nº 9.433/1997 (Silva, 2025). O CEMADEN, criado em 2011, tem como foco o monitoramento voltado à prevenção de desastres naturais, utilizando redes observacionais compostas por pluviômetros automáticos, radares meteorológicos e sensores de umidade do solo (Bacelar et al., 2020). Já o INMET, fundado em 1909, é responsável pelo monitoramento meteorológico nacional, disponibilizando informações fundamentais para previsão do tempo e estudos climáticos (INMET, 2025).

Em síntese, a atuação dessas instituições assegura ao Brasil um amplo conjunto de dados meteorológicos e hidrológicos, essenciais tanto para a pesquisa científica quanto para a gestão de riscos. Além de medir e organizar essas informações, tais instituições garantem sua confiabilidade e padronização, possibilitando sua aplicação em diferentes contextos, especialmente nos sistemas de previsão hidrológica.

2.4 Soluções de Acesso aos Dados Meteorológicos: APIs e Web-services

Os portais institucionais da ANA, CEMADEN e INMET foram desenvolvidos para acesso por meio de interfaces gráficas voltadas à interação humana. Conforme discutido por Cooksey (2014), websites são estruturados para facilitar a interpretação visual e a navegação por usuários. Entretanto, quando há necessidade de aquisição recorrente ou simultânea de dados de múltiplas estações, o acesso manual torna-se ineficiente.

Nesse contexto, as *Application Programming Interfaces* (APIs) constituem uma solução tecnológica para integração automatizada entre sistemas. Segundo Perron et al. (2025), uma API é um conjunto de regras que permite a comunicação estruturada entre programas distintos. Essa integração ocorre no modelo cliente-servidor, no qual o servidor disponibiliza os recursos e o cliente realiza requisições para acessar os dados (Cooksey, 2014).

A comunicação entre cliente e servidor é regida por protocolos padronizados, que definem a estrutura e o formato das mensagens trocadas. O protocolo amplamente utilizado nesse modelo é o *Hypertext Transfer Protocol* (HTTP), baseado no ciclo requisição–resposta. Para que uma requisição HTTP seja válida, são necessários quatro elementos principais: *Uniform Resource Locator* (URL), *method*, *headers* e *body*, cada um desempenhando função específica na identificação do recurso, definição da ação e envio de informações complementares (Cooksey, 2014).

Nesse sentido, tendo todos os componentes exigidos no processo, é possível se comunicar com o servidor e obter respostas de acordo com as especificações feitas na solicitação. As quatro componentes que tornam a chamada HTTP válida são ilustradas na Figura 1:

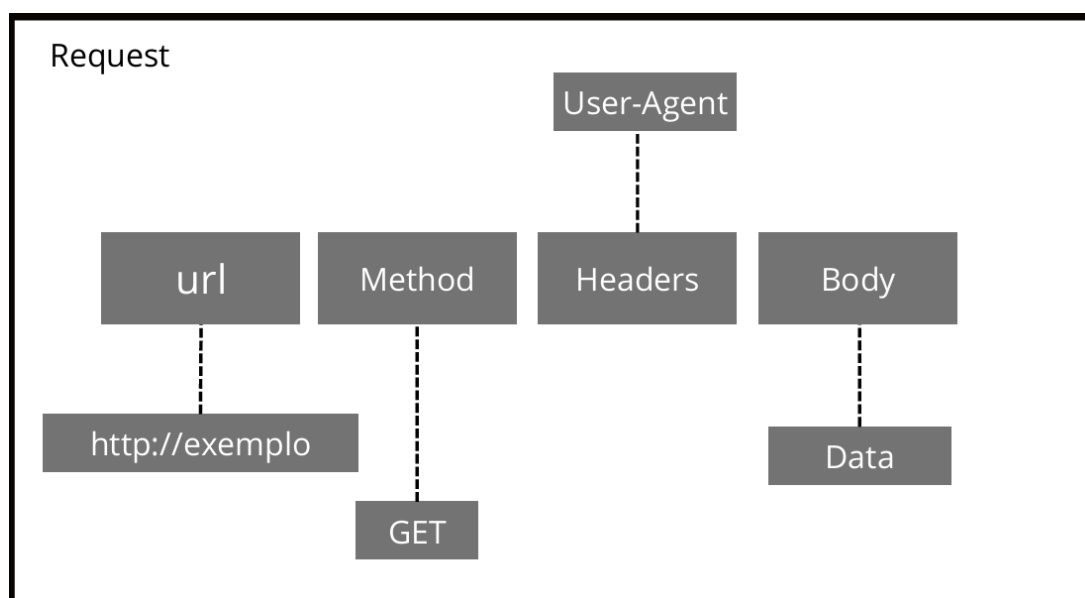


Figura 1: Estrutura de uma requisição HTTP.

Fonte: Elaboração própria, adaptado de (Cooksey, 2014).

Na Figura 1 a URL funciona como um endereço único para identificar um recurso específico, seja uma página, imagem ou dado em uma API, já o *method* define a ação que será realizada sobre esse recurso, podendo ser de obtenção, criação, atualização ou exclusão. Os *headers* fornecem informações complementares sobre a requisição. Por fim, o *body* contém os dados que o cliente envia ao servidor, apresentando formato

livre conforme a necessidade da aplicação (Cooksey, 2014).

Quando essa comunicação ocorre via internet, de modo a permitir a interoperabilidade entre sistemas distintos, utiliza-se o conceito de *web service*. De acordo com Vieira (2024), essa tecnologia viabiliza a troca de informações entre aplicações implementadas em diferentes linguagens ou plataformas. Essa interoperabilidade é possível devido ao uso de protocolos padronizados da internet (Cooksey, 2014). A Figura 2 representa esse funcionamento.

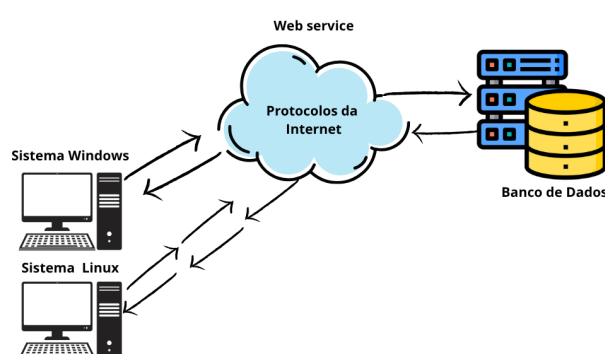


Figura 2: Esquema de comunicação via *web service* entre sistemas e banco de dados.

Fonte: Elaboração própria, adaptado de (Vieira, 2024).

No contexto brasileiro, instituições responsáveis pelo monitoramento hidrometeorológico disponibilizam APIs e *web services* para acesso automatizado aos seus dados. A instituição ANA oferece uma API para acesso automatizado às informações do portal Hidroweb (ANA, 2025a). Para o portal Hidrotelemetria, a ANA disponibiliza o *web service* denominado ServiceANA, que permite integração com sistemas externos (ANA, 2025b).

O ServiceANA disponibiliza catálogos como “DadosHidrometeorologicos”, “Lista-EstaçõesTelemetricas”, “HidroBaciaSubBacia”, “HidroEntidades”, “HidroEstado”, “HidroMunicipio”, “HidroRio” e “HidroSerieHistorica”. O catálogo “DadosHidrometeorológicos” permite a obtenção de dados das estações telemétricas mediante requisição HTTP válida, sendo obrigatório o envio do código da estação. As datas inicial e final são parâmetros opcionais, e os dados são retornados em formato *Extensible Markup Language* (XML) (ANA, 2013).

O CEMADEN também disponibiliza acesso automatizado por meio da Plataforma de Disponibilização de Dados (PED). Conforme o manual da plataforma, o acesso pode ocorrer via interface web ou por meio de chamadas ao *Web Services System* (WSS), sendo necessária autenticação por token gerado pelo Sistema de Gerenciamento de Autenticação e Autorização (SGAA) (Cemaden, 2023).

O SGAA controla o acesso ao SWS, permitindo a consulta de dados apenas a usuários autenticados. Dessa forma, só é possível consultar os dados utilizando o SWS se o usuário estiver cadastrado no SGAA, pois somente com o token gerado pelo sistema é possível consultar os dados (Cemaden, 2023). Na Figura 3 é possível visualizar esse fluxo:

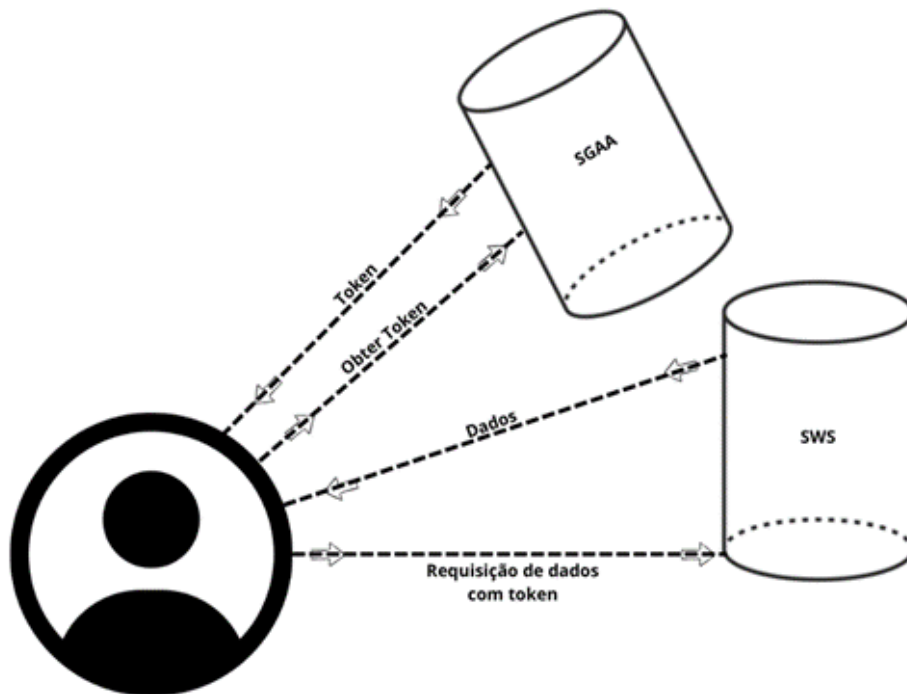


Figura 3: Interação entre o usuário, o SGAA e o SWS para autenticação e acesso aos dados.

Fonte: Elaboração própria, adaptado de (Cemaden, 2023).

O INMET igualmente disponibiliza API para facilitar consultas automatizadas aos seus dados, permitindo que sistemas e aplicações realizem requisições programadas sem necessidade de acesso manual ao portal institucional (INMET, 2025).

Assim, soluções tecnológicas como APIs e *web service* permitem integração automatizada entre sistemas, viabilizando a aquisição programada de dados hidrometeorológicos, reduzindo a necessidade de interação manual com portais institucionais e aumentando a eficiência no processo de obtenção de informações.

2.5 Sistema HEC-DSSVue para Gerenciamento de Séries Temporais

Após compreender as formas automatizadas de acesso aos dados fornecidos por instituições como ANA, CEMADEN e INMET, torna-se necessário discutir a ferramenta adotada para o gerenciamento dessas informações. Embora cada instituição opere diferentes tipos de estações, convencionais, automáticas ou telemétricas, e produza grande volume de dados, a organização e o tratamento dessas séries temporais representam um grande desafio.

As redes de monitoramento hidrológico geram volumes expressivos de informações, tornando o gerenciamento manual inviável diante da complexidade envolvida. Programas de uso geral, como o Excel, podem auxiliar em tarefas pontuais, porém não oferecem suporte adequado para lidar com falhas, lacunas, inconsistências e grandes volumes de dados hidrometeorológicos. Nesse contexto, torna-se necessária a utilização de softwares específicos, desenvolvidos para armazenamento, manipulação e integração eficiente de séries temporais (Silva et al., 2025a).

Entre essas ferramentas, destaca-se o sistema desenvolvido pelo *U.S. Army Corps of Engineers – Hydrologic Engineering Center* (HEC). De acordo com USACE (2024), o HEC-DSSVue é um programa de utilidade visual projetado para trabalhar com conjuntos de dados como precipitação, temperatura e vazão. O sistema permite visualização gráfica e tabular, edição, realização de cálculos automáticos e organização das informações, incluindo inserção, cópia, renomeação e exclusão de séries.

O desempenho do HEC-DSSVue está diretamente relacionado à estrutura de armazenamento adotada pelo HEC-DSS. Segundo USACE (2024), o sistema utiliza uma estrutura baseada em blocos de dados sequenciais, o que o torna eficiente para armazenamento e gerenciamento de séries temporais extensas. Cada conjunto de dados é identificado por um caminho exclusivo, funcionando como um endereço dentro do arquivo. Além disso, os registros armazenam tanto os valores quanto metadados, como unidades e histórico de gravações, tornando o sistema autodocumentado. Os dados são organizados em arquivos binários com extensão "dss", criados e gerenciados conforme a necessidade do usuário.

A aplicabilidade do HEC-DSSVue é evidenciada em diferentes estudos. No trabalho de Lukong et al. (2011), o sistema foi empregado como banco de dados intermediário para converter séries históricas de precipitação e vazão originalmente em formato Excel para o formato dss, viabilizando sua utilização no modelo HEC-HMS. De forma semelhante, Bravo et al. (2004) utilizaram o programa no gerenciamento e tratamento de dados meteorológicos e hidrométricos destinados ao módulo meteorológico do HEC-HMS, integrando-o a um fluxo de trabalho que envolveu também o HEC-RAS e sistemas de informação geográfica.

No estudo conduzido por Fürst et al. (2020), no qual são descritos a criação e o desenvolvimento da bacia experimental de Rosalia, como parte da iniciativa da Universidade BOKU para estudar processos hidrológicos, energéticos e de transporte de solutos no continuum solo-planta-atmosfera (SPAC), que representa o sistema integrado de fluxos de água e energia entre o solo, a vegetação e a atmosfera. O programa HEC-DSSVue foi utilizado para o armazenamento de séries temporais de alta resolução, bem como para a identificação de picos anômalos, falhas durante eventos de manutenção e a visualização dos dados.

Em um contexto mais recente, Unucka et al. (2023) destacam o uso do HEC-DSSVue como banco de dados primário em um sistema automatizado de previsão operacional de cheias na República Tcheca, atuando como intermediário entre sistemas de aquisição de dados e modelos hidrológicos como HEC-HMS, HEC-RAS e *River and Channel Modelling System* (MIKE). Esses estudos demonstram que o sistema não atua apenas como ferramenta de armazenamento, mas como elemento integrador em fluxos de modelagem hidrológica.

Dessa forma, no presente estudo, o HEC-DSSVue foi adotado como ferramenta central para o pré e pós-processamento das séries temporais hidrometeorológicas. Sua utilização permitiu o armazenamento estruturado, a edição, a interpolação, a conversão de unidades, o ajuste de intervalos temporais e a visualização gráfica dos dados. Assim, o programa atua como banco de dados base para o sistema proposto, integrando-se ao fluxo de modelagem e contribuindo para a construção de um ambiente organizado e confiável para previsão e alerta hidrológico.

2.6 Aplicações de Deep Learning (DL) na Previsão de Níveis de Rios

Previsões de curto prazo e precisas dos níveis dos rios são fundamentais, pois essa variável sintetiza de forma direta o estado hidrológico de um sistema fluvial e controla uma série de processos físicos, ambientais e socioeconômicos. Conforme destacado por Luo et al. (2025), esse parâmetro é de extrema relevância para os sistemas fluviais, uma vez que exerce influência direta sobre processos de tomada de decisão relacionados à mitigação de cheias, à geração de energia hidrelétrica, à navegação e ao abastecimento hídrico. Dessa maneira, a obtenção de previsões desse parâmetro torna-se essencial.

Em julho de 2021, a província de Henan, na China, foi atingida por um evento extremo de chuva com acumulados que atingiram 646 mm em apenas 24 horas, valor comparável à média anual do local. O episódio desencadeou inundações severas na área, resultando em 398 óbitos, interrupção prolongada das atividades urbanas e perdas econômicas diretas estimadas em US\$ 17,8 bilhões (He et al., 2023).

Esse infeliz fato, decorrente do avanço das mudanças climáticas, deixa claro que se torna necessário dispor de informações de curto prazo referentes à variável nível da água. De acordo com Hieu et al. (2023), a determinação precisa do nível da água constitui um elemento essencial para a implementação de mecanismos de alerta precoce de inundações, bem como para a mitigação dos danos causados por esses eventos.

Nessa previsão, existem alguns caminhos que os profissionais utilizam para a estimativa do nível da água. Destacam-se duas abordagens principais, segundo Hieu et al. (2023). O primeiro caminho está fortemente ligado aos modelos físicos, que oferecem maior precisão; no entanto, exigem um grande conjunto de dados, além de demandarem longos períodos de simulação. Já o segundo caminho emprega-se de modelos baseados em dados, nos quais os processos físicos não são representados explicitamente. Em vez disso, esses modelos aprendem padrões diretamente a partir dos dados observados.

