UNIVERSIDADE FEDERAL DO AMAZONAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GEOCIÊNCIAS

CÍNTIA DE LIMA ELEUTÉRIO MENDES

INTEGRAÇÃO DE DADOS DE MÚLTIPLAS ESTAÇÕES PARA A PREVISÃO DE EVENTOS HIDROLÓGICOS EXTREMOS NA BACIA DO RIO NEGRO COM APRENDIZADO PROFUNDO

Manaus
2025

CÍNTIA DE LIMA ELEUTÉRIO MENDES

INTEGRAÇÃO DE DADOS DE MÚLTIPLAS ESTAÇÕES PARA A PREVISÃO DE EVENTOS HIDROLÓGICOS EXTREMOS NA BACIA DO RIO NEGRO COM APRENDIZADO PROFUNDO

Dissertação apresentada ao Programa de Pós-Graduação em Geociências da Universidade Federal do Amazonas, como requisito parcial para obtenção do Título de Mestre em Geociências, área de concentração de Geociências.

Orientador: Dr. Naziano Pantoja Filizola Júnior

Manaus
2025

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

M538i Mendes, Cíntia de Lima Eleutério
Integração de dados de múltiplas estações para a previsão de eventos hidrológicos extremos na Bacia do Rio Negro com aprendizado profundo / Cíntia de Lima Eleutério Mendes. - 2025.
138 f. : il., color. ; 31 cm.

Orientador(a): Naziano Pantoja Filizola Júnior.
Dissertação (mestrado) - Universidade Federal do Amazonas, Programa de Pós-Graduação em Geociências, Manaus, 2025.

1. Previsão hidrológica. 2. rede neural convolucional. 3. influências geológicas e climáticas. 4. eventos extremos. I. Filizola Júnior, Naziano Pantoja. II. Universidade Federal do Amazonas. Programa de Pós-Graduação em Geociências. III. Título

---

CÍNTIA DE LIMA ELEUTÉRIO MENDES

INTEGRAÇÃO DE DADOS DE MÚLTIPLAS ESTAÇÕES PARA PREVISÃO DE CHEIAS E SECAS NA BACIA DO RIO NEGRO UTILIZANDO REDES NEURAIS CONVOLUCIONAIS

Dissertação apresentada ao Programa de Pós-Graduação em Geociências da Universidade Federal do Amazonas, como requisito parcial para obtenção do Título de Mestre em Geociências, área de concentração em Geociências.

Aprovada em: 06 de março de 2025.

BANCA EXAMINADORA

Prof. Dr. Naziano Pantoja Filizola J., Presidente.
Universidade Federal do Amazonas, PPGGEO

Prof. Dr. Ingo Daniel Wahnfried, Membro.
Universidade Federal do Amazonas, PPGGEO

Prof. Dr. José Luiz de Souza Pio, Membro.
Universidade Federal do Amazonas, ICOMP

Dedico este trabalho a Jeová Deus por nortear a minha vida e às minhas avós Ana Ivelin e Maria Enedina (in Memoriam), cujo amor e presença continuam a inspirar minha jornada.

Saudades eternas.

"Sem dados você é apenas mais uma pessoa com uma opinião".

W. Edwards Deming — Data Scientist.

Agradecimentos

Ao meu orientador pela simplicidade, confiança e pela oportunidade de realizar este sonho, minha eterna gratidão;

Agradeço aos professores do departamento de Geociências pelos ensinamentos nas disciplinas estudadas durante o processo;

Não poderia deixar de agradecer aos meus colegas de pós-graduação, principalmente da sala 03, pelas valiosas trocas de ideias durante todo o processo nas disciplinas, pelas conversas mirabolantes e dos muitos cafés da tarde com risadas .. Obrigada pelo apoio nesta caminhada;

Aos professores Marcus W. Beims (UFPR), Mircea Galiceanu (UFAM) e Alderlene Brito (UFAM) pela ajuda, questionamentos e participação na escrita dos artigos;

Agradeço à minha família pelo suporte, amor e incentivo, em especial ao meu esposo Carlos Fábio e filhos Ana e Felippe pela compreensão. Vocês são minha base e inspiração;

À UFAM pela oportunidade de realização do Mestrado em Geociências pela infraestrutura e o acesso aos recursos necessários para a realização da pesquisa;

À FAPEAM por ter me concedido bolsa de estudos sem a qual minha formação não seria possível;

Ao IRD (Institut de recherche pour le développement) pelo financiamento à mobilidade na França, com o estágio de 1 mês no GET (Géosciences Environnement Toulouse) via projeto FEFFACION;

E a todos os que contribuíram, mesmo que indiretamente, para minha formação e para realização dessa dissertação.

Um grande obrigada a todos!

Resumo

Apliquei redes neurais convolucionais com arquiteturas baseadas em sub-redes para prever cheias e secas extremas na bacia do Rio Negro, localizada na região norte da América do Sul. A bacia é caracterizada por formações geológicas diversas, incluindo uma variedade de tipos de rochas que influenciam a infiltração e os padrões de escoamento, além de fatores climáticos que moldam a dinâmica das precipitações. Adicionalmente, a hidrodinâmica da bacia, impulsionada pelas interações entre suas formações geológicas e o fluxo de água, bem como sua topografia, que governa a direção e a velocidade do escoamento superficial, desempenham papéis cruciais em seu comportamento hidrológico. O regime de chuvas e a sazonalidade determinam ainda a distribuição temporal dos níveis de água. Os tipos de cobertura do solo também desempenham um papel significativo ao alterar as taxas de infiltração e o escoamento superficial. Analisando dados hidrológicos de cinco estações, a saber, Cucuí, Serrinha, Caracaraí, Santa Maria do Boiaçú e Moura, esta pesquisa busca modelar e prever eventos hidrológicos extremos. A aplicação da abordagem de aprendizado profundo destaca a importância de integrar dados de múltiplas estações para capturar variações localizadas e diferenças regionais. Esses achados ressaltam o potencial da inteligência artificial como uma ferramenta complementar aos modelos existentes usados pelas autoridades competentes para monitoramento, auxiliando na mitigação dos impactos de inundações e secas extremas. Embora não seja uma solução autônoma, o modelo fornece compreensão valiosa e enfatiza a necessidade de melhorias adicionais, especialmente por meio de ajustes de hiperparâmetros, para aumentar sua confiabilidade e precisão.

Palavras-chave: Previsão hidrológica, rede neural convolucional, influências geológicas e climáticas.

Abstract

I applied convolutional neural networks with subnet-based architectures to predict extreme floods and droughts in the Negro River basin, located in the northern region of South America. The basin is characterized by diverse geological formations, including a variety of rock types that influence water infiltration and flow patterns, alongside climatic factors that shape precipitation dynamics. Additionally, the basin's hydrodynamics, driven by interactions between its geological formations and water flow, as well as its topography, which govern the direction and speed of runoff, play crucial roles in its hydrological behavior. Rainfall regime and seasonality further dictate the temporal distribution of water levels. Land cover types also play a significant role in altering infiltration rates and surface runoff. By analyzing hydrological data from five stations, namely Cucuí, Serrinha, Caracaraí, Santa Maria do Boiaçú, and Moura, this research aims to model and forecast extreme hydrological events. The application of deep learning approach underscores the importance of integrating multi-station data to capture localized variations and regional differences. These findings highlight the potential of artificial intelligence as a complementary tool to existing models used by competent authorities responsible for monitoring, aiding in mitigating the impacts of extreme floods and droughts. While not a standalone solution, the model provides valuable insights and emphasizes the need for further improvements, particularly through hyperparameter tuning, to enhance its reliability and accuracy.

Keywords: Hydrological forecasting, convolutional neural network, geological and climatic influences.

Lista de ilustrações

Lista de tabelas

Lista de Siglas

- IA - Inteligência Artificial
- ANN - Redes Neurais Artificiais (inglês - Artificial Neural Network)
- CNN - Rede Neural Convolucional (inglês - Convolutional Neural Network)
- ANA - Agência Nacional de Águas e Saneamento Básico
- HYBAM - Serviço de Observação Hidrológica da Bacia Amazônica
- SGB - Serviço Geológico do Brasil
- Adam - Adaptative Moment Estimation
- MSE - Erro Quadrático Médio (inglês - Mean Squared Error)
- MAE - Erro Absoluto Médio (inglês - Mean Absolute Error)
- DC - Correlação de Distância (inglês - Distance Correlation)
- ENOS - El-Ninõ Oscilação Sul

Sumário

1 Introdução

1.1 Bacia do Rio Negro

A Bacia do Rio Negro, um sistema hidrológico extenso e intrincado dentro da vasta bacia amazônica, é um componente complexo e vital tanto para a região amazônica quanto para o ecossistema global. O Rio Negro é o afluente que carrega a maior vazão para o Rio Amazonas (FRAPPART et al., 2005). Localizado no noroeste da Amazônia, o Rio Negro é o sexto maior rio do mundo em termos de descarga de água, sendo caracterizado por um sistema de canais múltiplos e uma estrutura geológica distinta (MARINHO et al., 2022). O Rio Negro, servindo como o principal curso de água da bacia, possui particular relevância, especialmente ao se unir ao Rio Solimões para formar o vasto Rio Amazonas (LARAQUE; GUYOT; FILIZOLA, 2009).

A dinâmica fluvial da Bacia do Rio Negro é excepcionalmente complexa, o que motiva esforços extensivos de pesquisa para desvendar suas peculiaridades e ampliar o entendimento sobre esta bacia (MARINHO; FILIZOLA; CREMON, 2020; MARINHO; ZANIN; FILIZOLA, 2022). A bacia possui uma importância socioeconômica substancial na região (REIS et al., 2022). Trabalhos de modelagem na Bacia do Rio Negro têm sido fundamentais para compreender suas dinâmicas complexas, com foco na prevenção de desastres decorrentes de cheias e secas extremas, que se tornaram mais frequentes nos últimos anos (MACIEL et al., 2022; CHEVUTURI et al., 2022; CHEVUTURI et al., 2023).

A Bacia do Rio Negro é uma sub-bacia da Bacia Amazônica, reconhecida como uma das maiores bacias tropicais do mundo. A Bacia Amazônica cobre aproximadamente 6,1 milhões de km^2, representando cerca de 5% da área continental do planeta. Este vasto sistema hidrológico descarrega cerca de $6,6 \times 10^{12}\ m^3$ de água doce no oceano anualmente (MOLINIER et al., 1996). Sua diversidade climática e topográfica é notável. A porção ocidental é dominada pelos Andes que contribuem com a maior parte dos sedimentos transportados pelos grandes rios para as terras baixas (FILIZOLA; GUYOT, 2009). Cercada por terrenos geologicamente antigos, como os Escudos Brasileiro e das Guianas, o relevo da bacia varia de áreas próximas ao nível do mar até altitudes que excedem 6.000 m. A precipitação também varia significativamente, com taxas anuais entre 100 mm a 5.000 mm. Cerca de 70% da bacia é coberta por floresta tropical, consolidando a Bacia Amazônica como um dos ecossistemas mais biodiversos e globalmente significativos (FILIZOLA; GUYOT, 2011).

A capacidade de prever o futuro com base em dados passados e presentes é de fundamental relevância em áreas distintas, como mudanças climáticas, eventos extremos e raros, que incluem ondas gigantes no oceano, fenômenos climáticos extremos, como enchentes e secas, entre outros. As dinâmicas em problemas realistas são tipicamente sistemas complexos de alta dimensão não

triviais, e a geomorfologia da Bacia do Rio Negro é um exemplo (LATRUBESSE; FRANZINELLI, 2005; MARINHO; ZANIN; FILIZOLA, 2022).

Nos últimos anos, eventos extremos como cheias e secas tornaram-se cada vez mais frequentes, especialmente em bacias hidrográficas, destacando a necessidade de compreender esses eventos (GARCIA; LIBONATI; NUNES, 2018; GRANATO-SOUZA; STAHLE, 2023). Estudar esses eventos extremos apresenta desafios significativos, especialmente em bacias caracterizadas por alta complexidade. Uma exploração mais aprofundada das dinâmicas dessas bacias é essencial para compreender e mitigar os impactos de eventos extremos. A presente pesquisa visa prever com precisão cheias e secas extremas por meio de aprendizagem profunda.

1.2 Inteligência artificial na previsão de dados hidrológicos

A Inteligência Artificial (IA) tornou-se um tema de significativa relevância científica. O aprendizado de máquina, uma subárea da IA, emprega algoritmos que aprendem e se adaptam à medida que são apresentados a mais dados ao longo do tempo. O aprendizado profundo (deep learning em inglês) é uma subárea que utiliza Redes Neurais Artificiais (ANNs, do inglês Artificial Neural Networks) com múltiplas camadas sobrepostas.

No campo da hidrologia, o deep learning tem encontrado diversas aplicações, contribuindo para métodos no setor hídrico, abrangendo monitoramento, gestão, governança e comunicação de recursos hídricos (SIT et al., 2020). Diversos estudos utilizam ANNs para previsão de eventos extremos em várias bacias, como a previsão de fluxos na Bacia do Rio Apure, na Venezuela (DIBIKE; SOLOMATINE, 2001), previsão de vazão no Rio Yangtzé, na China (HA; LIU; MU, 2021), modelagem de previsão de níveis de água na Bacia do Rio Carrión, na Espanha (LINEROS et al., 2021), previsão de cheias na Bacia do Rio Perkerra, no Quênia (CHEBII; MUKOLWE; ONG'OR, 2022), previsão de fluxos hidrológicos na Bacia do Rio Sot, na Índia (SHARMA; SINGH; SHARMA, 2022), e desenvolvimento de modelos de cheias na Bacia do Rio Var, na França (AHMAD et al., 2022). E também tem sido empregado para mapeamento espacial, avaliando a contaminação por arsênio em águas subterrâneas (CHOWDHURY; ALOUANI; HOSSAIN, 2010)

A previsão de séries temporais, com foco nas dinâmicas do nível de água na Bacia do Rio Negro, abrangendo fenômenos como cheias e secas, representa desafios de alta complexidade. Assim, com o objetivo principal de compreender tais fenômenos, neste trabalho empregamos as Redes Neurais Convolucionais (CNN, do inglês Convolutional Neural Network) estruturadas como sub-redes dentro de uma rede global. O objetivo é prever cheias e secas extremas na Bacia do Rio Negro, utilizando dados históricos observados em cinco estações de medições fluviométricas: Caracaraí, Cucuí, Santa Maria do Boiaçú, Serrinha e Moura. Fornecendo dados históricos, essas estações contribuem significativamente para o entendimento das dinâmicas hidrológicas da bacia.

O uso de sub-redes, com a integração de dados de todas as estações, é fundamental para permitir uma previsão mais detalhada e precisa dos padrões globais de dados. Elas permitem que o modelo capture variações temporais localizadas, diferenças regionais nas dinâmicas hidrológicas e a influência de características geológicas e climáticas. Em geral, a previsão de séries temporais tem sido usada para detectar mudanças de regimes de movimento em sistemas caóticos, como a mudança de estações climáticas utilizando o famoso modelo de convecção atmosférica de Lorenz (BRUGNAGO et al., 2020; BRUGNAGO; GALLAS; BEIMS, 2020b), sua generalização (BRUGNAGO; FELICIO; BEIMS, 2022), e para prever as reversões do campo magnético da Terra (BRUGNAGO; GALLAS; BEIMS, 2020a).

1.3 Questões de pesquisa

- Como a interação entre fatores geológicos, cobertura do solo e regime de precipitação obtida de múltiplas fontes influencia a medição dos padrões hidrológicos extremos na Bacia do Rio Negro?
- Qual é a eficácia das Redes Neurais Convolucionais na previsão de eventos hidrológicos extremos na Bacia do Rio Negro?

1.4 Objetivo

1.4.1 Objetivo Geral

O objetivo geral deste trabalho é realizar a previsão eficiente de eventos de cheias e secas extremas por meio do uso de Redes Neurais Convolucionais com uma arquitetura em sub-redes, a partir de dados históricos de cinco estações fluviométricas, visando contribuir para o entendimento dos processos hidrológicos e o aprimoramento de ferramentas de previsão de riscos associados a extremos climáticos.

1.4.2 Objetivos Específicos

Destacam-se os seguintes objetivos específicos:

- Prever eventos de cheias e secas extremas por meio do uso de Redes Neurais Convolucionais treinadas com dados históricos integrados de diferentes estações da bacia;
- Investigar o potencial das redes neurais no apoio a medidas de mitigação dos impactos de eventos hidrológicos extremos;
- Verificar o desempenho das redes neurais nas previsões dos diferentes comportamentos hidrológicos observados nas diferentes estações da Bacia do Rio Negro.