Referente ao primeiro caminho, os modelos hidrológicos têm apresentado bons resultados na previsão de vazão, pois, por meio desse resultado, é feita a previsão do nível de água dos rios (Prudhomme et al., 2017; Devia et al., 2015; Papacharalampous et al., 2020). No entanto, esses modelos fazem uso de uma gama variada de dados e variáveis de entrada, que incluem, por exemplo, precipitação, temperatura, características do solo, taxas de evaporação e velocidade do vento (Lian et al., 2007).

Para superar as dificuldades encontradas na utilização desses modelos, os pesquisadores buscam seguir o segundo caminho. Por ser uma metodologia que não necessita de uma gama diversificada de informações, esses profissionais estão cada vez mais utilizando modelos de ML a fim de aperfeiçoar a modelagem de rios, previsões relacionadas a alagamentos em áreas urbanas e a previsão de níveis de água, com significativos sucessos (Taormina et al., 2015; Zahura et al., 2020; Huang et al., 2023; Assem et al., 2017; Bonakdari et al., 2019).

Notavelmente, a utilização dessa técnica foi implementada no trabalho de Chieu et al. (2024), no qual foram empregados modelos de regressão baseados em dados, com o objetivo de prever o nível da água no rio Kien Giang. No contexto dos modelos de ML, destaca-se o DL, que tem se consolidado como uma abordagem promissora para a modelagem de sistemas complexos. De acordo com Luo et al. (2025), o DL é um subconjunto do ML que se baseia no emprego de redes neurais com múltiplas camadas, capazes de capturar e representar estruturas complexas presentes nos dados.

Dentro da área de DL, destacam-se os modelos de séries temporais. Esses modelos se caracterizam como um conjunto de abordagens matemáticas e estatísticas destinadas à análise e à previsão de valores futuros, utilizando como base o comportamento histórico de uma variável observada ao longo do tempo. Diferentemente

dos modelos clássicos baseados em dados, que pressupõem independência entre as observações, esses modelos incorporam explicitamente a dependência temporal, considerando a ordem cronológica e a sequência dos dados (Agostinho, 2025).

Nesse contexto, surgem alguns modelos que possuem a capacidade de captar variações e tendências nesses conjuntos de dados com dependência temporal, dentre eles as *Recurrent Neural Networks* (RNN) clássicas e suas variações, pois conseguem capturar dependências entre observações sucessivas, o que tem levado a excelentes resultados em múltiplos domínios de aplicação (Li et al., 2018).

No campo da hidrologia e da gestão de recursos hídricos, modelos baseados em ML e DL têm sido amplamente empregados para diferentes propósitos, especialmente na previsão de variáveis hidrológicas. Nesse contexto, o estudo desenvolvido por Vizi et al. (2023), aplicado ao rio Tisza, na Europa Central, teve como objetivo avaliar o desempenho de diferentes modelos preditivos na estimativa do nível de água, com horizonte de previsão de até sete dias, comparando abordagens baseadas em ML com o modelo hidrológico clássico *Discrete Linear Cascade Model* (DLCM).

O estudo utilizou para a avaliação do desempenho dos modelos, as métricas *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Coefficient of Determination* (R^2) e *Willmott Index* (WI). Os resultados mostraram que o modelo baseado em redes neurais recorrentes do tipo LSTM, utilizando uma arquitetura encoder–decoder, apresentou desempenho superior tanto em relação aos modelos tradicionais quanto aos métodos de ML mais simples, além de superar o desempenho do modelo DLCM na maioria dos horizontes de previsão analisados (Vizi et al., 2023).

Em outro contexto, Atashi et al. (2022) propuseram uma aplicação ao rio Red River of the North, na América do Norte, cujo foco foi comparar o desempenho de diferentes abordagens preditivas na estimativa do nível da água, com horizontes de previsão variando de seis horas até uma semana. Para isso, foram avaliados um modelo estatístico clássico *Seasonal Autoregressive Integrated Moving Verage* (SARIMA), um modelo ML *Random Forest* (RF) e um modelo DL baseado em RNN do tipo LSTM.

Para a verificação do desempenho dos modelos, foram utilizadas as métricas RMSE e o coeficiente de eficiência de Nash–Sutcliffe (NSE). Os resultados indicaram que o modelo LSTM apresentou desempenho superior em relação aos modelos SARIMA e RF em todos os horizontes de previsão e em todas as estações analisadas, destacando-se especialmente nas previsões de médio prazo, como a estimativa com uma semana de antecedência (Atashi et al., 2022).

Entre os estudos recentes, destaca-se o trabalho desenvolvido por Kim et al. (2025), aplicado à bacia do rio Nam Ngum, no Laos, com o objetivo de avaliar o desempenho de um modelo baseado em RNN do tipo LSTM na previsão do nível de água, com ênfase na otimização de hiperparâmetros. O estudo comparou diferentes estratégias de otimização, como *Grid Search*, *Random Search* e *Bayesian Search*,

aplicadas ao ajuste dos principais parâmetros do modelo.

No processo de avaliação do desempenho, foram empregadas as métricas MAE, RMSE e NSE. Os resultados indicaram que a utilização da chuva média da bacia como variável de entrada apresentou maior correlação com o nível do rio em comparação às chuvas pontuais, contribuindo para um melhor desempenho preditivo. Ademais, o modelo LSTM otimizado por meio da abordagem de *Bayesian Search* apresentou desempenho superior em relação às demais técnicas de otimização analisadas (Kim et al., 2025).

Em um cenário diferente, os autores Zakaria et al. (2023) realizaram um estudo no rio Muda, na Malásia, com o objetivo de avaliar o desempenho de diferentes algoritmos de ML na previsão do nível de água em escala diária. O estudo comparou três abordagens amplamente utilizadas na literatura, a saber, o modelo *Multi-Layer Perceptron* (MLP), a RNN do tipo LSTM e o modelo *Extreme Gradient Boosting* (XG-Boost), considerando diferentes combinações de variáveis de entrada e defasagens temporais, incluindo cenários univariados e multivariados com a incorporação de dados meteorológicos.

Para a avaliar o desempenho dos modelos, os autores utilizaram as métricas RMSE, MAE, R^2 e *Mean Absolute Percentage Error* (MAPE). Os resultados da aplicação indicaram que o modelo MLP mostrou melhor desempenho geral para previsões de um dia à frente, já o modelo LSTM demonstrou maior capacidade preditiva em horizontes de previsão mais longos, evidenciando sua aptidão para capturar dependências temporais de longo prazo (Zakaria et al., 2023).

Além disso, existem várias outras aplicações utilizando esses modelos com diferentes fins. No estudo de Huang et al. (2022), foram utilizadas Unidades Recorrentes com Portão (GRU), com o objetivo de prever as trocas entre rios e lagos e as condições da variável nível a jusante. Ainda nesse contexto, destaca-se não só o trabalho de Li et al. (2023), que acoplaram modelos RNN do tipo LSTM a modelos hidráulicos para previsões de níveis de rios, como também o estudo de Xie et al. (2021), que combinaram a transformada wavelet discreta com o modelo LSTM, visando ao aprimoramento da precisão na previsão de níveis de rios.

Não só essas aplicações, mas também diversas outras que utilizam esses modelos demonstram a relevância do uso dessas metodologias na previsão do nível de rios, lagos, entre outras aplicações. No entanto, no contexto da previsão de níveis, esses modelos ainda apresentam limitações. De acordo com Li et al. (2023), quando utilizados em cenários de eventos extremos, os modelos tornam-se menos precisos.

2.7 Origem da Arquitetura LSTM

Para compreender a arquitetura LSTM, é necessário entender sua origem. Conforme apresentado anteriormente, o modelo é uma variação das RNNs, porém a motivação dessa variação decorre de limitações inerentes ao treinamento das RNNs clássicas. Nesse contexto, torna-se necessário entender o funcionamento de uma RNN. A Figura 4 ilustra a arquitetura básica de uma RNN.

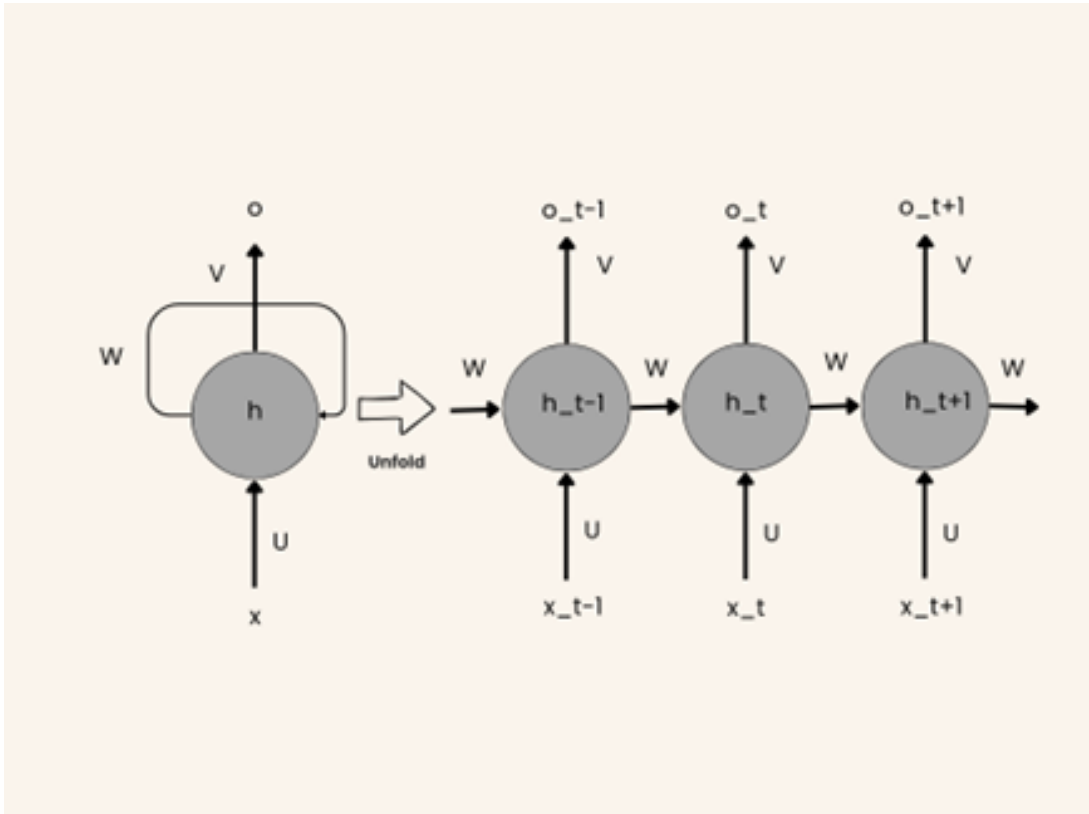


Figura 4: Arquitetura básica de uma RNN

Fonte: Elaboração própria, adaptado de Khan et al. (2023).

Segundo Khan et al. (2023), a arquitetura RNN é composta por uma matriz de entrada x , uma matriz de estado oculto h e uma matriz de saída o , além das matrizes de ponderação U , W e V , e dos termos viés b_i e b_o tanto nas camadas de entrada quanto na camada de saída. Para um instante de tempo t , o estado oculto atual h_t depende da entrada x_t e do estado oculto anterior h_{t-1} , e a saída predita \hat{y}_t é obtida a partir de funções de ativação ϕ_h e ϕ_o , conforme:

$$h_t = Ux_t + Wh_{t-1} + b_{eu} \quad (1)$$

$$o_t = \phi_h(Vh_t + b_o) \quad (2)$$

$$\hat{y}_t = \phi_o(o_t) \quad (3)$$

No processo de treinamento de uma arquitetura RNN, o erro, representado pela função de perda, é calculado e retropropagado ao longo da rede com o objetivo de ajustar as matrizes de ponderação e os termos de viés envolvidos. Entretanto, para determinadas funções de ativação, como a sigmoid e a tangente hiperbólica, a multiplicação sucessiva das derivadas ao longo das etapas temporais pode fazer com que o gradiente se torne excessivamente pequeno, causando um fenômeno conhecido como problema do desaparecimento do gradiente, em contraste quando esse gradiente se torna exageradamente grande, ocorre a explosão desse gradiente. Como consequência, o modelo apresenta dificuldades em aprender dependências de longo prazo presentes nos dados sequenciais (Khan et al., 2023).

Por conta desse problema foram criadas variações da RNN, dentre elas destaca-se as arquiteturas, LSTM e *Gated Recurrent Unit-LSTM* (GRU-LSTM), essas duas técnicas são utilizadas para resolver o problema do desaparecimento do gradiente (Feng, 2023; Ahmadzadeh et al., 2022). Assim, o próximo tópico detalhará a estrutura básica do modelo LSTM.

2.8 Estrutura Básica da arquitetura LSTM

Como discutido brevemente, as RNN são modelos projetados para lidar com dados organizados temporalmente, como dados meteorológicos, hidrológicos, sequências textuais, entre outros tipos de séries temporais. O principal diferencial dessas redes está em sua estrutura interna, que contém conexões recorrentes (loops), permitindo que informações observadas em instantes anteriores sejam armazenadas e utilizadas para orientar previsões futuras.

Assim, a RNN possui uma espécie de memória curta, mas, devido ao problema do desaparecimento do gradiente, que ocorre na etapa de treinamento, à medida que a rede está sendo treinada, as informações mais antigas tendem a perder força, deixando de influenciar corretamente nos resultados das previsões. Dessa forma, essa memória nem sempre funciona adequadamente.

Na busca por metodologias que solucionassem esse problema, foi criada a arquitetura LSTM. De acordo com Khan et al. (2023) essa rede surgiu com o propósito de superar limitações associadas ao aprendizado em dados sequenciais, buscando resolver o problema do desaparecimento do gradiente nas RNNs.

O diferencial dessa rede neural, de acordo com Mieres (2024), é a estrutura de sua memória, a qual é composta por pequenos blocos de memória responsáveis por armazenar informações relevantes por períodos mais longos. Esses blocos decidem quais informações devem ser lembradas e quais podem ser esquecidas, característica

que torna essa arquitetura diferente das demais RNNs. Na Figura 5, é exemplificado a sequência desses blocos.

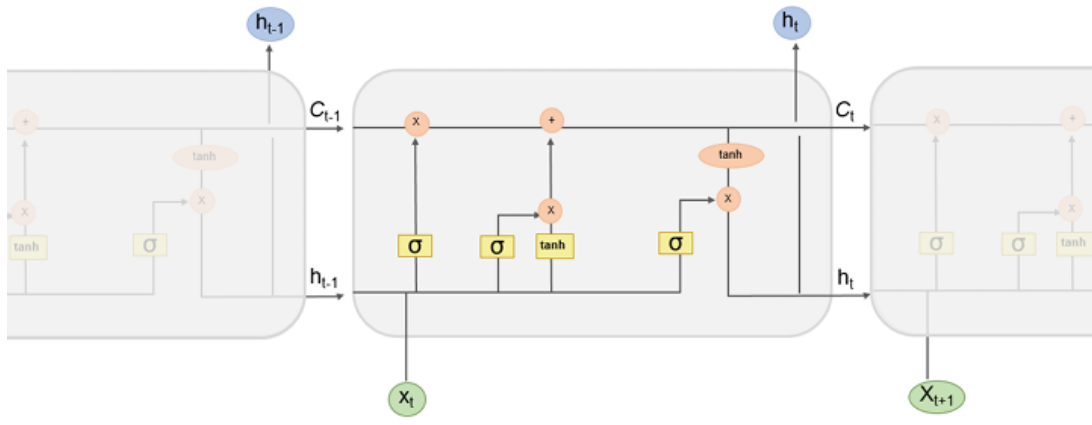


Figura 5: Estrutura geral da arquitetura LSTM ao longo dos passos temporais

Fonte: Mieres (2024).

A Figura 5 mostra o mapa geral da arquitetura LSTM ao longo dos diferentes passos temporais. No entanto, para uma compreensão mais aprofundada de seu funcionamento interno, faz-se necessária a análise detalhada de um único bloco de memória. Nesse contexto, a Figura 6 mostra uma ampliação de um único bloco, com o intuito de detalhar cada componente que constitui os blocos de memória, bem como o fluxo de informações entre eles.

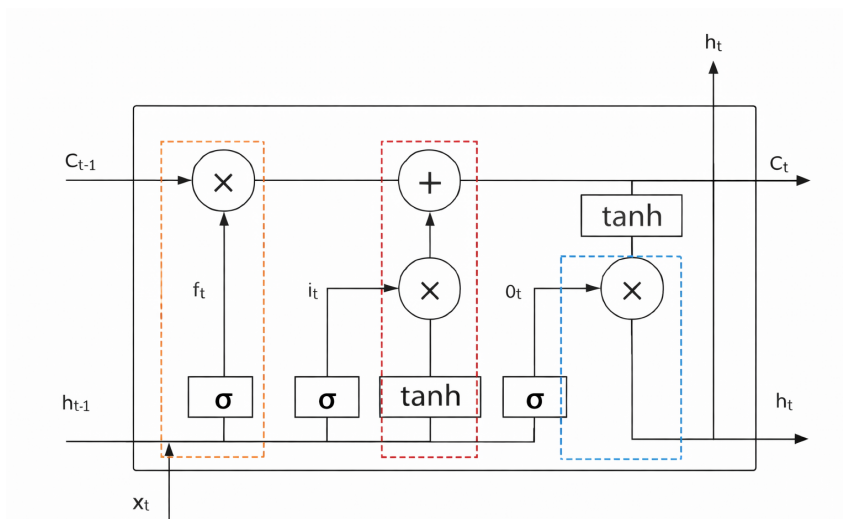


Figura 6: Estrutura interna de um bloco de memória da LSTM

Fonte: Elaboração própria, adaptado de Amaral (2025).