1.5 Informações adicionais e organização

Durante as pesquisas realizadas no mestrado, resultaram um artigo publicado em uma revista internacional indexada e dois artigos submetidos e em fase de avaliação. As modelagens numéricas que viabilizaram a obtenção dos resultados apresentados nesta dissertação foram realizadas utilizando a linguagem de programação Python. Para a construção dos códigos das Redes Neurais Convolucionais (CNNs), foram empregadas as bibliotecas TensorFlow e Keras. Além disso, bibliotecas amplamente utilizadas na manipulação e análise de dados, como Pandas, NumPy e SciPy, foram essenciais para o processamento e organização das informações. Os gráficos apresentados foram elaborados a partir do uso da biblioteca Matplotlib. Os mapas da bacia do Rio Negro foram construídos com o auxílio do software QGIS, versão 3.28, em conjunto com ferramentas de automação e processamento espacial desenvolvidas em Python, destacando-se a utilização de bibliotecas como Rasterio para análise e manipulação de dados raster. O diagrama esquemático apresentado foi construído utilizando a ferramenta online draw.io. Por fim, tanto os artigos quanto esta dissertação foram redigidos utilizando o sistema de preparação de documentos LaTeX, garantindo alta qualidade na formatação e apresentação dos textos e figuras.

Este trabalho está estruturado da seguinte forma:

- **Capítulo 2:** Apresenta a área analisada, incluindo uma descrição das características geológicas da Bacia do Rio Negro, cobertura do solo e regimes de precipitação, além de detalhes sobre as cinco estações utilizadas para a aquisição de dados históricos.
- **Capítulo 3:** Descreve o modelo das Redes Neurais Convolucionais utilizadas nesta pesquisa, bem como uma visão geral dos procedimentos estatísticos empregados para a avaliação de desempenho do modelo.
- **Capítulo 4:** É dedicado à apresentação e discussão dos resultados obtidos.
- **Capítulo 5:** Apresenta as conclusões deste trabalho, destacando as principais contribuições e perspectivas futuras.
- **Apêndices:**
  - Contém a primeira página do primeiro artigo submetido para publicação da pesquisa desta dissertação, na revista científica internacional indexada *Hydrological Processes*. Apenas a primeira página, pois todos os resultados deste artigo se encontram aqui nesta dissertação;
  - Contém o segundo artigo completo, que analisa o uso de Redes Neurais Artificiais (ANNs) na previsão de cheias e secas extremas na Bacia do Rio Negro, submetido para publicação na revista científica internacional indexada *Hydrological Sciences Journal*. Os resultados deste segundo artigo referem-se ao primeiro ano do mestrado, durante o qual foram desenvolvidos diferentes modelos de ANNs para comparar

a eficácia das previsões utilizando dados das estações de Cucuí e Moura, ambas estações localizadas no Rio Negro;

- Contém o terceiro artigo completo, que trata sobre o uso de Rede Neural Convolucional (CNN) na identificação e classificação de incêndios florestais na Amazônia, publicado e escolhido como destaque na revista internacional indexada *International Journal of Remote Sensing*. Os resultados deste terceiro artigo referem-se à disciplina de Geologia Ambiental, cursada no segundo semestre de 2023;
- Contém o quarto artigo completo que se encontra em fase de preparação para submissão em revista internacional indexada, da qual tenho colaboração.
- E ainda a descrição de trabalhos em andamentos e perspectivas futuras.

2 Área de estudo

2.1 Descrição e localização das estações fluviométricas

Os dados hidrológicos utilizados neste estudo foram obtidos da Rede Hidrométrica Brasileira, gerenciada pela Agência Nacional de Águas e Saneamento Básico (ANA) (ANA, 2024), e do Serviço de Observação Hidrológica da Bacia Amazônica (HYBAM) (HYBAM, 2024). Essas estações estão localizadas ao longo da bacia, abrangendo os Rios Negro e Branco. A Tabela 1 fornece informações detalhadas sobre cada estação, incluindo seus códigos, períodos de coleta de dados, áreas de drenagem e coordenadas geográficas. Os códigos das estações foram atribuídos pela ANA, e sua gestão é realizada em colaboração com o Serviço Geológico do Brasil (SGB) (SGB, 2024).

Tabela 1 – Algumas informações relevantes sobre as estações.

Estação	**Código**	**Período**	**Área de drenagem** (km^2)	**Rio**	**Latitude**	**Longitude**
Cucuí	14110000	1981-2022	74.300	Negro	02°04'24"N	66°51'09"W
Serrinha	14420000	1980-2022	293.000	Negro	00°28'54"S	64°49'38"W
Caracaraí	14710000	1980-2022	126.000	Branco	01°49'17"N	61°07'25"W
Santa Maria do Boiaçú	14790000	1980-2022	184.000	Branco	00°30'19"S	61°47'11"W
Moura	14840000	1980-2022	593.000	Negro (pós-confluência)	01°27'24"S	61°38'05"W

Essas estações fornecem dados hidrológicos — como níveis de água, vazões, cargas de sedimentos, qualidade da água, etc — que são essenciais para compreender a dinâmica da bacia. Elas desempenham um papel fundamental no monitoramento e na avaliação dos recursos hídricos da Bacia Amazônica. Esta pesquisa utiliza exclusivamente dados de nível de água. As localizações geográficas dessas estações podem ser vistas na Figura 1. E elas são muito importantes para compreender os resultados. Os dados foram selecionados a partir de 1980, mas para a estação Cucuí, os dados estão disponíveis no site da ANA apenas a partir de 1981. Como resultado, ela possui um ano a menos de dados em comparação com as outras estações.

As diferenças na altura dos níveis de água (H) normalizados da Bacia do Rio Negro medidos nas réguas das cinco estações fluviométricas podem ser vistas na Fig. 2. As dinâmicas superficiais dos níveis de água normalizados nas diferentes estações são ilustradas, com as cores representando as transições entre os períodos de cheia e seca. As tonalidades vermelhas indicam secas, enquanto o azul corresponde às cheias, de acordo com a paleta de cores. As maiores instabilidades, caracterizadas por grandes oscilações, são observadas nas estações Cucuí e Caracaraí (ver Figs. 2(a) e 2(c)). Essas dinâmicas são refletidas nas tonalidades vermelhas vibrantes entre janeiro-maio e setembro-dezembro, que indicam períodos intensos de seca com variações significativas nos níveis da água. Da mesma forma, o período entre junho e setembro, com tonalidades azuis, corresponde às cheias, mostrando flutuações consideráveis.

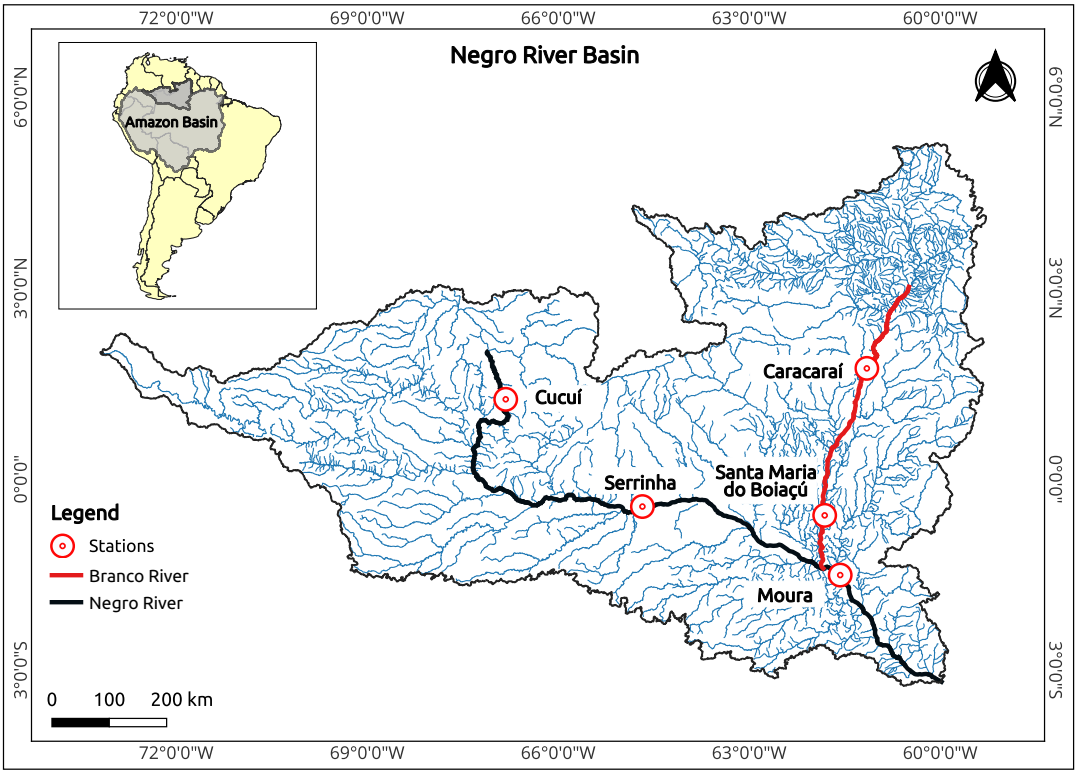


Figura 1 – Uma visão detalhada da Bacia do Rio Negro, ilustrando as localizações das cinco estações fluviométricas das quais os dados utilizados neste estudo foram coletados. As características destacadas dentro da bacia incluem os Rios Negro e Branco. Além disso, a figura mostra a posição da Bacia do Rio Negro como uma sub-bacia dentro da Bacia Amazônica na América do Sul.

Estações mais a jusante, como Serrinha e Santa Maria do Boiaçú, exibem um comportamento relativamente mais suave em comparação com as estações Cucuí e Caracaraí. Uma análise detalhada ao longo do período de 42 anos de dados, como mostrado nas Figuras 2(b) e 2(d), destaca a continuidade dos padrões fluviais das estações mais a montante para as mais a jusante. Isso indica que características significativas nas dinâmicas da água são transmitidas e preservadas ao longo da bacia. As irregularidades observadas nas estações Cucuí e Caracaraí são influenciadas pelas suas posições geográficas e condições climáticas, como as diferenças na frequência das chuvas entre as regiões do Alto Rio Negro e do Rio Branco. A interação das dinâmicas da água dos Rios Negro e Branco provavelmente desempenha um papel fundamental na formação do comportamento único observado em Moura (ver Fig. 2(e)) (MARINHO et al., 2022).

As características geológicas, os padrões de precipitação e a cobertura do solo dessas regiões contribuem ainda mais para essas variações, conforme discutido na subseção a seguir. É importante enfatizar, no entanto, que outros fatores não abordados neste estudo — como áreas desmatadas, fatores climáticos, regiões urbanizadas, terras utilizadas para atividades agrícolas e outras ações antrópicas — podem alterar significativamente os padrões hidrológicos da bacia.

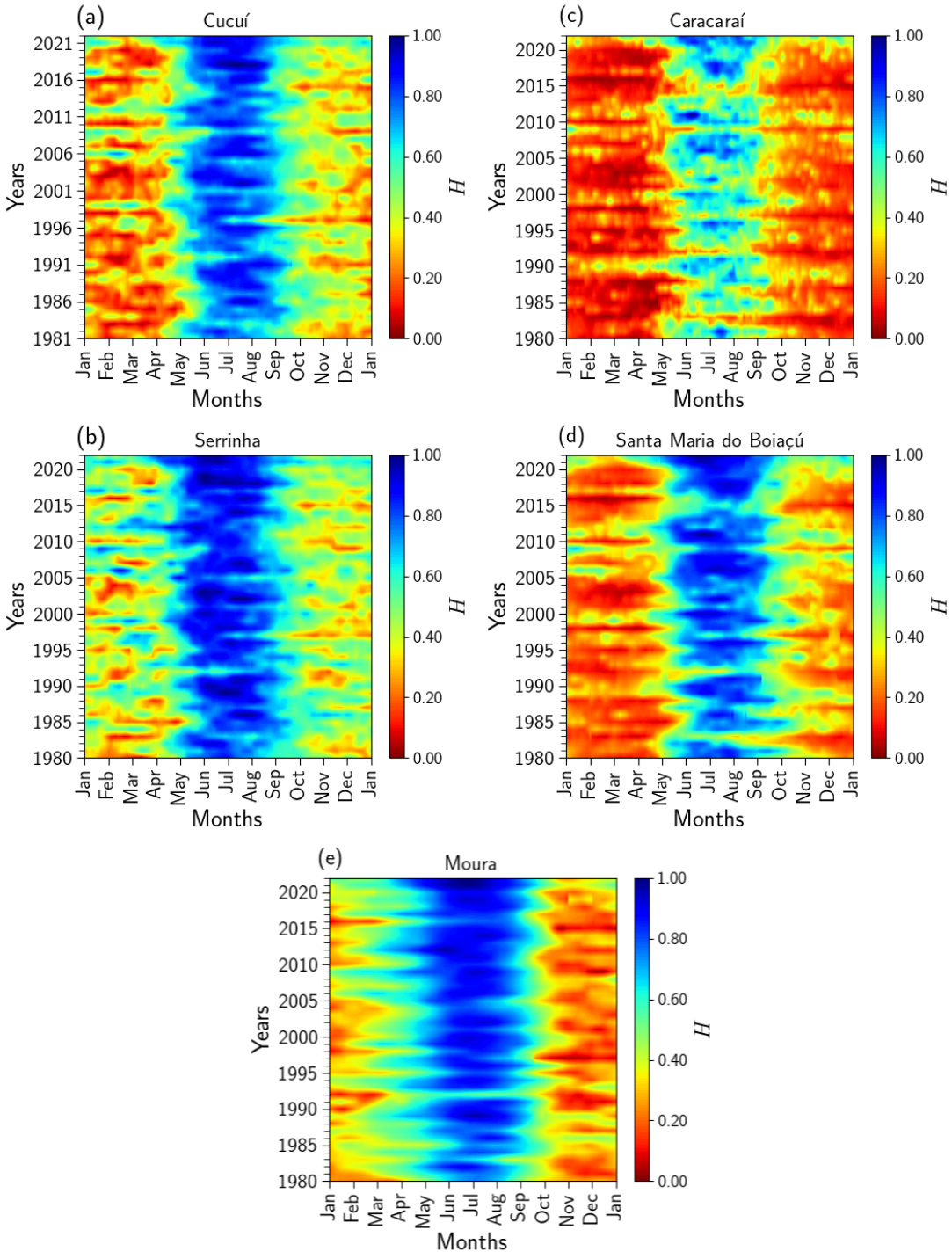


Figura 2 – Os mapas de calor ilustram as dinâmicas superficiais dos níveis de água na bacia do Rio Negro nas seguintes estações: (a) Cucuí, (b) Serrinha, (c) Caracaraí, (d) Santa Maria do Boiaçú e (e) Moura. Os gradientes de cores representam as variações dos níveis da altura da água (H) normalizados.

2.2 Contexto geral da Bacia do Rio Negro

Na Bacia Amazônica estão localizados quatro dos dez maiores rios do mundo em volume de vazão: Amazonas, Negro, Madeira e Japurá. Esses rios destacam-se por suas extensas e

complexas planícies de inundação e por apresentarem um padrão de drenagem do tipo multicanal (anabranching). Entre os principais afluentes do Rio Amazonas, os rios Madeira e Negro se destacam, abrangendo, respectivamente, 23% e 12% da área total da bacia. Quanto à contribuição para a vazão do Rio Amazonas, esses dois rios respondem por 15% e 14%, respectivamente (MARINHO, 2019).

A Bacia do Rio Negro ocupa 12% da Bacia Amazônica, abrangendo uma vasta área de drenagem de 712.000 km^2. Substancialmente, 82% de seu território está localizado dentro dos limites do Brasil, enquanto a porção restante se estende para outros países sul-americanos. Recebe o nome de Rio Negro a partir da confluência entre os rios Guainía e Casiquiare. No Brasil, esta bacia envolve os estados do Amazonas e Roraima. Com aproximadamente 900 km de extensão desde sua nascente até Manaus, o Rio Negro ocupa uma posição proeminente na região. Um de seus principais afluentes, o Rio Branco, contribui com uma área de drenagem significativa de 189.925 km^2. Durante os períodos de cheia, diversas planícies são inundadas, desempenhando um papel importante no ciclo hidrológico da bacia (GETIRANA, 2009). Embora as savanas ocorram em algumas áreas na Colômbia e Roraima, grande parte da Bacia do Rio Negro é coberta com floresta tropical densa. Estima-se que 15% da bacia são ocupadas por água durante os períodos de cheia (MARTINEZ; Le Toan, 2007). O ciclo hidrológico do Rio Negro é caracterizado pela chegada da estação de cheias que vai de maio a agosto, alcançando seu pico entre junho e julho (GUYOT et al., 1993).

O Rio Negro apresenta um padrão de drenagem do tipo anabranching, com arquipélagos como Anavilhanas e Mariuá, e uma rede fluvial que exibe consideráveis variações em termos de largura, sinuosidade e declividade (MARINHO, 2019). A partir da confluência com o Rio Uaupés, a largura média do Rio Negro é de cerca de 2 km, chegando a até 10 km em Manaus. Já o Rio Branco, a jusante de Caracaraí, possui uma largura média entre 1 e 2 km. Ao longo de grande parte de seu curso, os canais de ambos os rios são majoritariamente retilíneos, com uma sinuosidade média variando de 1 a 1,05 (MARINHO, 2019).