Na Figura 6, o termo x_t corresponde ao vetor de entrada no instante de tempo

t , enquanto h_{t-1} representa o estado oculto proveniente do passo temporal anterior $t - 1$. Os blocos de memória que compõem a arquitetura LSTM são estruturados a partir de três mecanismos de controle, denominados *forget gate*, *input gate* e *output gate* (Mieres, 2024).

Ainda nessa figura, o primeiro mecanismo, *forget gate* (porta de esquecimento), está destacado com tracejado laranja; o segundo componente, *input gate* (também chamado de porta de entrada), está delimitado com tracejado vermelho; e, por fim, o *output gate* (denominado porta de saída) está destacado pelo tracejado azul.

Para Abreu (2024), a primeira porta atua avaliando o estado previamente armazenado, permitindo que a rede descarte informações irrelevantes e preserve apenas aquelas essenciais. A segunda porta define quais novas informações devem ser incorporadas à célula, regulando tanto a geração de novos valores quanto a intensidade com que esses valores atualizam o estado interno. Por fim, a terceira porta determina quais partes do estado atualizado da célula serão utilizadas para produzir a saída da rede, garantindo que apenas as informações mais relevantes sejam propagadas adiante.

Na etapa *Forget Gate*, é realizada a combinação de x_t e h_{t-1} , aplicando-se uma função sigmoide. Essa aplicação gera um vetor f_t , conhecido como sinal, que produz como saída valores entre 0 e 1, em que o valor 0 descarta totalmente a informação e o valor 1 preserva totalmente a informação. Além disso, é esse sinal que controla, por meio de uma multiplicação elemento a elemento, o quanto do estado anterior C_{t-1} será preservado no estado atual C_t (Mieres, 2024).

O vetor de ativação associado tanto à porta *Input* quanto a porta *Forget* são determinados por duas equações, que de acordo com Amaral (2025) W_t e W_f são matrizes de pesos e b_i e b_f são vetores de *bias*, abaixo são apresentadas essas equações.

$$i_t = \sigma (W_t[h_{t-1}, x_t] + b_i) \quad (4)$$

$$f_t = \sigma (W_f[h_{t-1}, x_t] + b_f) \quad (5)$$

Após a realização do cálculo dos vetores f_t (*forget gate*) e i_t (*input gate*), o modelo LSTM realiza a atualização do estado interno da célula. Esse processo consiste na combinação controlada entre a informação proveniente do estado anterior C_{t-1} e o novo conteúdo candidato \tilde{C}_t . O vetor f_t determina, por meio de multiplicação elemento a elemento, qual fração de C_{t-1} será preservada. Já o vetor i_t regula a incorporação do novo conteúdo candidato \tilde{C}_t , controlando a intensidade com que essa nova informação será integrada ao estado da célula (Amaral, 2025). Essa atualização do estado da célula é expressa pela seguinte equação (6):

$$C_t = f_t C_{t-1} + i_t \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (6)$$

Como etapa final, a LSTM busca calcular a porta *Output* do modelo. De acordo com Mieres (2024) , nessa fase, o estado da célula é novamente submetido à função tangente hiperbólica, resultando em um vetor de ativação. Em seguida, esse vetor é multiplicado elemento a elemento pelo sinal proveniente da porta de saída. Esse procedimento define a parcela do estado atual da célula que será efetivamente transmitida como saída do modelo. Abaixo, é apresentada a expressão (7) referente à porta *Output*, bem como a expressão (8) da saída final.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t \tanh(C_t) \quad (8)$$

Cabe destacar que, na expressão (7), W_o é uma matriz de pesos e b_o é um vetor de *bias*. Assim, segundo Lara-benítez et al. (2021), essa arquitetura permite que a rede LSTM represente dependências temporais de longo alcance, ao mesmo tempo em que preserva a capacidade de capturar padrões associados a variações de curto prazo.

3 METODOLOGIA

3.1 Estrutura Geral do Trabalho

Esta dissertação está dividida em duas etapas. A primeira etapa está vinculada ao projeto financiado pelo Edital FAPERGS nº 06/2024 e visou desenvolver três algoritmos, com a finalidade de automatizar a integração de dados hidrometeorológicos das instituições ANA, CEMADEN e INMET no sistema de banco de dado HEC-DSSVue, de forma a constituir, futuramente, um banco de dados que servirá para a operacionalização do sistema de previsão e alerta antecipado da bacia do rio Piratini. Além disso, o banco de dados poderá ser utilizado para outros fins, não limitando seu uso apenas a esse sistema.

Assim, embora esses algoritmos tenham sido desenvolvidos para aplicação na bacia do rio Piratini, nesta pesquisa eles foram testados na Bacia Hidrográfica do Rio Jaguarão (BHRJ). Nesse contexto, será apresentado, de forma detalhada, o desenvolvimento dessa infraestrutura, responsável pela aquisição, tratamento e integração de dados provenientes de diferentes instituições ao sistema HEC-DSSVue.

Finalizada a primeira etapa, deu-se início à segunda; nesse momento, buscou-se utilizar o algoritmo desenvolvido para o *web service* da ANA com o propósito de obter um banco de dados com séries históricas de precipitação e nível. Além disso, são explicadas as funções internas do HEC-DSSVue utilizadas para a conversão das séries regulares de 15 minutos em séries regulares diárias. Esse processo foi essencial para a consolidação da base de dados de entrada do modelo LSTM.

Nessa segunda etapa, a finalidade principal é aplicar a arquitetura LSTM para a previsão do nível máximo diário, buscando avaliar se o modelo apresenta bom desempenho na predição da variável. Assim, a seguir serão apresentados todos os passos do caminho metodológico, desde a escolha da estação utilizada, passando pelo pré-processamento dos dados coletados até o desenvolvimento da arquitetura empregada. to dos dados coletados e o desenvolvimento da arquitetura empregada.

3.2 Área de Estudo

Uma das sub-bacias que compõem a bacia hidrográfica Mirim-São Gonçalo (BHMSG) é conhecida como BHRJ. Localizando-se na divisa dos territórios Brasil e Uruguai e sendo umas das sub-bacias mais relevantes da estrutura da BHMSG, abrangendo áreas da planície costeira e do escudo uruguaio-sul-rio-grandense. Sua rede hidrográfica é organizada em torno do rio Jaguarão, considerado o principal corpo hídrico (Santos, 2022).

Do ponto de vista socioeconômico, nos dois lados da BHRJ a economia é estruturada principalmente em torno da agropecuária, isto é, a agricultura e a pecuária constituem a base do desenvolvimento econômico. A partir disso, os polos industriais da região se expandem, evidenciando que não há uma separação entre o campo e a indústria (Betancur, 2015).

Além dos aspectos socioeconômicos, informações mais específicas, como a posição da BHRJ, são essenciais para sua compreensão de sua localização. De acordo com Santos (2022), a BHRJ está situada especificamente entre as coordenadas 31°30' a 34°35' de latitude Sul e 52°15' a 55°15' de longitude Oeste, abrangendo toda a área de interesse para este estudo. Nesse contexto, esta bacia foi adotada como área de aplicação da pesquisa, sendo representada na Figura 7.

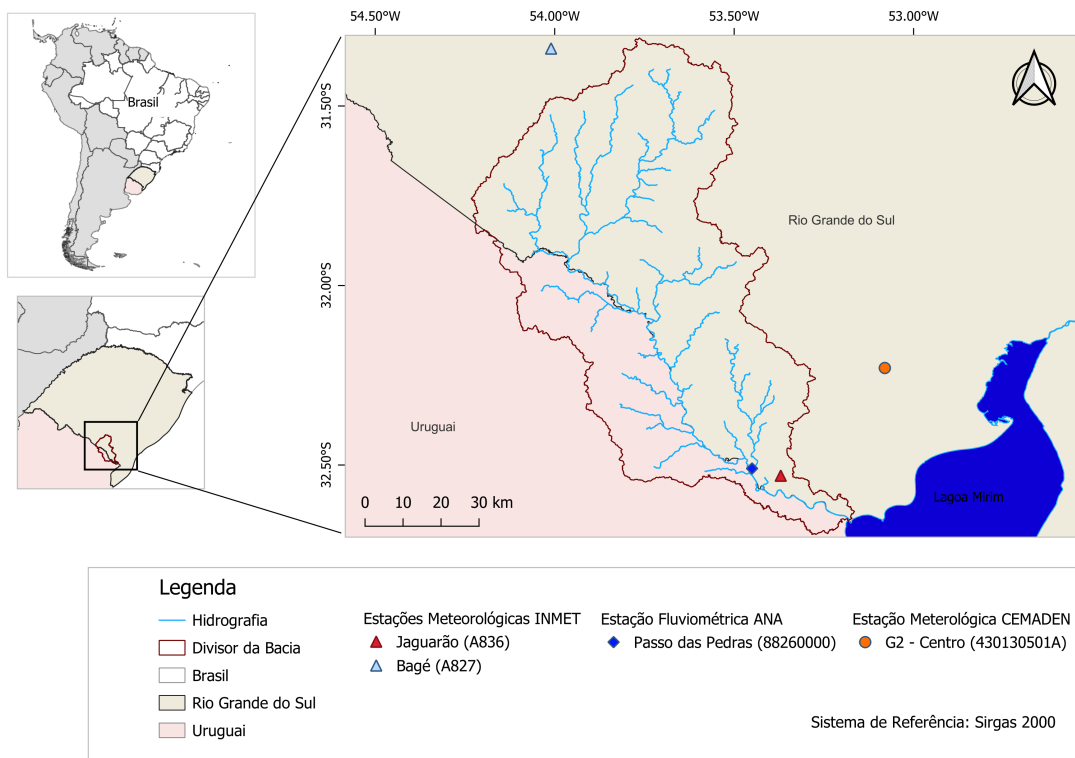


Figura 7: Distribuição das estações hidrometeorológicas utilizadas da BHRJ.

Fonte: Elaboração própria .

Na BHRJ, tanto em seu interior quanto em áreas adjacentes, existem estações de monitoramento hidrometeorológico, conforme ilustrado na Figura 7. Essas estações foram selecionadas para a primeira etapa do trabalho. Assim, para a fase de aplicação dos três algoritmos, foram utilizadas quatro estações no total, sendo duas estações meteorológicas do INMET, uma estação fluviométrica automática ANA e uma estação meteorológica do CEMADEN.

Esse detalhamento é importante, pois a primeira etapa deste trabalho consistiu em desenvolver e testar os algoritmos propostos utilizando a BHRJ, buscando avaliá-los antes de sua utilização pelo sistema de previsão e alerta antecipado da bacia do rio Piratini.

3.3 Etapa de Desenvolvimento dos Algoritmos

O desenvolvimento da automatização seguiu um processo estruturado, garantindo organização, clareza e reprodutibilidade, o que futuramente permitirá a otimização de tarefas referentes aquisição de dados das instituições ANA, CEMADEN e INMET. Inicialmente será definido as ferramentas, como linguagem, bibliotecas e logica utilizada.

Para a implementação, foi utilizada a linguagem Python. Essa linguagem é empregada em diferentes áreas por ser de fácil adaptação e por atender a diversas finalidades, sendo comumente utilizada em projetos que necessitam automatizar tarefas, analisar grandes volumes de dados e até mesmo no desenvolvimento de inteligência artificial (Gonçalves, 2025).

Especificamente a versão do Python utilizada foi 3.13, em conjunto com o ambiente de desenvolvimento integrado *Visual Studio Code*, esse ambiente é um programa utilizado para escrever e testar códigos, oferecendo ferramentas como depuração, integração com o *Git* e gerenciamento de ambientes Python (Gonçalves, 2025).

Durante o processo de desenvolvimento, foram utilizadas bibliotecas internas do Python, como os, *time*, *json*, *logging*, *datetime*, *xml.etree.ElementTree* e *sys*, além de bibliotecas externas, como *Pandas*, *requests*, *numpy*, *tqdm* e *hecdfs*. Essas bibliotecas foram responsáveis por diferentes etapas do fluxo do algoritmo, desde a comunicação com o serviço até a etapa final, o armazenamento dos dados.

A comunicação com o *web service*, foi feita por meio a biblioteca *requests*. De acordo com Silva et al. (2025b), esse modulo possibilita a realização de requisições ou chamadas HTTP, permitindo o acesso a APIs e a obtenção de dados disponíveis de fontes externas. Para o tratamento dos dados recebidos das requisições, foi utilizada a biblioteca *pandas*, considerada a mais relevantes para a manipulação de dados, oferecendo DataFrames que são estruturas complexas que permitem organizar as informações em formato tabela, operações como seleção, filtragem, agregação, limpeza e transformação de dados (Casagrande, 2024).

A organização das informações foi intermediada pela biblioteca *hec_dss*, desenvolvida pelo HEC. O pacote é um wrapper em Python para trabalhar com arquivos no formato ".dss"; na teoria, ele funciona como um intermediário que permite a utilização das bibliotecas nativas. Assim, é possível a leitura, escrita e manipulação dos dados utilizando estruturas mais legíveis.

Para futuramente evitar conflitos de versões dessas dependências, foi criado um ambiente virtual. De acordo com Castro (2024) esse ambiente serve como uma caixa separada, onde ficarão o Python e as dependências do projeto, sendo isoladas do python do sistema principal, evitando conflitos. O ambiente é importante pois diferentes projetos podem precisar de versões diferentes das mesmas bibliotecas, e o ambiente virtual garante esse funcionamento.

Além desses requisitos de configuração, os algoritmos utilizam a mesma lógica. Dessa forma, todos eles utilizam dois modos de execução, denominados "SérieHistórica" e "Operacional". No primeiro modo, o algoritmo é capaz de requisitar séries históricas de vários anos, sendo implementado apenas para carregar o HEC-DSSVue com os dados históricos. Já o segundo modo foi desenvolvido para realizar a atualização dos dados em tempo real. Além disso, a metodologia de processamento segue diferentes formas de tratamento das respostas retornadas pelas APIs.

Foram desenvolvidos três algoritmos nessa primeira etapa, todos implementados em projetos distintos. Apesar de possuírem a mesma lógica, diferem quanto ao fluxo de execução, uma vez que cada um contém suas próprias funções. A seguir, serão apresentadas, nas subseções, a descrição de cada um deles e suas respectivas finalidades.

3.3.1 Algoritmo de integração com ServiceANA

O primeiro algoritmo foi implementado com a finalidade de acessar os dados *web service* da ANA, especificamente o serviço "DadosHidrometeorológicos". Além disso, esse algoritmo, assim como os demais, utiliza as mesmas dependências específicas e configurações especificadas anteriormente.

Para que as informações, como o código das estações acessadas pelo algoritmo, fossem organizadas, foi criado no projeto um arquivo JSON responsável por armazenar as informações das bacias e das estações vinculadas a cada uma delas. Basicamente, ele organiza as estações que serão utilizadas, sendo sempre lido no início da execução para o carregamento da lista de bacias e estações que serão processadas.

Além desse arquivo, foi feita a configuração do sistema de registro, isto é, a forma como o programa vai registrar logs através do módulo *logging*. Esse módulo nativo do Python permite registrar eventos e mensagens gerados por uma aplicação ou por suas bibliotecas, sendo seu principal objetivo centralizar e padronizar o registro desses eventos (Python, 2025).

Para realizar as requisições ao *ServiceANA*, foi criada uma função que, para cada estação do arquivo JSON, realiza automaticamente a aquisição dos dados. Seu uso se destaca pela automatização da comunicação com o serviço, incluindo tentativas sucessivas em caso de falha e a conversão dos dados obtidos para o formato Data-Frame do *Pandas*. Além disso, a função cuida para que os valores retornados estejam no formato numérico.

Para não haver conflitos no processo de organização dos dados no programa HEC-DSSVue, foi implementada uma rotina de padronização temporal, responsável por organizar as observações em intervalos regulares de 15 minutos. Essa etapa é necessária para garantir a consistência das séries temporais, uma vez que conjuntos de dados hidrológicos podem apresentar falhas ou registros em intervalos irregulares. Uma série temporal é considerada regular quando suas observações ocorrem em intervalos igualmente espaçados no tempo (USACE, 2024).

Com o conjunto de dados tratados, faltava apenas organizá-los no programa. Para isso, foi desenvolvida uma função para armazená-los. A função recebe o DataFrame com os dados padronizados e grava-os em um único arquivo dss. Sua utilização se destaca pela metodologia automatizada de organizar as informações de cada variável, desde a criação específica de *pathnames* até o tratamento de valores ausentes.

Com as funções definidas, foi implementado um bloco responsável por controlar todo o fluxo do algoritmo. Essa parte do algoritmo é o núcleo do sistema: todo o processo, desde o modo de operação até a etapa final da execução, acontece nesse bloco. O fluxograma do bloco principal é representado na Figura 8.