A dinâmica hidrológica do Rio Negro, incluindo suas variações sazonais e a interação com seus afluentes, é significativamente influenciada pelos efeitos hidráulicos. O efeito hidráulico ocorre quando um rio com maior descarga impede o fluxo de um afluente menor, fazendo com que a água se acumule na bacia do afluente e alterando a velocidade do fluxo e os níveis da água em ambos os rios (MEADE et al., 1991; SIQUEIRA; FILIZOLA, 2021; MARINHO; FILIZOLA; CREMON, 2020). Esse fenômeno é particularmente significativo na Amazônia, onde os rios apresentam variações sazonais pronunciadas na descarga. Durante os períodos de cheia, a água represada eleva os níveis dos rios, levando à inundação das áreas adjacentes; enquanto em períodos de seca, a descarga reduzida pode diminuir os níveis da água, afetando negativamente os ecossistemas aquáticos e as comunidades ribeirinhas. Além disso, esse efeito pode influenciar a dinâmica dos sedimentos, promovendo a sedimentação e o acúmulo em áreas represadas, o que pode degradar a qualidade da água ao reduzir a velocidade do fluxo.

2.3 Geologia, geomorfologia e solo

O Rio Negro dividide-se em seis trechos, considerando seu estilo geomorfológico e controles tectônicos. O primeiro abrange a parte superior da bacia, origem de grande parte dos sedimentos e do escoamento. O segundo apresenta canais anabranching em uma vasta planície holocênica. O terceiro segue um lineamento noroeste-sudeste até a confluência com o Rio Branco. O quarto, entre o Rio Branco e a montante de Anavilhanas, é estreito e possui afloramentos rochosos. O quinto, no arquipélago de Anavilhanas, cruza rochas cretáceas da formação Alter do Chão. O sexto trecho vai de Anavilhanas até Manaus, onde se encontra com o Rio Solimões (LATRUBESSE; FRANZINELLI, 2005).

A região do Alto Negro drena uma área compacta de planícies sedimentares na Colômbia, cobertas por savanas e rochas graníticas e granulíticas do Escudo das Guianas, sobrepostas por densa floresta tropical. Este setor da bacia, acompanhado por seus afluentes da margem direita, drena uma planície de baixo relevo, com elevações variando entre 60 e 160 m. Os relevos mais elevados são caracterizados por vários inselbergs isolados ou agrupados, atingindo até 700 m. É nas cabeceiras dos afluentes da margem esquerda que as elevações ultrapassam os 2.500 m (LATRUBESSE; FRANZINELLI, 2005). A elevação do terreno na Bacia do Rio Negro é ilustrada na Fig. 3. Os dados mostrados foram obtidos no OpenTopography (OpenTopography, 2024).

No setor norte da bacia, o terreno se eleva para altitudes variando de 500 a quase 3.000 m. Indo em direção às extremidades ocidentais da bacia, as altitudes oscilam entre 100 e 300 m. Da parte central à sul da bacia, a topografia vai gradualmente descendo, apresentando terrenos mais baixos com elevações que chegam até 100 m. Além disso, essas variações na elevação contribuem significativamente para modelar a dinâmica hidrológica da bacia, influenciando os padrões de fluxo e contribuindo para as interações complexas entre o rio e a paisagem ao seu redor.

No Hemisfério Norte, a bacia do Rio Negro abrange o Escudo das Guianas, enquanto ao sul da linha do Equador se estende a planície amazônica. Essa composição dual de hemisfério exerce uma influência profunda sobre o regime hidrológico da bacia, resultando em comportamentos hidrológicos distintos em comparação com outros afluentes dentro da Bacia do Rio Amazonas. As flutuações mais notáveis no volume de água ocorrem nas confluências entre o rio principal e seus afluentes. A Bacia do Rio Negro apresenta vastas expansões de planícies de inundação durante os períodos de cheia. Durante a estação de cheias, os níveis da água dos afluentes são ditados pelos cursos principais, resultando em inundações significativas nas regiões adjacentes às confluências (MARTINEZ; Le Toan, 2007; FRAPPART et al., 2005).

De maneira geral, os solos da Bacia do Rio Negro são de baixa fertilidade, originados em saprolitos profundos e com predominância de áreas arenosas, cobertas por uma densa vegetação de floresta tropical. Um solo típico da região é o Espodossolo (Podzol), encontrado principalmente nas áreas do alto Rio Negro e Rio Branco. Esses solos têm baixa capacidade nutritiva devido à

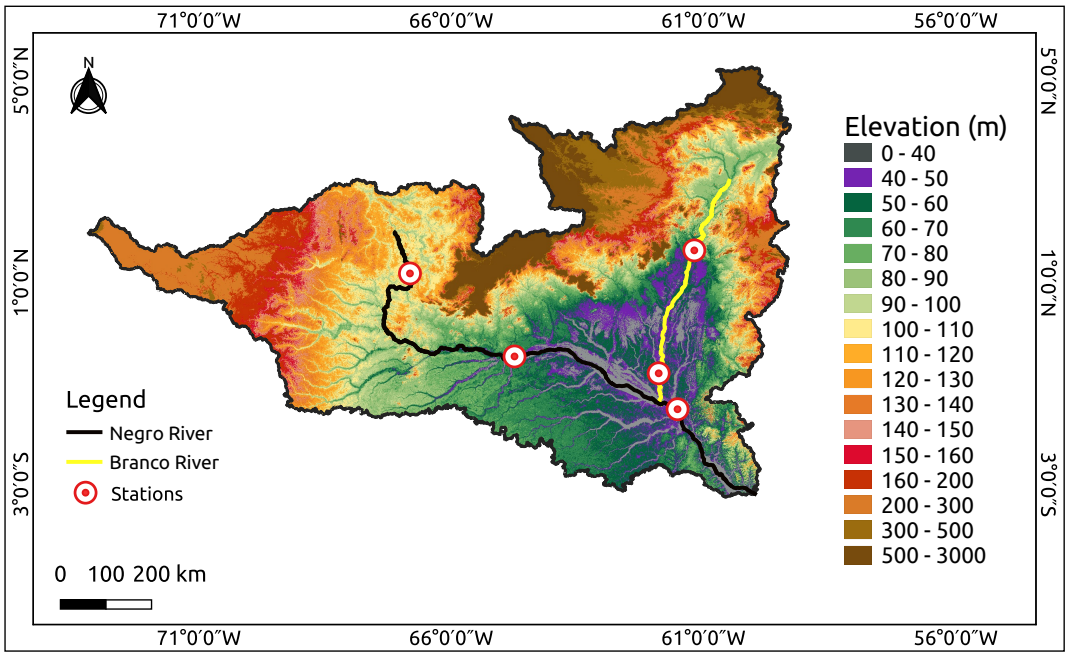


Figura 3 – Elevação do terreno na Bacia do Rio Negro. As cores na legenda representam os diferentes níveis de elevação.

alta acidez, apresentam grande quantidade de quartzo e quase nenhuma argila. Já os solos ao longo das planícies dos Rios Negro e Branco são classificados como Greissolos, com tonalidade cinza e permanecem saturados por longos períodos, conforme indicado por (QUESADA et al., 2011). Essas características do solo desempenham um papel fundamental na dinâmica ecológica e nos processos hidrológicos da região, influenciando a biodiversidade e os padrões de uso do solo. Detalhes a respeito das características dos solos podem ser vistos na Tabela 2 (MAIA, 2010; HOLANDA; MARMOS; MAIA, 2014; DUBROEUCQ; VOLKOFF, 1998).

Tabela 2 – Características dos Solos na Bacia do Rio Negro.