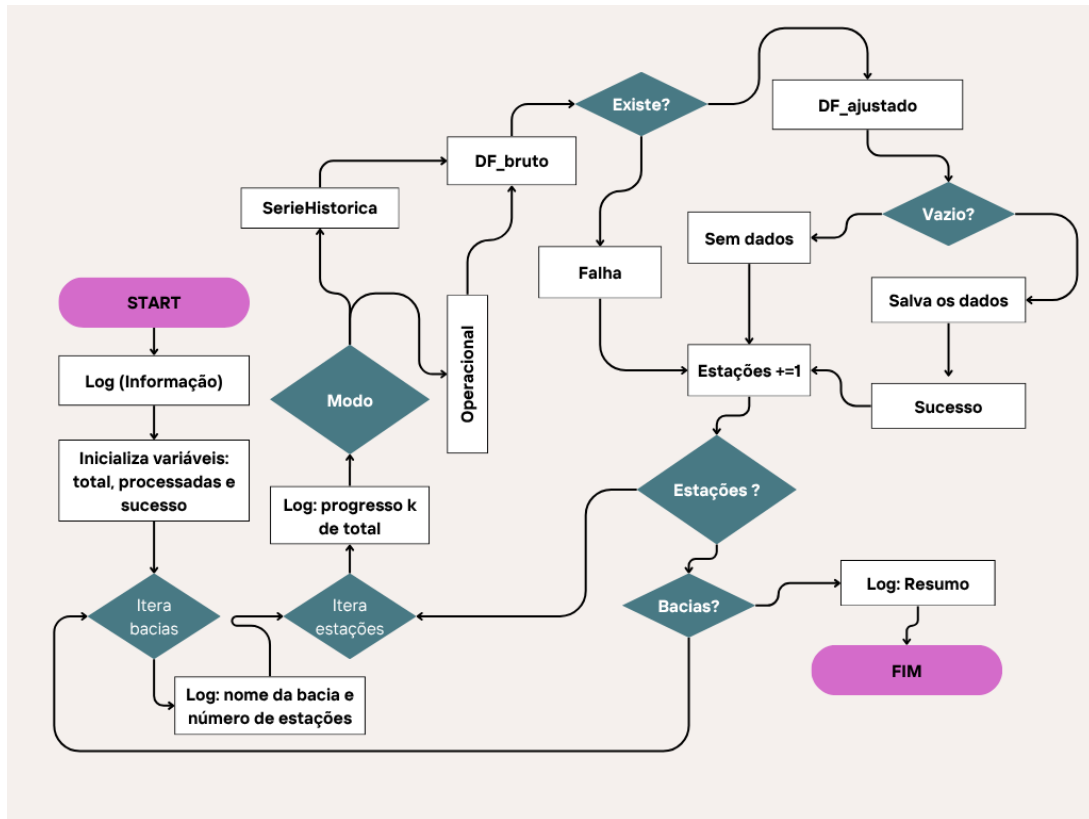


Figura 8: Bloco principal do algoritmo

Fonte: Elaboração própria.

3.3.2 Algoritmo de integração com a API do CEMADEN

O segundo algoritmo foi implementado para acessar dados da PED do CEMADEN, utilizando os SWS. Além disso, assim como a lógica anterior, ele opera em dois modos e utiliza o arquivo json, entretanto a estrutura do algoritmo é totalmente diferente, seguindo as regras impostas para a utilização da PED.

A PED utiliza o sistema SGAA e, tendo o cadastro nesse sistema, a próxima etapa consistiu na implementação de uma função responsável por solicitar o token. Dessa forma, a função realiza uma requisição, enviando as credenciais de e-mail e senha registradas no sistema; para esse envio de credenciais, é utilizado o método POST. Caso não ocorra nenhum erro, a função extrai do corpo da resposta o token, que será utilizado nas requisições. Caso a autenticação falhe, a função interrompe a execução e informa o código de erro.

Após a criação da função que gera o token, foi criada a função dedicada a realizar as requisições para a PED. Para a solicitação, a função utiliza um controle de tentativas e tratamento de erros; dessa maneira, caso o serviço retorne uma resposta válida, é encerrada a execução e retornado o conteúdo. Porém, caso ocorra alguma falha na solicitação, a função realiza até cinco tentativas, aguardando um tempo de 60 segundos entre cada tentativa.

Em seguida, foi implementada a função responsável por requisitar séries históricas, esta função serve exclusivamente para contornar a restrição quanto ao volume de dados solicitado. Assim, o algoritmo divide o período total solicitado em janelas de até 31 dias, realizando múltiplas chamadas consecutivas por meio da função anterior até completar o intervalo definido.

Para garantir a regularidade temporal das respostas, foi implementada a função cuja finalidade é a padronização temporal dos dados, assegurando que o eixo temporal siga uma sequência contínua e regular, evitando, posteriormente, conflitos no processo de organização dos dados, uma vez que o método utilizado espera uma série regular.

Na próxima etapa de desenvolvimento, foi criada a função dedicada em organizar os dados regulares no programa HEC-DSSVue, sua responsabilidade consiste somente no armazenamento dos dados, para isso, realiza vários processos, como verificação e conversão dos dados para o formato numérico, além de construir os pathnames.

Para a finalização do algoritmo, foi desenvolvido duas estruturas, a primeira tem a finalidade de gerenciar os modos de aquisição, de forma que, dependendo da configuração, o algoritmo utilizará o modo série histórica ou operacional. Além disso, é nesse processo que as respostas são organizadas em um DataFrame, no qual são filtradas somente as informações de interesse. Por fim, esse DataFrame é armazenado em uma lista que servirá para a consolidação dos dados. A Figura 9 ilustra o fluxograma correspondente a estrutura:

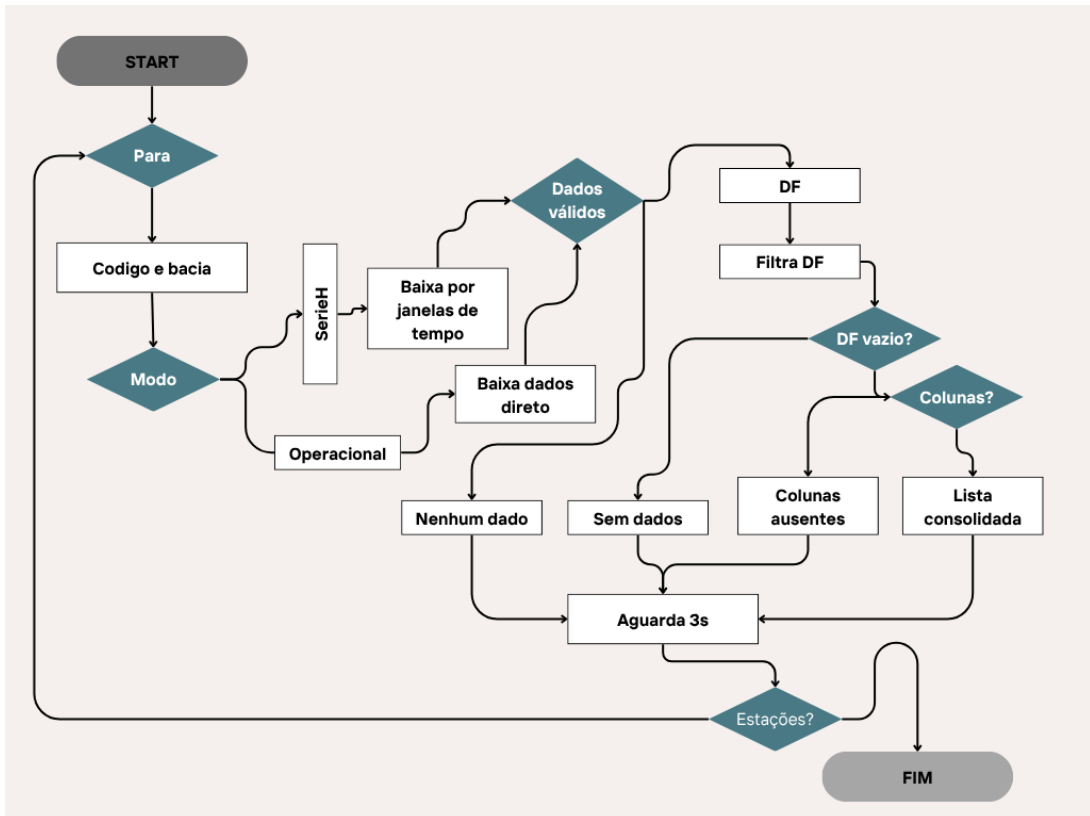


Figura 9: Processo de aquisição e tratamento dos dados

Fonte: Elaboração própria.

A segunda estrutura é responsável pela realização do processamento final da resposta. Nesse processo, o algoritmo utiliza a função de regularização e a função que organiza os dados no programa HEC-DSSVue. Além disso, é realizado o ajuste de fuso horário para o horário padrão, pois as respostas recebidas seguem o padrão de tempo universal (UTC). A Figura 10 apresenta o fluxograma desse processo:

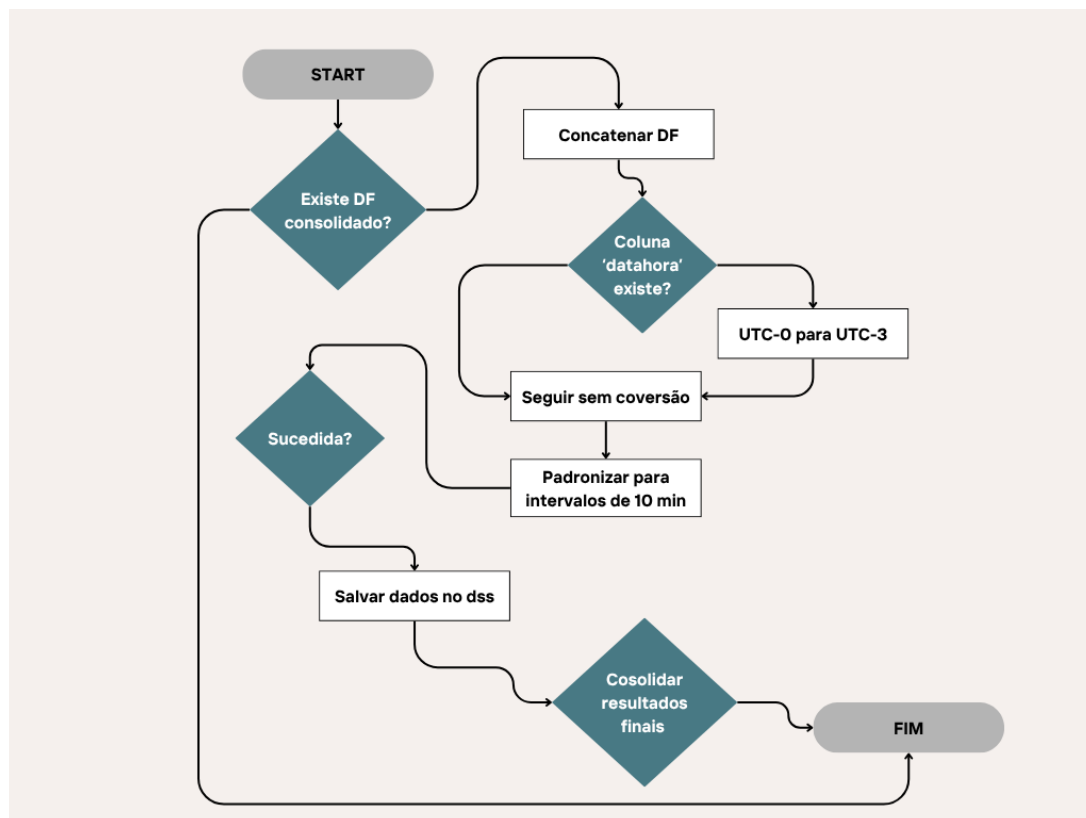


Figura 10: Processo de armazenamento dos dados no HEC-DSSVue

Fonte: Elaboração própria.

3.3.3 Algoritmo de integração com a API do INMET

O terceiro e último algoritmo foi desenvolvido com as mesmas finalidades dos anteriores, tendo como foco requisitar, processar e armazenar dados meteorológicos provenientes da API do INMET. Assim como nas etapas anteriores, o processo de aquisição opera em dois modos e utiliza o arquivo json, bem como a lógica de algumas funções já implementadas. No entanto, a estrutura deste algoritmo apresenta um fluxo totalmente distinto, uma vez que o serviço do INMET utiliza um sistema de autenticação diferente, baseado em chave de acesso única, não sendo necessária a implementação de uma função específica de autenticação, como ocorreu no algoritmo do CEMADEN.

Nesse algoritmo, foram implementadas oito funções, cada uma com finalidades específicas. Assim, a primeira etapa consistiu na criação da função dedicada a ler o arquivo JSON; ela realiza a abertura do arquivo e converte as informações para uma estrutura de dicionário do Python. Além disso, para evitar conflitos com caracteres especiais e acentos no processo de salvamento das respostas no HEC-DSSVue, foi implementada a função responsável por remover essas inconsistências, evitando erros no processo de criação dos *pathnames*. Também foram definidas, no algoritmo, funções voltadas ao processo de conversão dos valores, considerando que os dados

recebidos são originalmente do tipo *string*, o que exigiu o tratamento antes de serem armazenados.

Para a padronização das respostas recebidas, foi implantada uma função cuja finalidade é estruturar a resposta para o formato de armazenamento do programa. Durante o processo, a função trata a questão do fuso horário, utiliza as funções de conversão de valores e realiza cálculos de conversão de unidades, uma vez que o HEC-DSSVue utiliza, para algumas variáveis, unidades específicas. Ao final, a função devolve as informações organizadas em um dicionário.

Após essa etapa, foi criada a função responsável por salvar os dados já organizados. Para o desenvolvimento dessa função foram utilizadas as mesmas classes dos algoritmos anteriores, classes como *HecDss* e *RegularTimeSeries* da biblioteca *hecdss*, além da lógica de tratamento de valores ausentes. Definidas todas as funções, a última etapa consistiu na implementação do bloco destinado a controlar a execução do algoritmo, seguindo o mesmo princípio dos blocos de controle anteriores. Assim, toda a operação é conduzida dentro de um laço de repetição, no qual, para cada estação, são realizadas a requisição, organização e salvamento dos dados no programa HEC-DSSVue. O fluxograma desse bloco de controle é apresentado na Figura 11.

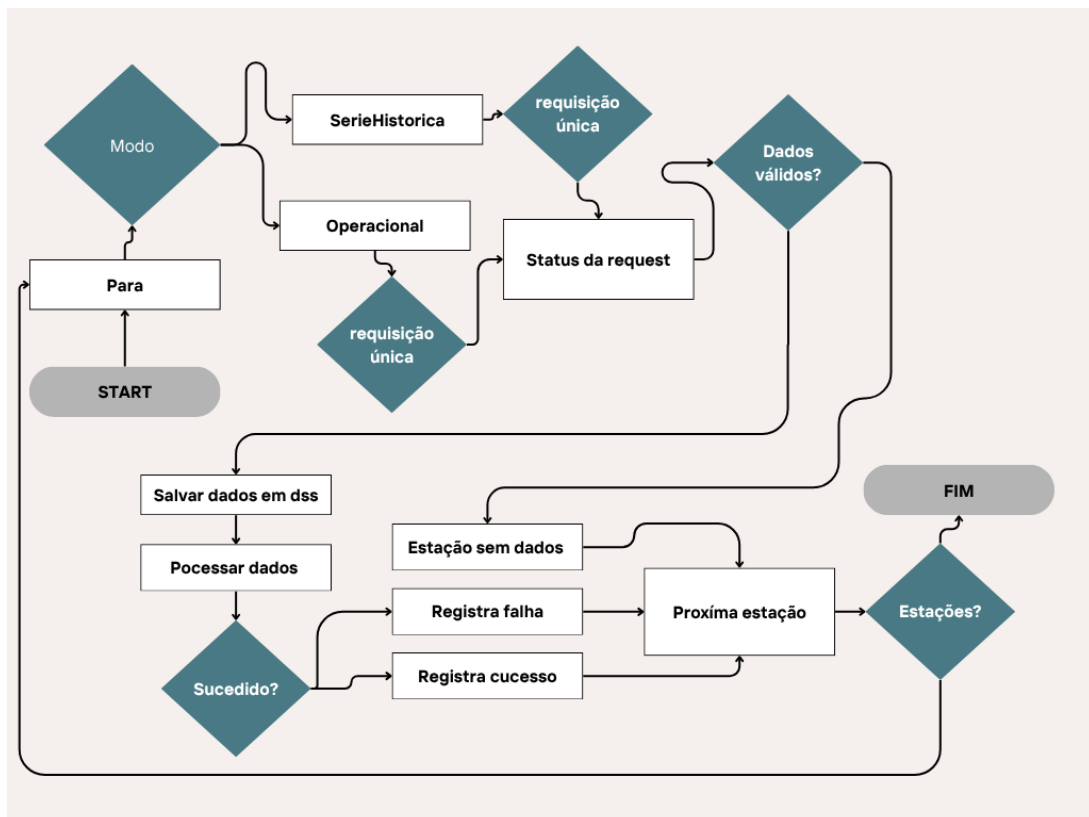


Figura 11: Estrutura lógica de controle do algoritmo

Fonte: Elaboração própria.

3.4 Consolidação da Base de Dados

Para compor a base de dados para a arquitetura LSTM, foram selecionados dados de nível e precipitação da estação Passo das Pedras, código 88260000, uma estação hidrometeorológica automática localizada na porção a jusante da bacia. A Figura 12 apresenta a BHRJ com a estação selecionada.

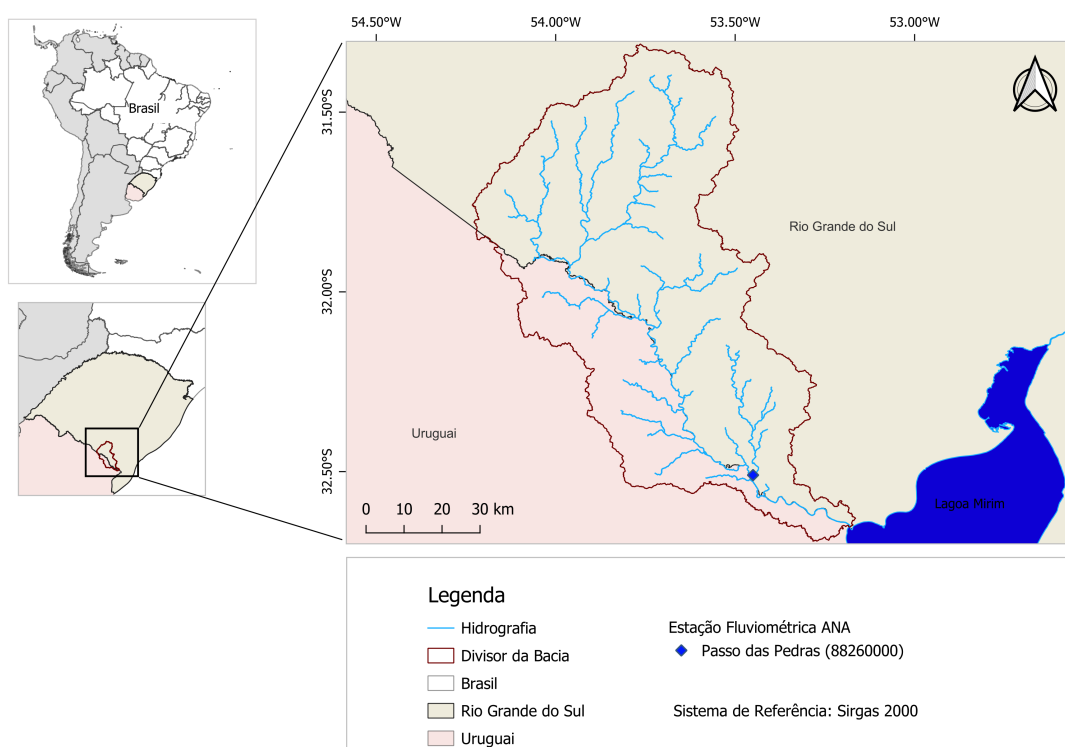


Figura 12: BHRJ com a estação selecionada

Fonte: Elaboração própria.

Os dados dessa estação foram obtidos por meio do algoritmo de integração com o *Web Service* da ANA, cuja utilização buscou facilitar o gerenciamento das informações antes da etapa de pré-processamento. Como resultado, foi gerado um arquivo no formato DSS, denominado “passoDasPedras.dss”, com 2,67 MB, contendo as séries históricas de nível e precipitação.

A partir desse arquivo contendo as duas séries, deu-se início à etapa seguinte, na qual foram aplicadas ferramentas do sistema HEC-DSSVue para a transformação temporal das séries, procedimento descrito na seção subsequente. .

3.5 Pré-processamento das Séries

Este tópico é dedicado a abordar tanto a fase de conversão das séries históricas de nível e precipitação quanto a identificação de dados ausentes nesses conjuntos.

Primeiramente, foi realizado o ajuste temporal das séries históricas, pois a arquitetura LSTM necessitava de uma base de dados organizada em ordem temporal diária. Antes desses ajustes, os dados encontravam-se em intervalos regulares de 15 minutos.

Para realizar essa operação, utilizaram-se as ferramentas do HEC-DSSVue, uma vez que o software conta com um amplo aparato de funções matemáticas. De acordo com Silva et al. (2025a), o sistema disponibiliza mais de 60 funções matemáticas, distribuídas nas categorias: *Arithmetic*, *Geneal*, *Time Functions*, *Hydrologic*, *Smoothing* e *Statistics*.

Para criar uma série regular diária, foram utilizadas duas funções da categoria *Time Functions*. Para conjunto de dados do parâmetro precipitação, foi utilizada a função *Change Time Interval*; já para a série de nível, utilizou-se a função *Maximum for Period*.

Inicialmente, buscou-se obter uma série regular diária de precipitação a partir de uma série regular com intervalo de 15 minutos. Para isso, utilizou-se a ferramenta *Change Time Interval*, disponível em *Time Functions* do HEC-DSSVue. Essa função gera uma nova série regular com intervalo de tempo distinto (1DAY), mantendo a consistência temporal dos dados.

Conforme USACE (2024), o comportamento da transformação depende do data type da série original, sendo essa característica determinante para o método de cálculo adotado. Para séries de *type INST-VAL* ou *INST-CUM*, o valor no novo intervalo é obtido por interpolação linear. Para séries com *type PER-AVER*, o valor correspondente ao novo intervalo é calculado como a média do período. Já para séries do tipo *PER-CUM*, o valor no novo intervalo corresponde ao acumulado dentro do período, ou seja, à soma dos valores contidos no intervalo diário.

O mesmo ocorre na aplicação da função *Maximum for Period*, cujo procedimento de interpolação de um conjunto de dados depende diretamente da natureza do dado, ou seja, do data type. No caso do parâmetro nível, o *type* é classificado como *INST-VAL*. Assim, de acordo com USACE (2024), essa função é calculada identificando-se o maior valor entre os pontos da série original compreendidos no novo intervalo de tempo e, adicionalmente, realizando-se a interpolação linear nos instantes inicial e final desse intervalo, a fim de verificar se os valores interpolados correspondem ao valor máximo ao longo do período considerado.

Após a aplicação dessas funções, duas novas séries regulares diárias foram criadas: uma série de precipitação diária e outra de nível máximo diário. A partir desse ponto, buscou-se analisa-las, com o propósito de identificar o número de dias em cada uma e a quantidade de dados ausentes. Abaixo, a Tabela 1 apresenta a análise feita nesses conjuntos de dados.

Tabela 1: Resumo das séries temporais diárias utilizadas no estudo

Estação	Nome	Variável	Período	Frequência	Total de dias	Dias com Ausência
88260000	Passo das Pedras	Chuva Diária	2014- 2025	Diária	4186	58
88260000	Passo das Pedras	Nível Diário Máximo	2014- 2025	Diária	4186	313

A análise das séries revelou a presença de dados ausentes tanto no conjunto precipitação diária quanto no nível máximo diário. Para a precipitação, as falhas correspondem a 1,39% do total de observações, representando baixa expressividade estatística frente ao conjunto de dados. Em contraste, a série de nível máximo apresentou 7,48% de valores ausentes, percentual significativamente superior, decorrente da existência de um intervalo contínuo de falhas na série original.

3.6 Desenvolvimento do modelo LSTM

Nesta seção descreve-se a implementação do modelo LSTM para previsão do nível máximo diário com horizonte de um dia à frente, a partir das séries de precipitação diária e nível máximo diário. Convém destacar que essa rede é uma extensão da RNN amplamente empregada na modelagem de séries temporais por sua capacidade de representar dependências de curto e longo prazo. Os fundamentos teóricos relacionados a essa arquitetura foram apresentados na seção de 2.9.

A implementação foi realizada em Python, utilizando as bibliotecas *NumPy* e *pandas* para manipulação de dados e o ecossistema *TensorFlow/Keras* para construção e treinamento do modelo. O *TensorFlow* fornece suporte ao treinamento de modelos de aprendizado profundo com cálculo automático de gradientes, enquanto o *Keras* atua como API de alto nível para definição e ajuste de redes neurais, integrada ao *TensorFlow* (TensorFlow, 2026).

Inicialmente, as séries de precipitação diária e nível máximo diário foram organizadas em arquivos no formato .csv, contendo duas colunas: a primeira correspondente às datas e a segunda aos respectivos valores observados. Posteriormente, essas séries foram carregadas e mescladas com base na variável temporal, resultando em um único DataFrame multivariado contendo ambas as variáveis de interesse.

Cabe destacar que uma série temporal é um conjunto de valores observados ao longo do tempo, registrados em intervalos específicos e organizados em ordem cronológica (Tzoumpas et al., 2024). Nesse contexto a fim de garantir consistência temporal

das duas series que foram mescladas, realizou-se a padronização da frequência da série para periodicidade diária.

Assim, foi criado um índice temporal completo entre a menor e a maior data disponível, reindexando-se a serie temporal para assegurar a presença explícita de todos os dias no intervalo analisado. Esse procedimento permite a correta identificação de lacunas na série histórica.

Antes da criação do DataFrame, foi verificado na fase de pré-processamento, que as duas séries apresentam valores ausentes. Para resolver esse problema, utilizou-se o método de interpolação linear para estimar os valores ausentes nessas colunas, fazendo com que as séries ficassem contínuas, sem lacunas, conforme descrito no estudo de (Candido et al., 2025).

Para a aplicação da interpolação linear nas duas séries temporais, utilizou-se inicialmente a função *interpolate()* da biblioteca *pandas*. Essa função recebe como argumento o parâmetro *method*, responsável por definir o método de interpolação a ser utilizado. Neste trabalho foi adotado o método "*linear*", no qual os valores ausentes são estimados a partir de uma relação linear entre os pontos adjacentes conhecidos da série temporal. Esse método assume que a variação entre duas observações consecutivas pode ser representada por uma reta, permitindo estimar valores intermediários de forma contínua ao longo da série (Pandas Development Team, 2026).

Posteriormente, para tratar possíveis lacunas remanescentes nas extremidades do período analisado, foram utilizadas as funções *forward fill (ffill)* e *backward fill (bfill)*, também disponíveis na biblioteca *pandas*. Essas funções permitem propagar o último valor válido anterior ou o próximo valor válido na série, respectivamente, garantindo a continuidade temporal das séries utilizadas no processo de modelagem (Pandas Development Team, 2026).

Logo em seguida, foi realizada a normalização dos dados. De acordo com Mieres (2024), esse processo consiste na aplicação de uma transformação matemática com o objetivo de reescalar os valores das variáveis para um intervalo previamente definido, esse procedimento torna-se essencial pelo fato de as variáveis de precipitação e nível apresentarem unidades de medida distintas. Ao realizar o escalonamento, ambas passam a compartilhar a mesma escala numérica, o que favorece o processamento pela rede neural, mantendo-se, entretanto, a representatividade dos valores extremos.

Assim, visando aplicar essa normalização para o aprimoramento da estabilidade numérica e ao melhor desempenho do processo de treinamento do modelo, as variáveis de precipitação diária e nível máximo diário foram normalizadas por meio da classe *MinMaxScaler* da biblioteca *scikit-learn*, reescalando os valores para o intervalo $[0,1]$. A transformação aplicada pode ser expressa pela seguinte equação:

$$x_{\text{normalizado}} = \frac{x - X_{\min}}{X_{\max} - X_{\min}} \quad (9)$$

Em que x representa o valor original da variável, X_{\min} e X_{\max} correspondem, respectivamente, aos valores mínimo e máximo observados no conjunto de dados utilizado para ajuste do escalonamento. Além disso, cabe destacar que os parâmetros de normalização X_{\min} e X_{\max} foram estimados exclusivamente a partir do conjunto de treinamento.

Posteriormente, esses mesmos parâmetros foram utilizados para transformar os conjuntos de validação e teste, evitando vazamento de informação futura no processo de treinamento. Após a geração das previsões pelo modelo, realizou-se a transformação inversa dos valores normalizados para o domínio físico original (centímetros), de modo que as métricas de desempenho fossem calculadas na unidade real do fenômeno hidrológico.

Definida o processo de normalização dos dados, foi necessário adequar a matriz de entrada ao formato esperado pela arquitetura LSTM. Inicialmente os dados encontravam-se organizados em uma matriz bidimensional, contendo apenas as observações nas linhas (amostras) e duas colunas (features) correspondentes às variáveis precipitação diária e nível máximo diário. Entretanto, as redes LSTM requerem entradas tridimensionais, nas quais, além do número de amostras e variáveis, deve-se definir explicitamente o número de passos temporais (time steps) considerados em cada amostra (Abreu, 2024).

Para tornar as entradas compatíveis com o modelo, foi utilizado o método sliding window, também conhecido como janela deslizante. Segundo Martins (2024), essa técnica, denominada na literatura de estatística e séries temporais como método de latência, consiste em utilizar um conjunto de observações passadas, cuja quantidade é definida pela largura da janela (ou tamanho do atraso), para estruturar o problema de previsão. A janela é deslocada sequencialmente ao longo da série, permitindo reorganizar os dados temporais em um formato compatível com problemas de aprendizado supervisionado.

Com base nesse princípio metodológico, adotou-se neste estudo uma janela temporal deslizante com comprimento de 20 dias. Essa configuração foi baseada na mesma configuração adotada no trabalho de Candido et al. (2025). Em seguida, também foram utilizados outros tamanhos de janela; contudo, essas variações não ocasionaram aumento significativo no desempenho do modelo LSTM. Assim, a estrutura original dos dados, anteriormente bidimensional, passou a ser representada por uma matriz tridimensional com formato (número de amostras, 20, 2), em que 20 corresponde aos passos temporais considerados e 2 ao número de variáveis de entrada.

O próximo passo consistiu na divisão do conjunto de dados. De posse da série multivariada, a separação foi realizada de forma sequencial, preservando a ordem cronológica das séries temporais. O conjunto total foi dividido da seguinte forma: 70% para treinamento, 15% para validação e 15% para teste. A divisão ocorreu de ma-

neira estritamente sequencial, sem qualquer embaralhamento das observações, assegurando que o modelo fosse treinado apenas com dados passados e avaliado em períodos futuros, respeitando a estrutura causal do fenômeno analisado.

Com a divisão dos dados estabelecida e as entradas organizadas em uma matriz tridimensional, a próxima etapa consistiu na implementação do modelo LSTM para previsão do nível máximo diário. A arquitetura foi estruturada com duas camadas LSTM empilhadas, seguidas por camadas densas totalmente conectadas, incorporando técnicas de regularização por meio de Dropout com o objetivo de reduzir o sobreajuste.

A primeira camada LSTM foi configurada com 128 unidades e ativação ReLU, retornando sequências completas, permitindo que sua saída fosse utilizada como entrada para a segunda camada LSTM. A segunda camada, composta por 32 unidades e também com ativação ReLU, foi configurada para retornar apenas o último estado oculto, consolidando a representação temporal extraída da sequência de 20 dias.

Após as camadas recorrentes, foi adicionada uma camada densa com 8 neurônios e ativação ReLU, seguida por Dropout, culminando em uma camada de saída composta por um único neurônio com ativação linear, responsável por fornecer a estimativa do nível máximo diário normalizado correspondente ao dia seguinte. No treinamento utilizou-se o otimizador Adam e a função de perda MAE, abaixo a tabela mostra o resumo da arquitetura LSTM e sua configuração.

Tabela 2: Arquitetura LSTM e configuração

Camada	Parâmetros
Entrada	20 dias, chuva diária e nível máximo diário
LSTM 1	128 unidades, ativação ReLU, com retorno de sequências
Dropout	20% de neurônios
LSTM 2	32 unidades, ativação ReLU, sem retorno de sequências
Dropout	20% de neurônios
Dense	8 neurônios, ativação ReLU
Dropout	10% de neurônios
Saída	1 neurônio

Para avaliar a acurácia do modelo LSTM, foram utilizadas as métricas RMSE, NSE e MAE, que, de acordo com Atashi et al. (2022) essas métricas são amplamente utilizadas em hidrologia para comparar valores observados com valores previstos pelos modelos.

A métrica RMSE permite quantificar o desvio médio quadrático entre os valores observados e os valores simulados pelo modelo, sendo sensível a erros de maior magnitude. Conforme descrito por Machado (2025), essa métrica é expressa pela seguinte equação.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2} \quad (10)$$

Em que Y_i representa o valor observado, Y'_i o valor estimado pelo modelo e n o número total de observações.

A métrica NSE, conforme descrita por Medeiros (2022), avalia o desempenho do modelo ao comparar o erro das previsões com o erro obtido ao utilizar apenas a média histórica como estimativa. É um índice adimensional que varia de $-\infty$ a 1, sendo que valores próximos de 1 indicam excelente desempenho; valores próximos de 0 mostram que o modelo não supera a média; e valores negativos indicam desempenho inferior ao uso da média histórica como previsão, abaixo, apresenta-se a formulação matemática.

$$NSE = 1 - \frac{\sum_{i=1}^n (O_t - P_t)^2}{\sum_{i=1}^n (O_t - \bar{O})^2} \quad (11)$$

Na equação, O_t representa o valor observado do nível de água no instante t , enquanto P_t indica o valor previsto pelo modelo para o mesmo instante. O termo \bar{O} corresponde à média dos valores observados do nível de água, e n representa o número total de observações consideradas no conjunto de dados.