Região da Bacia	**Tipo de Solo**	**Características**	**Vegetação**	**Cor**	**Textura**
Escudo das Guianas e Planalto de Roraima	Espodossolo	Solos ácidos com forte lixiviação, horizontes claros, acúmulo de óxidos de ferro e alumínio.	Florestas de terra firme em áreas elevadas e úmidas, mas exige correção de acidez para uso agrícola.	Cinza-claro	Argilosa a arenosa
Planícies e Várzeas ao longo do Rio Negro	Latossolo	Solos profundos, bem drenados, com coloração vermelha/amarelada, devido a óxidos de ferro.	Adequado para a agricultura em áreas de terra firme, mas requer correção de acidez e adubação para melhores rendimentos.	Amarela a vermelha	Argilosa ou arenosa
Planícies e Várzeas ao longo do Rio Negro	Gleissolo	Solos encharcados, com características hidromórficas (alta saturação de água), rica matéria orgânica.	Vegetação de igapó e várzea, ideal para culturas adaptadas a solos inundáveis (ex.: arroz, frutais).	Escura	Fina, argilosa
Margens de rios e depósitos aluviais	Neossolo	Solos jovens formados por depósitos aluviais recentes, com textura variável.	Pode ser fértil dependendo dos sedimentos acumulados, com potencial para uso agrícola em áreas de planícies inundáveis.	Variável (geralmente clara)	Arenosa ou argilosa
Relevo Intermediário e Terrenos Menos Elevados	Plintossolo	Solos derivados de rochas ricas em ferro e silício, compactados, com acúmulo de óxidos de ferro.	Vegetação tropical e cerrado. Solo pouco fértil, precisa de correção para a agricultura.	Vermelha ou amarelada	Compacta a argilosa

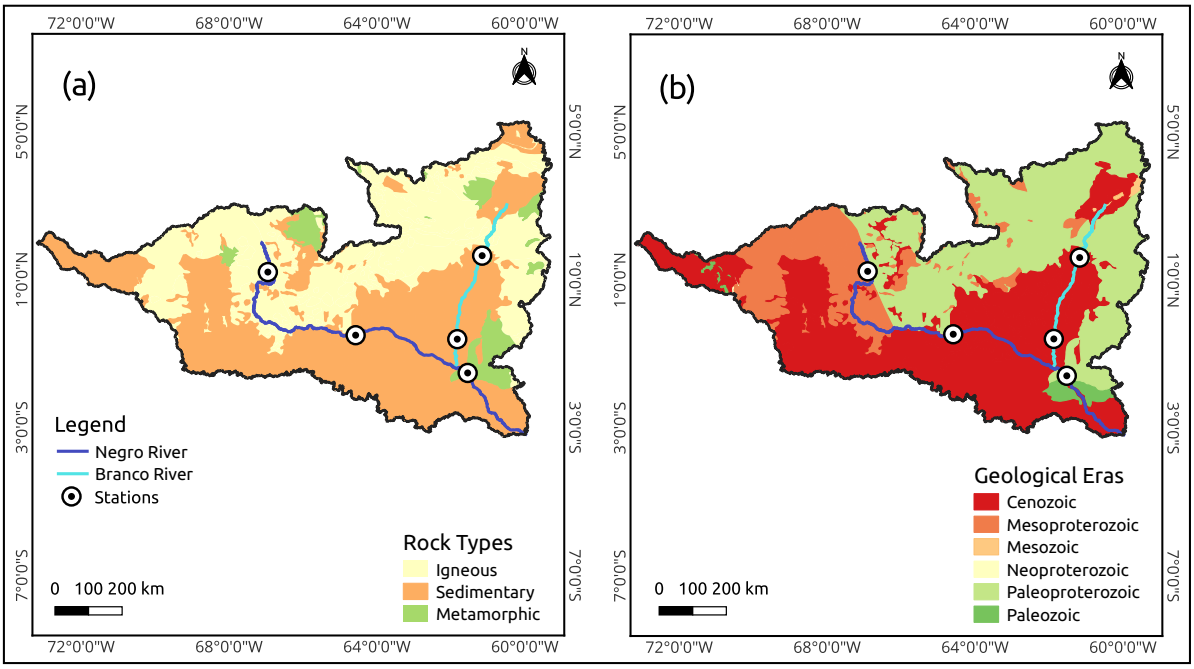


Figura 4 – (a) Tipos de rochas na Bacia do Rio Negro, destacando rochas ígneas, sedimentares e metamórficas, juntamente com os Rios Negro e Branco e as localizações das estações de monitoramento. (b) Eras geológicas representadas na bacia.

A Fig. 4(a) ilustra a distribuição dos tipos de rochas na Bacia do Rio Negro, destacando rochas ígneas, sedimentares e metamórficas. Os dados mostrados na figura foram obtidos do SGB (GeoSGB, 2024). As rochas sedimentares, cobrindo 225.815, 701 km^2 (39,39%) (veja na Tabela 3), dominam as regiões sul e parte da região oeste. Caracterizadas por camadas estratificadas e conteúdo fossilífero, essas rochas refletem ambientes deposicionais como deltas de rios e planícies de inundação. As rochas ígneas, como granitos e basaltos, conhecidas pela sua durabilidade e baixa porosidade, cobrem 304.737, 875 km^2 (53,15%) e predominam nas regiões norte, leste e oeste. As rochas metamórficas, encontradas em áreas menores (42.805, 099 km^2, 7,47%), principalmente no norte, leste, sul e extremo oeste, incluem formações de grau médio a alto formadas sob intensa pressão e temperatura, bem como rochas de grau baixo a médio indicativas de recristalização e transformação menos intensas de materiais preexistentes.

Tabela 3 – Distribuição da área total por tipo de rocha e eras geológicas na Bacia do Rio Negro.

Tipos de Rochas	**Área (km^2)**	**Porcentual (%)**	**Eras Geológicas**	**Área (km^2)**	**Porcentual (%)**
Ígnea	304.737,875	53,15	Paleoproterozoica	250.328,928	43,66
Sedimentar	225.815,701	39,39	Cenozóica	191.349,001	33,37
Metamórfica	42.805,099	7,47	Mesoproterozoica	116.769,151	20,37
-	-	-	Paleozoica	12.127,351	2,12
-	-	-	Mesozoica	2.323,279	0,41
-	-	-	Neoproterozoica	460,967	0,08

A Figura 4(b) apresenta o mapa geocronológico da Bacia do Rio Negro, destacando seis eras geológicas: Paleoproterozóica, Cenozóica, Mesoproterozóica, Paleozóica, Mesozóica e Neoproterozóica. A era Paleoproterozóica domina a bacia, cobrindo 250.328, 928 km^2 (43,66%)

(veja a Tabela 3), com rochas ígneas e metamórficas formando o embasamento cristalino. A era Cenozóica, ocupando 191.349, 001 km^2 (33,37%), apresenta sedimentos recentes típicos de planícies e áreas de sedimentação ativa. A era Mesoproterozóica corresponde a 116.769, 151 km^2 (20,37%), caracterizada por formações vulcânicas e intrusões graníticas. As eras restantes — Paleozóica, Mesozóica e Neoproterozóica — ocupam áreas menores, representando coletivamente menos de 3% da bacia. Essas formações refletem uma história de ambientes marinhos e glaciais, bem como processos tectônicos e sedimentares. A diversidade geológica dessas eras ressalta a complexidade tectônica e a história evolutiva da bacia (MASLOV, 2010; MAIA, 2010; HOLANDA; MARMOS; MAIA, 2014).

A diversidade geológica da Bacia do Rio Negro desempenha um papel fundamental na formação de suas características tectônicas e hidrológicas. A predominância de rochas ígneas nas áreas norte e leste, juntamente com as formações sedimentares no sul e oeste, reflete uma história tectônica complexa que influenciou as características hidrográficas locais. Essa diversidade litológica também impacta a dinâmica hidrológica, como pode ser observado nas estações Cucuí e Caracaraí (Figura 2). As rochas ígneas, com baixa porosidade, promovem drenagem rápida e amplificam as flutuações, enquanto as rochas sedimentares retêm água, moderando esses efeitos. Os eventos geotectônicos moldaram ainda mais a morfologia do terreno e a permeabilidade do solo, com as rochas resistentes do Paleoproterozóico limitando a infiltração, enquanto os sedimentos Cenozóicos e Mesoproterozóicos aumentam a retenção de água e a formação de áreas de planície de inundação (MENDES et al., 2021).

A Figura 5 mostra a divisão da área a montante de cada estação, quantificando as contribuições geológicas por tipos de rochas e eras geológicas em cinco regiões, com o objetivo de identificar as principais influências no comportamento hidrológico dessas estações. A Tabela 4 descreve as contribuições das eras geológicas e tipos de rochas para as áreas a montante das estações. Na região A (estação de Cucuí), as influências predominantes vêm da era Mesoproterozoica (49,19%) e das rochas ígneas (55,93%), seguidas pela era Cenozóica (31,36%) e rochas sedimentares (35,49%). A predominância de rochas ígneas, com sua menor capacidade de retenção de água, pode explicar a variação significativa nos níveis de água na estação Cucuí, resultando em flutuações mais abruptas. Na região C (estação de Caracaraí), a era Paleoproterozoica (79,01%) e as rochas ígneas (71,69%) também são predominantes, seguidas pela era Cenozóica (13,35%) e rochas sedimentares (21,68%). A grande presença de rochas ígneas contribui para as grandes flutuações nos níveis de água, levando a um comportamento hidrológico mais instável.