Outra métrica adotada foi o MAE. Segundo Machado (2025), essa métrica consiste em um indicador estatístico que expressa a média dos desvios absolutos entre os valores estimados pelo modelo e aqueles efetivamente observados. Dessa forma, fornece uma avaliação clara da magnitude média dos erros, permitindo interpretar, de maneira direta, o quanto as previsões se afastam dos dados reais. Matematicamente, essa métrica é representada pela seguinte expressão.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - Y'_i| \quad (12)$$

Em que Y_i são os valores observados, Y'_i são os valores previstos pelo modelo e n é o número total de observações.

4 RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados os resultados obtidos na primeira e na segunda etapa do trabalho. Como primeiro resultado, destacam-se os três algoritmos desenvolvidos para a aquisição de dados hidrometeorológicos, cuja principal relevância consiste na eliminação do processo manual de aquisição e organização dos dados para o sistema HEC-DSSVue.

Para demonstrar a relevância desses algoritmos, foram utilizadas as estações anteriormente selecionadas na etapa metodológica. Essas estações serviram para testar os algoritmos desenvolvidos, sendo utilizadas apenas como apoio na Etapa 1 da dissertação. Abaixo, a Tabela 3 apresenta informações complementares de cada uma das estações.

Tabela 3: Estações utilizadas para teste de aplicação dos algoritmos

Nome	Código	Responsável	Latitude	Longitude	Atitude (m)
Passo das Pedras	88260000	ANA	-32,51	-53,45	2
G2-Centro	430130501A	CEMADEN	-32,23	-53,08	3
Jaguarão	A836	INMET	-32,53	-53,37	31,48
Bagé	A827	INMET	-31,34	-54,01	226,19

Para possibilitar uma aplicação mais dinâmica de ambos os algoritmos, foram configuradas rotinas de execução no Agendador de Tarefas do Windows. Esse processo foi utilizado para testar a operacionalização em tempo real. Dessa forma, os algoritmos foram programados para serem executados a cada 24 horas. Essa foi a configuração adotada para o teste; no entanto, poderia ser ajustada para outro intervalo de tempo, conforme a necessidade.

Como foi explicado na seção de desenvolvimento desses algoritmos, eles possuem dois modos. Assim, para a primeira execução, os algoritmos foram configurados para o

modo de série histórica. Esse modo foi utilizado uma única vez, servindo apenas para o carregamento das séries históricas de cada estação no HEC-DSSVue. Dessa forma, nas execuções subsequentes, os algoritmos passaram a utilizar o modo operacional, com o propósito de atualizar as informações das últimas horas.

Na primeira execução, foram gerados três arquivos DSS, cada um contendo dados das respostas retornadas de cada serviço. Quando esse arquivo é gerado, automaticamente as séries temporais de cada variável são organizadas em pathnames. De acordo com (USACE, 2024), esses pathnames funcionam como endereços padronizados compostos por seis partes rotuladas de “A” a “F”, separadas por barras (“/”). Cada parte representa um atributo da série; por exemplo, a parte “C” é referente ao tipo de dado.

Além disso, é gerado automaticamente um arquivo de catálogo com extensão “.dsc”, que atua como um índice de referência para todos os pathnames armazenados. Esse catálogo contém a lista completa dos registros presentes no arquivo, incluindo informações como o pathname de cada série, a data e a hora da última atualização e o nome do programa responsável pela gravação dos dados. De acordo com (USACE, 2024), o catálogo é geralmente organizado em ordem alfabética com base nas partes dos pathnames, o que facilita a localização e a gestão das séries dentro do ambiente *HEC-DSSVue*. Cada registro recebe ainda um identificador e um número de referência, que podem ser utilizados em substituição ao pathname completo em determinadas operações do sistema, agilizando consultas e manipulações de dados.

O primeiro arquivo DSS, foi obtido utilizando o algoritmo de integração com o Web-Service da ANA, esse arquivo denominado por “passoDasPedras.dss”, possui tamanho de 2,67 MB, contendo dados de chuva e de nível; no sistema DSS, a precipitação é denominada *PRECIP-INC* e o nível, *STAGE*. O algoritmo organizou a série histórica compreendida entre 4 de julho de 2014 e 29 de dezembro de 2025. A Figura 13 apresenta o catálogo das variáveis no HEC-DSSVue.

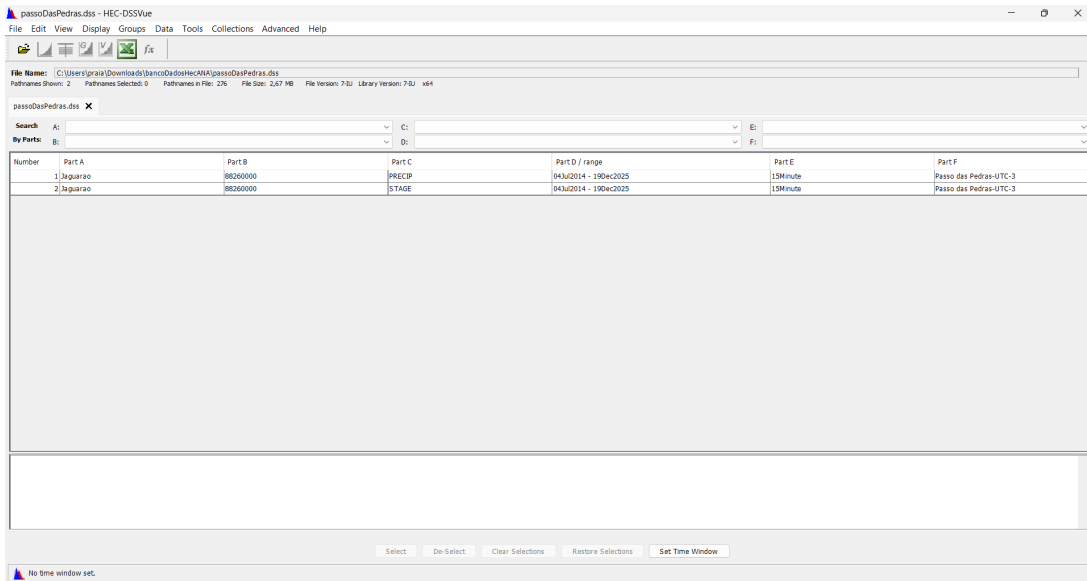


Figura 13: Catálogo do conjunto de dados da estação 88260000

Fonte: Elaboração própria.

Nessa figura é possível visualizar os dois caminhos que armazenam a série histórica das variáveis precipitação e nível. Além disso, é possível analisar a série histórica de cada variável por meio de gráficos ou tabelas; neste trabalho, optou-se pela apresentação gráfica. A Figura 14 apresenta o comportamento da variável *PRECIP-INC*.

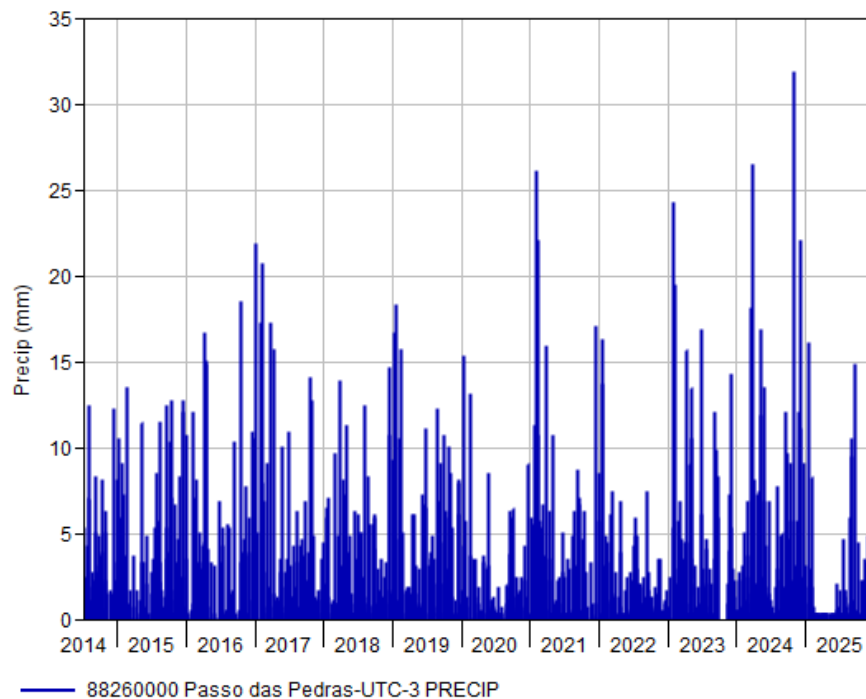


Figura 14: Plotagem da variável *PRECIP-INC*

Fonte: Elaboração própria.

Em continuidade, é apresentado o gráfico da variável *STAGE*, no qual é possível observar toda a série temporal da variável nível em intervalos regulares de 15 minutos ao longo dos anos. A Figura 15 apresenta a visualização gráfica dessa variável.

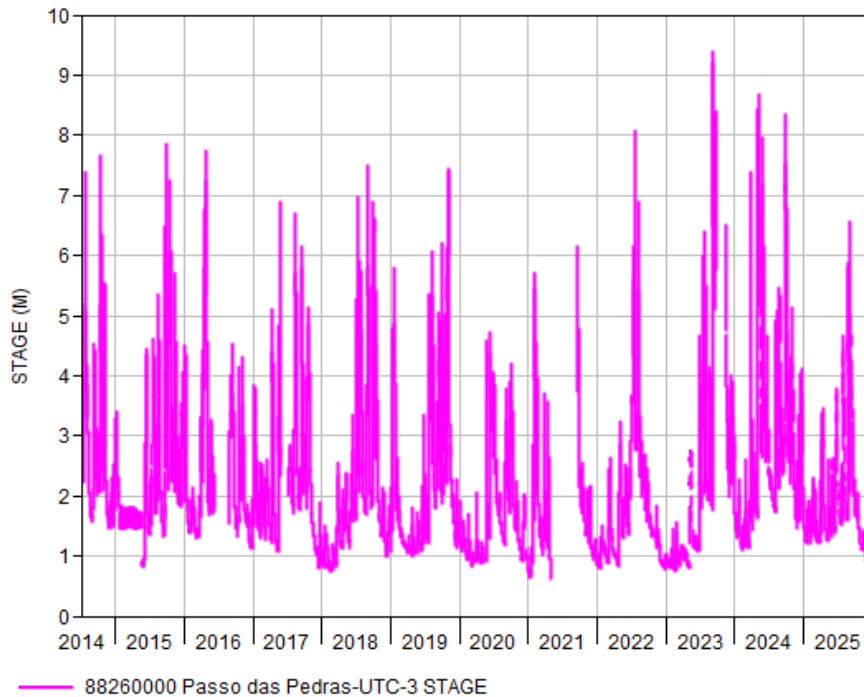


Figura 15: Plotagem da variável *STAGE*

Fonte: Elaboração própria.

O segundo arquivo DSS, denominado "G2-Centro.dss", foi obtido por meio do algoritmo de aquisição de dados do CEMADEN. Esse arquivo contém apenas dados de *PRECIP-INC*. Nesse caso, o algoritmo organizou a série histórica com início em 30 de setembro de 2017 e término em 19 de outubro de 2025. Além disso, os dados estão organizados em intervalos regulares de 10 minutos. Na Figura 16, é possível visualizar o catálogo com a variável correspondente.

Number	Part A	Part B	Part C	Part D / range	Part E	Part F
1	G2-Centro	430130501A	PRECIP-INC	30Sep2017 - 19Oct2025	10Minute	CENTRO-ARROIO GRANDE-UTC-3

Figura 16: Catálogo da estação 430130501A

Fonte: Elaboração própria.

O terceiro arquivo DSS, denominado “INMET.dss”, foi gerado por meio do algoritmo de integração com a API do INMET. O arquivo contém informações hidrometeorológicas das estações A827 e A836. O algoritmo organizou as séries históricas abrangendo o período de 4 de janeiro de 2007 até 17 de setembro de 2025. Além disso, as informações estão estruturadas em intervalos regulares de 1 hora. Ademais, o número de variáveis é maior, pois as duas estações são meteorológicas; cada estação possui um total de oito variáveis, totalizando 16 variáveis armazenadas na estrutura DSS. Na Figura 17, é possível visualizar essas variáveis.

Number	Part A	Part B	Part C	Part D / range	Part E	Part F
1	MRRM-SG	A827	GUST	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
2	MRRM-SG	A827	HUMIDITY	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
3	MRRM-SG	A827	PRECIP-INC	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
4	MRRM-SG	A827	PRESSURE	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
5	MRRM-SG	A827	RAD_KJ	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
6	MRRM-SG	A827	RAD_W	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
7	MRRM-SG	A827	TEMP	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
8	MRRM-SG	A827	WINDSPEED	04Jan2007 - 13Aug2025	1Hour	BASE-UTC-3
9	MRRM-SG	A836	GUST	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3
10	MRRM-SG	A836	HUMIDITY	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3
11	MRRM-SG	A836	PRECIP-INC	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3
12	MRRM-SG	A836	PRESSURE	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3
13	MRRM-SG	A836	RAD_KJ	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3
14	MRRM-SG	A836	RAD_W	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3
15	MRRM-SG	A836	TEMP	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3
16	MRRM-SG	A836	WINDSPEED	10Jan2007 - 17Sep2025	1Hour	JAGUARAO-UTC-3

Figura 17: Catálogo do conjunto de dados das estações A827 e A836

Fonte: Elaboração própria.

Na Figura 17, é possível notar as séries históricas das variáveis *GUST*, *HUMIDITY*, *PRECIP-INC*, *PRESSURE*, *RAD-kj*, *RAD-w*, *TEMP* e *WINDSPEED*. Cabe destacar

que os nomes das variáveis estão de acordo com a descrição exigida pelo sistema de armazenamento de dados HEC-DSSVue.

Esses foram os três arquivos gerados a partir dos algoritmos desenvolvidos para aquisição e organização automatizada no sistema HEC-DSSVue. Entretanto, para a aplicação do modelo LSTM, foi utilizado apenas o arquivo "passoDasPedras.dss". Apesar de ter sido utilizado apenas esse arquivo, o ganho de tempo na aquisição de dados, não só dessa estação, mas também das demais utilizadas, demonstra a relevância desses três algoritmos em uma situação na qual seja exigida eficiência no gerenciamento dessas dados.

Nesse momento será avaliado o modelo LSTM desenvolvido para prever nível máximo diário, como foi detalhado construção da arquitetura utilizou-se 70% do conjunto total dos dados, esse bloco compreende o período de 04/07/2014 a 11/07/2022. O bloco subsequente, destinado à validação, abrange o intervalo de 12/07/2022 a 31/03/2024. Por fim, a faixa de dados compreendida entre 01/04/2024 e 19/12/2025 foi reservada exclusivamente para a etapa de teste do modelo.

Para avaliar o desempenho desse modelo, utilizou-se as métricas MAE, RMSE e NSE ambas descritas na seção de desenvolvimento do modelo de rede neural. Além disso, os cálculos dessas métricas foram feitos separadamente para os conjuntos de treino, validação e teste, permitindo avaliar tanto a capacidade de ajuste do modelo aos dados utilizados no treinamento quanto sua habilidade de generalização para dados não vistos.

De acordo com Burzykowski et al. (2023), a fase de treinamento é utilizada para ajustar diferentes modelos, seja com diferentes estruturas ou complexidades. O erro nessa etapa tende a ser otimista demais, visto que o modelo é avaliado nos mesmos dados que utilizou para aprender. O processo de validação é empregado para selecionar, entre os modelos já treinados, aquele que apresenta o menor erro de predição. Por fim, tem-se a fase de teste, que é mantida totalmente separada das demais etapas. Nessa etapa, com o modelo definido de acordo com a validação, ele é aplicado ao conjunto de teste com o propósito de medir o erro real de generalização, isto é, o desempenho do modelo em dados nunca vistos.

Referente ao bloco de treinamento, a arquitetura apresentou RMSE igual a 0,27 m e MAE de 0,15 m. Esses valores apontam que o erro médio absoluto é bastante reduzido entre os níveis máximos diários observados e preditos. A métrica NSE obtida foi de 0,9377, valor considerado bom segundo os critérios de classificação dessa métrica. De acordo com Suryaningtyas et al. (2020), $NSE < 0,36$ indica resultado não qualificado; $0,36 < NSE < 0,75$ corresponde a resultado qualificado; e $NSE > 0,75$ é classificado como bom.

No conjunto de validação, a LSTM apresentou RMSE de 0,32 m e MAE de 0,17 m. Em relação ao conjunto de treinamento, houve um discreto aumento nos valores des-

sas métricas. Entretanto, para a métrica NSE, obteve-se o valor de 0,9661, indicando desempenho ainda mais elevado do que na etapa de treinamento. Esse resultado sugere boa estabilidade do modelo durante o processo de ajuste de hiperparâmetros, sem indícios relevantes de overfitting.

Por fim, no bloco de teste, o modelo LSTM apresentou RMSE de 0,38 m e MAE de 0,20 m. Embora tenha ocorrido um discreto aumento dos erros em relação aos demais conjuntos, os valores permanecem baixos em termos absolutos. A métrica NSE foi igual a 0,9065, mantendo-se acima de 0,75, o que indica excelente capacidade preditiva, segundo a literatura. Na Tabela 3, é apresentado o resumo das métricas em cada conjunto.

Tabela 4: Síntese comparativa das métricas de erro nos conjuntos de treino, validação e teste

Medida de desempenho	Treino	Validação	Teste
RMSE (m)	0,27	0,32	0,38
MAE (m)	0,15	0,17	0,20
NSE	0,9377	0,9661	0,9065

Observa-se que o modelo LSTM apresenta desempenho consistente entre os conjuntos de validação e teste, indicando adequada capacidade de generalização. A proximidade entre os valores de RMSE e MAE nos diferentes conjuntos sugere ausência de sobreajuste, uma vez que não há aumento abrupto dos erros quando o modelo é aplicado a dados não utilizados no processo de treinamento.

Um outro detalhe importante está relacionado à métrica NSE. Nota-se que o valor dela no conjunto de validação foi ligeiramente superior aos valores obtidos nos conjuntos de treino e teste. Esse comportamento pode ocorrer em função das características estatísticas específicas de cada subconjunto de dados, especialmente no que se refere à variabilidade da série observada. Nesse caso, o conjunto de validação apresenta uma série com comportamento mais suavizado, com menor ocorrência de picos extremos, o que pode favorecer o desempenho do modelo e, conseqüentemente, resultar em um valor de NSE ligeiramente superior em comparação aos demais conjuntos.

Além disso, a persistência de valores elevados da métrica NSE nos três conjuntos reforça que o modelo conseguiu preservar a variabilidade da série observada, mantendo elevada eficiência preditiva mesmo fora do conjunto de ajuste. Esse comportamento é indicativo de estabilidade estrutural da arquitetura adotada e coerência na escolha dos hiperparâmetros.

O modelo foi treinado por até 40 épocas em ambiente computacional local, utilizando um processador Intel® Core™ i7-9750H (2,60 GHz), com 16 GB de memória RAM e arquitetura de 64 bits. Durante o processo de treinamento, a melhor configuração foi obtida na 21ª época, momento em que se registrou o menor valor de perda no conjunto de validação, indicando o ponto de melhor capacidade de generalização do modelo. A Figura 18 apresenta esse histórico de perda, conhecido como loss, ao longo das fases de treinamento e validação.

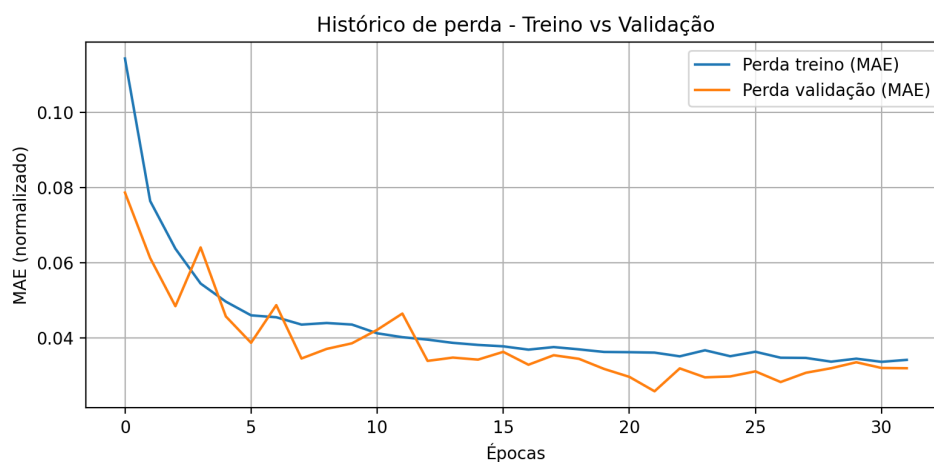


Figura 18: Histórico de perda no treino e validação

Fonte: Elaboração própria.

Na figura 18, observa-se que o treinamento termina por volta das épocas 30 e 31. Embora o número máximo de épocas fosse 40, o treinamento foi interrompido antecipadamente pelo mecanismo *Early Stopping*, o qual monitora o loss no conjunto de validação. Como não houve melhora relevante após aproximadamente a 21ª época, o algoritmo interrompeu automaticamente o treinamento após 10 épocas consecutivas sem ganho adicional, evitando um possível *overfitting*.

Dando continuidade à análise, na Figura 19 é apresentado o gráfico que realiza a comparação entre os níveis máximos diários observados e aqueles estimados pela rede LSTM na fase de treinamento. Nota-se uma elevada concordância entre as duas séries temporais ao longo de todo o período utilizado no treino. É possível visualizar como a curva laranja, correspondente aos valores previstos, acompanha de forma consistente a dinâmica da série real, reproduzindo tanto os períodos de estiagem quanto os eventos de cheia, inclusive os picos mais pronunciados.

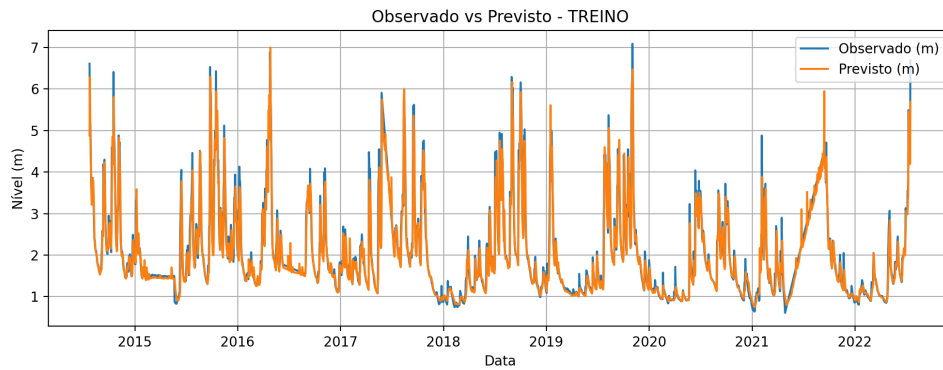


Figura 19: Comparação entre os valores observados e previstos do nível máximo diário (conjunto de treinamento)

Fonte: Elaboração própria.

De maneira geral, a forte sobreposição entre as curvas confirma o elevado desempenho indicado pelas métricas RMSE, MAE e NSE que foram apresentadas anteriormente, reforçando a capacidade que o modelo possui em capturar a estrutura temporal e as não linearidades presentes na série durante o treinamento.

Na Figura 20 é apresentada a mesma comparação, mas, diferentemente do conjunto de treinamento, essa fase contempla dados que não foram utilizados diretamente no ajuste dos pesos, permitindo avaliar a capacidade de generalização do modelo. Além disso, é possível observar-se que a curva prevista, destacada pela cor laranja, acompanha de forma bastante próxima a dinâmica da curva azul, que representa a série observada.

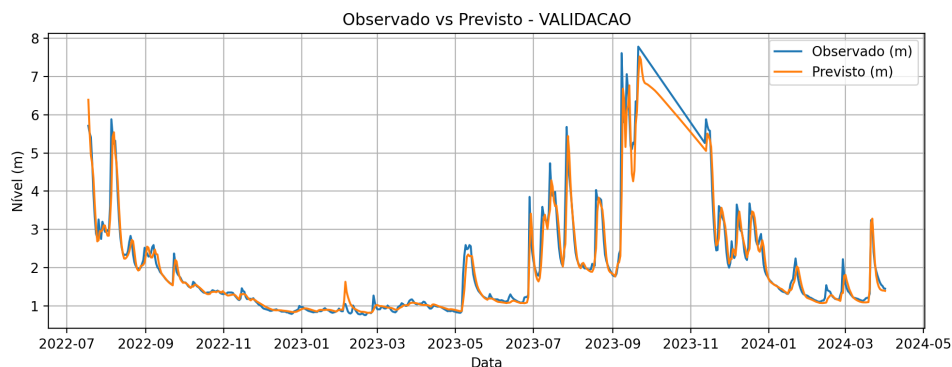


Figura 20: Comparação entre os valores observados e previstos do nível máximo diário (conjunto de validação)

Fonte: Elaboração própria.

Ainda analisando a Figura 19, nota-se que, em picos maiores, especialmente entre meados e o final de 2023, o modelo tende a suavizar levemente os valores máximos, subestimando discretamente alguns extremos. Isto é, em momentos em que a variável

observada apresenta valores muito abruptos, o modelo tende a prever valores um pouco menores que os realmente registrados, acompanhando a tendência de subida, mas reduzindo levemente a magnitude do extremo.

Nota-se ainda que no período de 2023-09 a 2023-11 o erro do observado com o previsto foi maior em relação aos demais períodos, isso ocorre pelo de nesse período ter uma número maior de falhas e foi nesse período de foi aplicado a interpolação, isso pode ter afetado o desempenho do modelo LSTM.

A Figura 21 mostra, assim como nos gráficos anteriores, a comparação entre os níveis máximos diários observados com os previstos pelo modelo LSTM, sendo que, neste caso, a análise refere-se ao conjunto de teste. Observa-se elevada aderência entre as séries observadas e previstas, com adequada representação das oscilações e dos eventos de pico. Esse comportamento visual é consistente com os indicadores estatísticos obtidos nesta etapa, cujos valores corroboram a boa capacidade de generalização do modelo para dados não utilizados no treinamento.

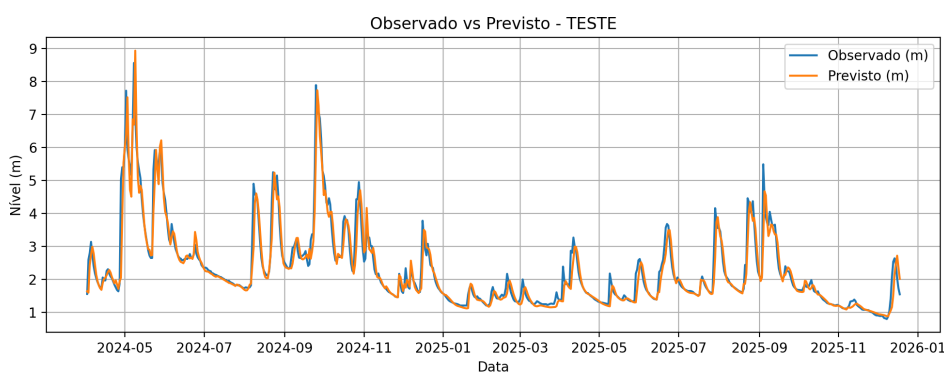


Figura 21: Comparação entre os valores observados e previstos do nível máximo diário (conjunto de teste)

Fonte: Elaboração própria.

A estabilidade das métricas no conjunto de teste reforça a robustez do modelo. Caso houvesse *overfitting*, seria observado uma degradação significativa no desempenho fora do conjunto de treinamento, com aumento expressivo de RMSE e MAE e redução substancial da NSE.

Entretanto, a diferença relativamente pequena entre os erros obtidos nos três conjuntos indica que o modelo conseguiu aprender padrões estruturais da série temporal, e não apenas memorizar oscilações específicas do período de treinamento. Esse comportamento confirma que o uso do *Early Stopping* contribuiu para interromper o processo de ajuste no ponto de melhor capacidade de generalização.

Nota-se também que os maiores desvios entre valores observados e previstos concentram-se em períodos de rápida elevação do nível, característicos de eventos hidrológicos intensos. Tal comportamento é esperado, uma vez que transições abruptas

tas tendem a amplificar incertezas associadas à modelagem baseada exclusivamente em dados históricos (Xie et al., 2021).

Eventos extremos normalmente apresentam dinâmica não linear mais acentuada e forte dependência de fatores externos instantâneos, como intensidade e distribuição espacial da precipitação. Como o modelo utiliza apenas informações passadas das variáveis observadas, a capacidade de antecipar variações abruptas pode apresentar leve suavização dos picos, fenômeno visualmente identificado nos gráficos apresentados anteriormente.

Apesar do desempenho satisfatório do modelo em cada conjunto, sua aplicação restringiu-se a uma única estação hidrométrica, limitando inferências espaciais. Dessa forma, não foi investigada sua capacidade de generalização para outras seções da própria bacia hidrográfica, as quais podem apresentar dinâmicas hidrológicas locais distintas.

Outra limitação do modelo está relacionada às variáveis de entrada. Não foram incorporadas previsões meteorológicas futuras como informações exógenas ao sistema. Dessa forma, o horizonte temporal de aplicação permanece condicionado ao padrão histórico das séries observadas, restringindo o potencial de antecipação de eventos extremos com base em cenários meteorológicos previstos.

5 CONCLUSÃO

Esta dissertação demonstrou que a integração entre uma infraestrutura automatizada de aquisição de dados hidrometeorológicos e a modelagem preditiva baseada em Deep Learning constitui uma abordagem viável e promissora para aplicações de previsão hidrológica de curto prazo. Nesse contexto, a utilização de arquivos binários no formato “dss” evidencia uma forma automatizada e estruturada de armazenamento de informações hidrometeorológicas por meio do sistema HEC-DSSVue.

Dessa forma, em situações operacionais nas quais seja necessário utilizar o HEC-DSSVue como banco de dados, a infraestrutura desenvolvida pode apoiar profissionais que demandam grandes volumes de dados hidrometeorológicos. Embora esses dados estejam disponíveis publicamente, encontram-se distribuídos em diferentes bases institucionais, o que dificulta sua integração. Assim, o sistema proposto contribui para aumentar a agilidade na preparação dos dados de entrada, tanto para modelos baseados em dados quanto para modelos baseados em processos físicos.

No que se refere à modelagem preditiva, a aplicação da arquitetura LSTM mostrou-se adequada para capturar as dependências temporais presentes nas séries hidrológicas de nível máximo diário. Os resultados obtidos indicaram bom desempenho preditivo, evidenciado pelas métricas RMSE, MAE e NSE, demonstrando a capacidade do modelo de representar as relações existentes entre as séries de precipitação diária e nível máximo diário.

Observou-se também que os maiores desvios ocorreram em períodos de variação abrupta do nível, comportamento compatível com a natureza altamente não linear dos eventos hidrológicos extremos. Essa constatação reforça que, embora modelos baseados exclusivamente em dados apresentem desempenho satisfatório, a previsão de extremos ainda representa um desafio inerente à complexidade do sistema hidrológico.

Os resultados obtidos corroboram estudos recentes da literatura, discutidos na fundamentação teórica, que apontam o potencial das redes LSTM para previsão de níveis de rios, especialmente em cenários nos quais a modelagem baseada em processos físicos apresenta elevada complexidade computacional e maior demanda de parame-

trização.

Por fim, a principal contribuição do trabalho reside na proposição de um fluxo integrado e tecnicamente replicável, que articula aquisição automatizada, organização e armazenamento estruturado de dados com modelagem preditiva orientada por dados. Dessa forma, o estudo não apenas confirma o potencial do uso de redes neurais na previsão hidrológica, mas também estabelece uma base operacional consistente para futuras aplicações em monitoramento e alerta hidrológico em tempo quase real.

6 PERSPECTIVAS FUTURAS

Como perspectiva futura desse trabalho, recomenda-se expandir a aplicação do modelo para múltiplas estações da bacia, assim fazendo a utilização da infraestrutura automatizada desenvolvida neste trabalho para contemplar uma base de dados maior, integrando um número maior de estações hidrometeorológicas e incorporando novas fontes de dados. A consolidação de uma base de dados mais abrangente poderá permitir o desenvolvimento de modelos com maior capacidade de generalização espacial, além de possibilitar a avaliação da robustez do sistema em diferentes contextos hidrológicos.

Outra direção a ser explorada está relacionada ao aprimoramento do modelo preditivo. Investigações futuras podem avaliar arquiteturas mais complexas, como modelos híbridos que combinem LSTM com camadas convolucionais (CNN-LSTM), mecanismos de atenção (Attention Mechanisms) ou ainda abordagens baseadas em Transformers aplicadas a séries temporais, buscando capturar de forma mais eficiente dependências espaciais e temporais presentes nos dados.

Por fim, destaca-se a importância da operacionalização completa do sistema em ambiente de tempo real, integrando a aquisição automática dos dados, o processamento no HEC-DSSVue e a geração automática de previsões com emissão de alertas. A incorporação de um módulo de avaliação contínua do desempenho do modelo, com atualização periódica dos parâmetros à medida que novos dados sejam incorporados, poderá transformar a infraestrutura proposta em um sistema adaptativo de suporte à decisão, contribuindo efetivamente para a mitigação de riscos hidrológicos e para o fortalecimento das estratégias de alerta antecipado.

REFERÊNCIAS

- ABREU, A. O. d. **Utilização de redes neurais de memória de longo e curto prazo (LSTM) para previsão da arrecadação mensal de receita orçamentária em Cajazeiras, estado da Paraíba.** 2024. Trabalho de Conclusão de Curso (Graduação em Análise e Desenvolvimento de Sistemas) — Instituto Federal da Paraíba (IFPB), Cajazeiras, PB.
- ADAMO, N. et al. Review of climate change impacts on human environment: past, present and future projections. **Engineering**, v.13, n.11, p.605–630, 2021.
- ADAMOWSKI, J.; SUN, K. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. **Journal of Hydrology**, v.390, n.1, p.85–91, 2010.
- AGOSTINHO, E. D. **Formação de carteira de ativos de renda variável usando técnicas computacionais para tratamento de séries temporais.** 2025. Trabalho de Conclusão de Curso (Graduação em Engenharia de Controle e Automação), Universidade Federal de Santa Catarina. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/265903>>. Acesso em: 2026-03-10.
- AHMADZADEH, E. et al. A Deep Bidirectional LSTM-GRU Network Model for Automated Ciphertext Classification. **IEEE Access**, v.10, p.3228–3237, 2022.
- ALVES, J. M. et al. Evaluation of the AR4 CMIP3 and the AR5 CMIP5 Model and Projections for Precipitation in Northeast Brazil. **Frontiers in Earth Science**, v.4, p.44, 2016.
- AMARAL, L. S. **Predição de demanda de energia elétrica de curto-prazo empregando séries temporais.** 2025. Tese de Doutorado — Universidade Federal de Santa Catarina, Florianópolis.
- AMPAS, H. et al. Hybrid Hydrological Forecasting Through a Physical Model and a Weather-Informed Transformer Model: A Case Study in Greek Watershed. **Applied Sciences**, 2025.