Na região B, a era Cenozóica domina com 44,86% da área, seguida pelas rochas sedimentares com 48,49%. Essa região assemelha-se com a forma da cabeça de um elefante. A era Mesoproterozóica (34,75%) e as rochas ígneas (47,51%) também contribuem de maneira significativa. A prevalência de rochas sedimentares sugere um comportamento hidrológico mais estável, pois essas rochas têm maior capacidade de retenção de água, levando a flutuações mais

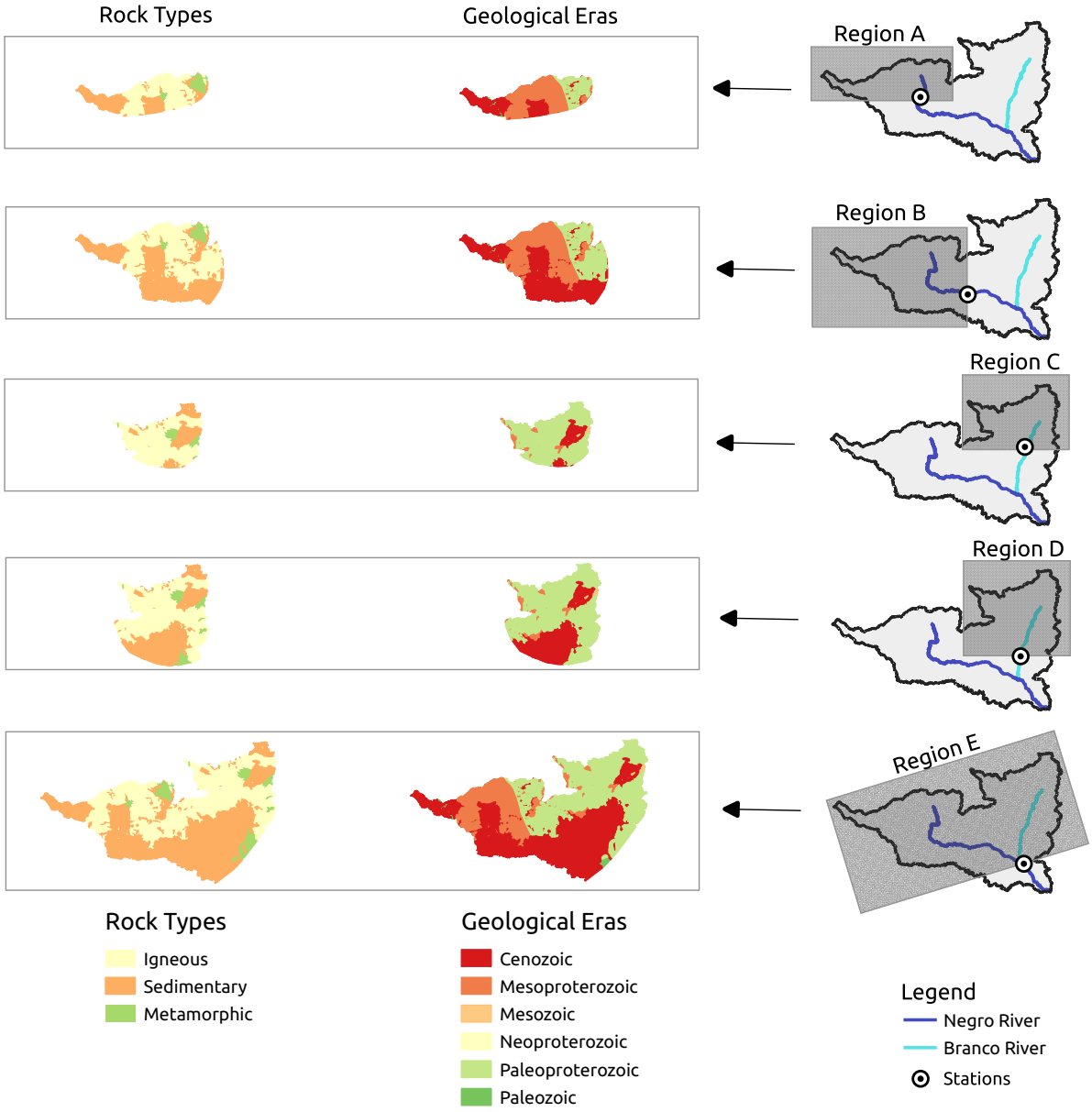


Figura 5 – Representação das regiões de contribuição das áreas a montante associadas aos tipos de rochas e eras geológicas de cada estação na bacia.

graduais. Na região D, a era Paleoproterozóica (62,52%) e as rochas ígneas (57,37%) são as mais dominantes, com a era Cenozóica contribuindo com 31,80% e as rochas sedimentares com 36,73%. Essa combinação de características geológicas provavelmente resulta em flutuações menos pronunciadas nos níveis de água devido ao equilíbrio entre as rochas ígneas e sedimentares. Por fim, na região E, a era Paleoproterozóica (44,44%) e as rochas ígneas (55,92%) também desempenham um papel significativo. No entanto, a presença de áreas de rochas sedimentares adiciona estabilidade aos níveis de água.

Tabela 4 – Contribuição das eras geológicas e tipos de rochas para a área total a montante de cada estação na bacia.

Região	Eras	Área (km^2)	Porcentual (%)	Rochas	Área (km^2)	Porcentual (%)
A	Mesoproterozóica	68.187,205	49,19	Ígnea	77.542,991	55,93
(montante	Paleoproterozóica	24.199,569	17,46	Sedimentar	49.205,265	35,49
de Cucuí)	Cenozóica	43.481,289	31,36	Metamórfica	11.882,815	8,57
	Paleozóica	2.307,148	1,66	-	-	-
	Neoproterozóica	455,860	0,33	-	-	-
B	Cenozóica	133.065,232	44,86	Sedimentar	143.838,569	48,49
(montante	Mesoproterozóica	103.072,672	34,75	Ígnea	140.925,623	47,51
de Serrinha)	Paleoproterozóica	57.740,988	19,46	Metamórfica	11.882,815	4,01
	Paleozóica	2.307,148	0,78	-	-	-
	Neoproterozóica	460,967	0,16	-	-	-
C	Paleoproterozóica	121.653,321	79,01	Ígnea	110.385,082	71,69
(montante	Cenozóica	20.560,878	13,35	Sedimentar	33.383,623	21,68
de Caracaraí)	Mesoproterozóica	9.443,551	6,13	Metamórfica	10.212,324	6,63
	Mesozóica	2.323,279	1,51	-	-	-
D	Paleoproterozóica	178.901,381	62,52	Ígnea	164.161,933	57,37
(montante	Cenozóica	90.979,252	31,80	Sedimentar	105.089,823	36,73
de Santa Maria	Mesoproterozóica	13.933,888	4,87	Metamórfica	16.886,044	5,90
do Boiaçú)	Mesozóica	2.323,279	0,81	-	-	-
E	Paleoproterozóica	238.574,811	44,44	Ígnea	300.224,246	55,92
(montante	Cenozóica	173.729,568	32,36	Sedimentar	201.087,092	37,46
de Moura)	Mesoproterozóica	116.769,151	21,75	Metamórfica	35.564,611	6,62
	Paleozóica	5.018,174	0,93	-	-	-
	Mesozóica	2.323,279	0,43	-	-	-

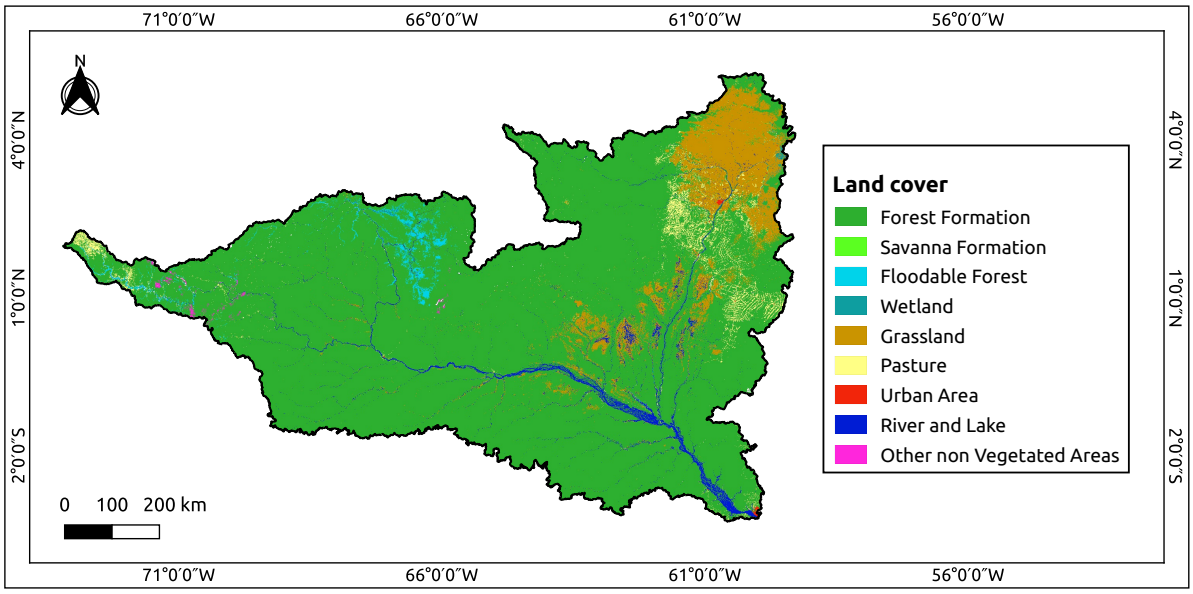


Figura 6 – Distribuição espacial dos tipos de cobertura do solo da Bacia do Rio Negro referente ao ano de 2020. Fonte: (MapBiomas, 2024).