ANA. **Descrição para disponibilizar dados hidrometeorológicos dos sistemas Telemetria 1 e HIDRO**. 2013. Atualizado em junho de 2013. Implementado em maio de 2011.

ANA. **Apresentação**. Disponível em: <<https://www.snirh.gov.br/hidroweb/apresentacao>>. Acesso em: 06 mar. 2026.

ANA. **Política Nacional de Recursos Hídricos**. Disponível em: <<https://www.gov.br/ana/pt-br/acesso-a-informacao/institucional>>. Acesso em: 06 mar. 2026.

ASSEM, H. et al. Urban Water Flow and Water Level Prediction Based on Deep Learning. In: **Machine Learning and Knowledge Discovery in Databases**. 2017. p.317–329.

ATASHI, V. et al. Water Level Forecasting Using Deep Learning Time-Series Analysis: A Case Study of Red River of the North. **Water**, v.14, n.12, 2022.

BACELAR, L. C. S. D. et al. Limiares de chuva deflagradores de inundações bruscas: metodologia, aplicação e avaliação em ambiente operacional. **Revista DAE**, v.68, n.221, p.71–86, 2020.

BARROS, S. N. M.; SILVA, B. C. d. Recursos Hídricos em Mudanças Climáticas: Uma Análise da Abordagem no Brasil. **Congresso Nacional de Meio Ambiente**, Poços de Caldas, MG, Brasil, 2023.

BETANCUR, M. O. **Análise da silvicultura na bacia hidrográfica do Rio Jaguarão: diagnóstico para gestão geoambiental em bacia hidrográfica do Rio Jaguarão**. 2015. 165p. Dissertação (Mestrado em Geografia) — Universidade de Brasília, Instituto de Ciências Humanas, Brasília.

BONAKDARI, H. et al. Lake water-level fluctuations forecasting using minimax probability machine regression, relevance vector machine, Gaussian process regression, and extreme learning machine. **Water Resources Management**, v.33, n.11, p.3965–3984, 2019.

BRAVO, N. et al. Estudo dos Processos Fluviais com Aplicação de Sistemas de Informação Geográfica. **Congresso Internacional sobre o Rio Minho 2004: Uma Nova Cultura da Água para o Rio Minho**, Melgaço, Portugal, 2004.

BURZYKOWSKI, T. et al. Validation of machine learning algorithms. **American journal of orthodontics and dentofacial orthopedics**, v.164, n.2, p.295–297, 2023.

CANDIDO, L. et al. **Previsão do nível da água do Rio Mundaú em Rio Largo com Redes Neurais Profundas**. 2025. 344–347p. Disponível em: <<https://sol.sbc.org.br/index.php/wcama/article/view/36113>>.

CASAGRANDE, V. M. Uma introdução à análise de dados usando pandas, matplotlib e seaborn. **Anais da Escola Regional de Informática de Mato Grosso (ERI-MT)**, 2024.

CASTRO, C. C. G. e. **Criação de dashboards via processamento de linguagem natural**. Disponível em: <<https://hdl.handle.net/11449/258579>>.

CEMADEN. **Manual de utilização de sistema**: Plataforma de Entrega de Dados (PED). 2023.

CHIEU, T. Q. et al. Prediction of the water level at the Kien Giang River based on regression techniques. **River**, v.3, n.1, p.59–68, 2024.

LANDERS, B.; SCHREIBER, D. (Ed.). **An Introduction to APIs**. Zapier, Inc., 2014. Cover art by Stephanie Briones.

DAI, Z. et al. A Hydrological Data Prediction Model Based on LSTM with Attention Mechanism. **Water**, v.15, n.4, 2023.

DEVIA, G. K. et al. A review on hydrological models. **Aquatic procedia**, v.4, p.1001–1007, 2015.

FAKHRUDDIN, B. S. et al. Benefits of economic assessment of cyclone early warning systems-A case study on Cyclone Evan in Samoa. **Progress in Disaster Science**, v.2, p.100034, 2019.

FENG, L. Predicting output responses of nonlinear dynamical systems with parametrized inputs using LSTM. **IEEE Journal on Multiscale and Multiphysics Computational Techniques**, v.8, p.97–107, 2023.

FÜRST, J. et al. Rosalia: an experimental research site to study hydrological processes in a forest catchment. **Earth System Science Data Discussions**, v.2020, p.1–18, 2020.

GOLDING, B. **Towards the “perfect” weather warning**: Bridging disciplinary gaps through partnership and communication. Springer Nature, 2022.

GONÇALVES, V. T. **Introdução à criação de APIs**: conceitos fundamentais e implementação em linguagem Python. Universidade Federal de Ouro Preto. Disponível em: <<https://www.monografias.ufop.br/handle/35400000/8436>>. Acesso em: 2026-03-08.

GOODARZI, M. R. et al. Evaluation of a weather forecasting model and HEC-HMS for flood forecasting: case study of Talesh catchment. **Applied Water Science**, v.14, n.2, p.34, 2024.

GRÜNDEMANN, G. J. et al. Extreme precipitation return levels for multiple durations on a global scale. **Journal of Hydrology**, v.621, p.129558, 2023.

HE, J. et al. Deep learning enables super-resolution hydrodynamic flooding process modeling under spatiotemporally varying rainstorms. **Water Research**, v.239, p.120057, 2023.

HIEU, T. T. et al. Water level prediction using deep learning models: A case study of the Kien Giang River, Quang Binh Province. **River**, v.2, n.4, p.468–479, 2023.

HUANG, J. et al. Short-term prediction method of reservoir downstream water level under complicated hydraulic influence. **Water Resources Management**, v.37, n.11, p.4475–4490, 2023.

HUANG, S. et al. Coupling Machine Learning Into Hydrodynamic Models to Improve River Modeling With Complex Boundary Conditions. **Water Resources Research**, v.58, n.10, p.e2022WR032183, 2022.

INMET. **Sobre o INMET**. Disponível em: <<https://portal.inmet.gov.br/sobre>>. Acesso em: 26 out. 2025.

IPCC. **Scientific Assessment of Climate Change**: The Policymakers' Summary of the Report of Working Group I to the Intergovernmental Panel on Climate Change. World Meteorological Organization/United Nations Environment Programme, 1990.

IPCC. Human Settlements, Infrastructure, and Spatial Planning. In: SETO, K. C.; DHAKAL, S.; BIGIO, A. G.; BLANCO, H.; DELGADO, G. C.; DEWAR, D.; HUANG, L.; INABA, A.; KANSAL, A.; LWASA, S.; MCMAHON, J. E.; MUELLER, D.; MURAKAMI, J.; NAGENDRA, H.; RAMASWAMI, A. (Ed.). **Climate Change 2014**: Mitigation of Climate Change: Working Group III Contribution to the IPCC Fifth Assessment Report. 2014. p.923–1000.

IPCC. **Mudança do Clima 2023**: Relatório Síntese. Contribuição dos Grupos de Trabalho I, II e III para o Sexto Relatório de Avaliação do Painel Intergovernamental sobre Mudança do Clima. 2023. Acesso em: 08 mar. 2026. Disponível em: <<https://www.ipcc.ch/report/ar6/syr/>>.

IYIOLA, A. O. et al. Climate change and water crisis in the global south. In: **Water Crises and Sustainable Management in the Global South**. Springer, 2024. p.111–140.

KHAN, A. et al. Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends. **IEEE Access**, v.11, p.94371–94391, 2023.

KIM, C.-S. et al. Quasi-Optimized LSTM Approach for River Water Level Forecasting. **Water**, v.17, n.14, 2025.

KIM, D. et al. Improvement of Deep Learning Models for River Water Level Prediction Using Complex Network Method. **Water**, v.14, n.3, 2022.

LARA-BENÍTEZ, P. et al. An experimental review on deep learning architectures for time series forecasting. **International journal of neural systems**, v.31, n.03, p.2130001, 2021.

LESK, C. et al. Influence of extreme weather disasters on global crop production. **Nature**, v.529, n.7584, p.84–87, 2016.

LI, G. et al. A combined hydrodynamic model and deep learning method to predict water level in ungauged rivers. **Journal of Hydrology**, v.625, p.130025, 2023.

LI, S. et al. Independently Recurrent Neural Network (IndRNN): Building a Longer and Deeper RNN. **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, Salt Lake City, UT, USA, p.5457–5466, 2018.

LIAO, J. et al. Numerical Simulation and Forecasting of Water Level for Qinghai Lake Using Multi-Altimeter Data Between 2002 and 2012. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v.7, n.2, p.609–622, 2014.

LUKONG, T. K. et al. Predictive Model of Rainfall-Runoff: A Case Study of the Sanaga Basin at Bamendjin Watershed in Cameroon. **Energy and Environment Research**, v.1, n.1, p.193, 2011.

LUO, J. et al. Classification-enhanced LSTM model for predicting river water levels. **Journal of Hydrology**, v.650, p.132535, 2025.

MACHADO, J. A. **Análise comparativa de modelos de previsão do nível da água no rio Guaíba**. 2025. Trabalho de Conclusão de Curso (Graduação) — Universidade Federal do Rio Grande do Sul, Porto Alegre.

MARENGO, J. A. et al. O maior desastre climático do Brasil: Chuvas e inundações no estado do Rio Grande do Sul em abril-maio 2024. **Estudos Avançados**, v.38, n.112, p.203–228, 2024.

MARTINS, D. W. d. P. **Inteligência artificial aplicada em séries temporais**. 2024. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo.

MATTIUZI, C. D. P. et al. **Relatório de operação do Sistema de Alerta Hidrológico da bacia do Rio Uruguai 2023**. Serviço Geológico do Brasil-CPRM, 2023.

MEDEIROS, M. S. **Previsão de vazão e energia natural afluyente**: um estudo de caso avaliando métricas de desempenho. 2022. Trabalho de Conclusão de Curso — Universidade Federal do Rio Grande do Sul, Porto Alegre.

MIERES, L. d. S. **Aplicação da Rede Neural LSTM para Predição de Nível do Rio Caí/RS**. 2024. Trabalho de Conclusão de Curso (MBA em Inteligência Artificial e Big Data) — Universidade de São Paulo (USP), São Carlos.

MOHAPATRA, M. et al. Cyclone Warning System in India: A journey of success over 25 years. **Weather and Forecasting**, v.40, n.6, p.829–855, 2025.

NAJAFI, H. et al. High-resolution impact-based early warning system for riverine flooding. **Nature communications**, v.15, n.1, p.3726, 2024.

OLSEN, N. R. B. et al. Three-dimensional numerical flow modelling for estimation of spillway capacity. **Journal of Hydraulic Research**, v.36, n.5, p.775–784, 1998.

OMBADI, M. et al. A warming-induced reduction in snow fraction amplifies rainfall extremes. **Nature**, v.619, n.7969, p.305–310, 2023.

PALASH, W. et al. A data-driven global flood forecasting system for medium to large rivers. **Scientific Reports**, v.14, n.1, p.8979, 2024.

Pandas Development Team. **pandas.DataFrame.interpolate**. Disponível em: <<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html>>. Acesso em: 2026-03-04.

PAPACHARALAMPOUS, G. et al. Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. **Journal of Hydrology**, v.590, p.125205, 2020.

PERRON, B. E. et al. Demystifying Application Programming Interfaces (APIs): Unlocking the Power of Large Language Models and Other Web-Based AI Services in Social Work Research. **Journal of the Society for Social Work and Research**, v.16, n.2, p.275–294, 2025.

PRUDHOMME, C. et al. Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales. **Hydrological Sciences Journal**, v.62, n.16, p.2753–2768, 2017.

Python. **logging — Logging facility for Python**. Disponível em: <<https://docs.python.org/3/library/logging.html>>. Acesso em: 6 nov. 2025.

REINKING, J. GNSS-SNR water level estimation using global optimization based on interval analysis. **Journal of Geodetic Science**, v.6, n.1, 2016.

RIBEIRO, L. d. S. Análise da relação entre a ocorrência de deslizamentos e os atributos do relevo no evento extremo de precipitação no sul do Brasil, maio de 2024. **Sociedade & Natureza**, 2025.

ROLIM, L. Z. R. et al. Machine learning strategies for multiannual rainfall prediction and drought early warning: insights from Ceara, Brazil. **Natural Hazards**, v.121, n.6, p.7229–7263, 2025.

SAITO, S. M. et al. Sistema de monitoramento e alertas de desastres naturais: práticas e desafios. **Eventos extremos no Rio Grande do Sul: inundações e movimentos de massa**, p.179–192, 2013.

SANTOS, G. B. d. **Balço hídrico quantitativo das águas superficiais na bacia hidrográfica do Rio Jaguarão**. 2022. 117p. Dissertação (Mestrado em Recursos Hídricos) — Universidade Federal de Pelotas, Pelotas.

SILVA, G. M. A obtenção de dados de precipitação nos principais repositórios nacionais brasileiros. **Contribuciones a las Ciencias Sociales**, v.18, n.2, p.44, 2025.

SILVA, J. M. A. et al. O USO DA LINGUAGEM PYTHON APLICADA À PERÍCIA FINANCEIRA. **Revista Gestão e Conhecimento**, v.19, n.1, p.e412–e412, 2025.

SILVA, M. E. da et al. APPLICABILITY OF HEC-DSSVue FOR MANAGEMENT OF HYDROLOGICAL GAGING NETWORK: A CASE STUDY USING SUB-HOURLY RAINFALL DATASETS FROM SOUTHERN BRAZIL. **Engenharia Agrícola**, v.45, p.e20240143, 2025.

SOUZA, A. et al. Flood prediction in Brazil: A review of trends, gaps and methodological perspectives. **Journal of South American Earth Sciences**, v.164, p.105637, 2025.

SURYANINGTYAS, L. S. et al. Hydrological Analysis of TRMM (Tropical Rainfall Measuring Mission) Data in Lesti Sub Watershed. **Civil and Environmental Science Journal (CIVENSE)**, v.3, n.1, p.18–30, Feb. 2020.

SUSHANTH, K. et al. Near-real-time forecasting of reservoir inflows using explainable machine learning and short-term weather forecasts. **Stochastic Environmental Research and Risk Assessment**, v.37, n.10, p.3945–3965, 2023.

TAORMINA, R. et al. Neural network river forecasting through baseflow separation and binary-coded swarm optimization. **Journal of Hydrology**, v.529, p.1788–1797, 2015.

TENG, L. Y. H. **Future Weir Water Level Forecasting using Novel Neural Network Considering Climate Change Impacts**. 2024. Tese (Doutorado em Ciência da Computação) — Swinburne.

TensorFlow. **Keras: The High-Level API for TensorFlow**. Disponível em: <<https://www.tensorflow.org/guide/keras>>. Acesso em: 2026-02-12.

TESSEMA, B. H. et al. Dam breach analysis and flood inundation mapping of Dire Dam, using HEC-HMS and HEC-RAS models. **Sustainable Water Resources Management**, v.10, n.2, p.45, 2024.

TRAN, C. K. et al. Real-time flood forecasting using time-varying parameter hydrological model: case study for Ta Trach reservoir. **Applied Water Science**, v.15, n.7, p.152, 2025.

TZOUMPAS, K. et al. A Data Filling Methodology for Time Series Based on CNN and (Bi)LSTM Neural Networks. **IEEE Access**, v.12, p.31443–31460, 2024.

UNUCKA, J. et al. Application for the parametrization and automatic running of the HEC-HMS rainfall-runoff model. **Vodohospodářské technicko-ekonomické informace**, v.65, n.4, p.14–17, 2023.

USACE. **HEC Data Storage System Visual Utility Engine (HEC-DSSVue): User's Manual**. 2024.

VIEIRA, M. E. **Automação da Equoterapia por Meio de Web Services**. 2024. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação), Instituto Federal Goiano. Disponível em: <<https://repositorio.ifgoiano.edu.br/handle/prefix/4777>>. Acesso em: 2026-03-09.

VIZI, Z. et al. Water level prediction using long short-term memory neural network model for a lowland river: a case study on the Tisza River, Central Europe. **Environmental Sciences Europe**, v.35, n.1, p.92, 2023.

XIE, K. et al. Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. **Journal of Hydrology**, v.603, p.127043, 2021.

ZAHURA, F. T. et al. Training machine learning surrogate models from a high-fidelity physics-based model: Application for real-time street-scale flood prediction in an urban coastal community. **Water Resources Research**, v.56, n.10, p.e2019WR027038, 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.