2.4 Cobertura do solo

A distribuição espacial dos tipos de cobertura do solo é apresentada na Fig. 6 para o ano de 2020, com base nos dados do MapBiomas (MapBiomas, 2024). A classificação inclui categorias como formação florestal, formação de savanas, floresta inundável, áreas úmidas, campos, pastagem, área urbana, rios e lagos, e outras áreas não vegetadas, as quais são quantificadas na Tabela 5. Cada categoria é representada por cores distintas, conforme indicado na legenda da figura, oferecendo uma visão detalhada das diversas características de cobertura do solo

em toda a bacia. A bacia é predominantemente coberta por formações florestais, que ocupam uma vasta área de $612.904,623\ km^2$, representando 85,91% de seu território. As formações florestais são coberturas vegetativas densas compostas principalmente por árvores e outras espécies vegetais, desempenhando um papel fundamental na manutenção da biodiversidade regional e na regulação do ciclo hidrológico. Essas porcentagens estão acordo com as observadas na referência (MARTINEZ; Le Toan, 2007).

Tabela 5 – Área de cobertura do solo e distribuição percentual em toda a bacia.

Cobertura do Solo	**Área (km^2)**	**Porcentagem (%)**
Formação Florestal	612.904,623	85,91
Campo	60.228,432	8,44
Rio e Lago	15.579,103	2,18
Pastagem	14.159,443	1,98
Floresta Inundável	6.716,334	0,94
Área Úmida	2.032,318	0,28
Outras Áreas Não Vegetadas	1.090,699	0,15
Formação de Savana	364,652	0,05
Área Urbana	369,532	0,05

Na porção norte da bacia, especialmente no estado de Roraima, as áreas de campos e pastagens são mais prevalentes. Os campos, caracterizados por paisagens abertas dominadas por vegetação herbácea, cobrem $60.228,432\ km^2$ (8,44%), enquanto as pastagens, áreas utilizadas para o pastoreio de gado, ocupam $14.159,443\ km^2$ (1,98%). A terceira cobertura do solo mais extensa é representada por rios e lagos, que abrangem $15.579,103\ km^2$, equivalendo a 2,18% da bacia. Em proporções menores, outras coberturas incluem áreas úmidas, florestas inundáveis, formações de savanas, áreas urbanas e outras regiões não vegetadas, que juntas somam $10.573,535\ km^2$, ou 1,47% da bacia.

2.5 Regime de precipitação

O regime de precipitação média anual da Bacia do Rio Negro de 1981 a 2023, ilustrado na Fig. 7, apresenta notável variabilidade. A precipitação média anual, estimada com base nos dados do CHIRPS (CHIRPS, 2024), varia de níveis mais baixos a mais elevados, como representado pelo gradiente de cores na figura. A região oeste, próxima à estação Cucuí, recebe a maior precipitação, enquanto a região norte, ao redor de Caracaraí, apresenta níveis mais baixos. Estações centrais, como Serrinha, Santa Maria do Boiaçú e Moura, experimentam precipitações intermediárias.

Essas variações alinham-se com três vetores regionais identificados por (GUYOT et al., 1993): a primeira região, até Serrinha, recebe de 2000 a 3600 mm/ano, com 30% ocorrendo de maio a julho; a segunda, cobrindo a bacia do Rio Branco até Caracaraí, recebe de 1400 a 2300 mm/ano, com 57% de maio a julho; e a terceira, a jusante de Manaus, recebe 40% da precipitação de março a maio. A Fig. 7 ilustra esses contrastes por meio das linhas de contornos da intensidade de precipitação. Isso destaca regiões-chave, particularmente as áreas mais ao norte e a porção oeste da bacia, que exibem contrastes pronunciados nos padrões pluviométricos.

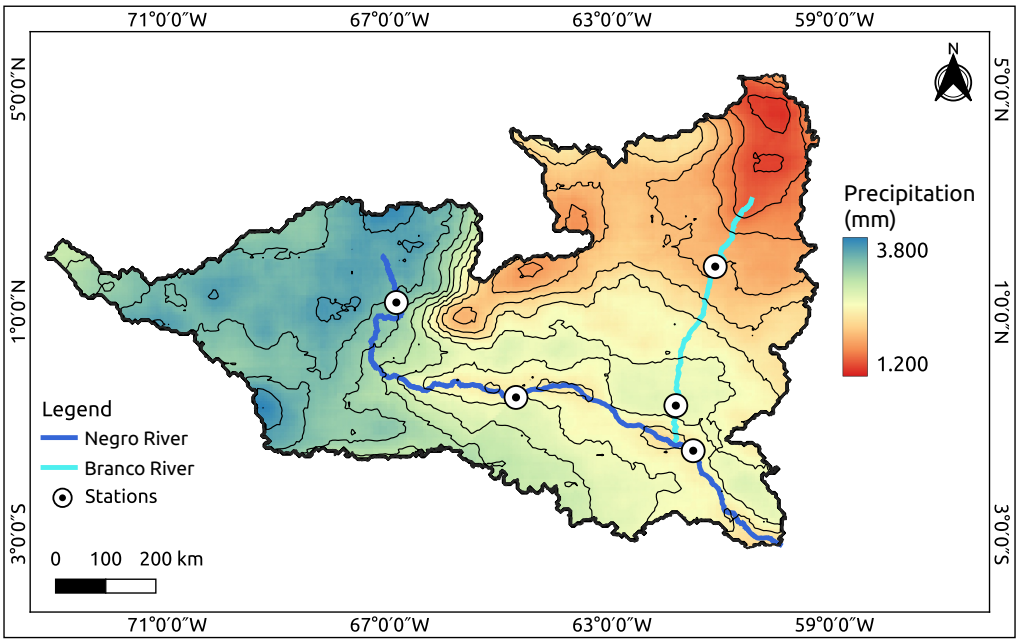


Figura 7 – (a) Precipitação média anual na Bacia do Rio Negro de 1981 a 2023. (b) Curvas de nível da precipitação para diferentes regiões da bacia. Fonte: (CHIRPS, 2024).

Tabela 6 – Tabela de precipitação média anual por região.

Região	**Precipitação Média** (*mm*)
A (montante de Cucuí)	3.409,503
B (montante de Serrinha)	3.214,063
C (montante de Caracaraí)	1.877,571
D (montante de Santa Maria do Boiaçú)	2.088,835
E (montante de Moura)	2.656,319

A Tabela 6 apresenta os valores de precipitação média anual para diferentes regiões da Bacia do Rio Negro conforme a Fig. 5, entre os anos de 1981 à 2023. A região A, correspondente a montante de Cucuí, apresenta a maior precipitação média com 3.409,503 mm, seguida pela região B (Serrinha), com 3.214,063 mm. A região C, associada a montante de Caracaraí, tem uma precipitação média de 1.877,571 mm. A região D, que abrange Santa Maria do Boiaçú, registra uma precipitação média de 2.088,835 mm, enquanto a região E, representando Moura, apresenta um valor de 2.656,319 mm.

Portanto, a Bacia do Rio Negro é uma das principais bacias hidrográficas da Bacia Amazônica, cobrindo uma vasta área, sendo o Rio Branco seu principal afluente. Sua dinâmica hidrológica é complexa, influenciada por diversos fatores, como variação de elevações do terreno, que vão de áreas de baixa altitude a regiões montanhosas, e uma geologia diversificada, incluindo

rochas ígneas, sedimentares e metamórficas. A cobertura do solo na região varia entre áreas de floresta densa e regiões com pastagens, impactando a absorção de água e o escoamento. A precipitação é um fator chave, com taxas de chuva elevadas variando pela bacia, afetando os regimes de cheia e o transporte de sedimentos.

3 Redes neurais artificiais

3.1 Informações elementares

A Inteligência Artificial (IA) é uma área da computação que busca desenvolver sistemas capazes de realizar tarefas que normalmente requerem inteligência humana, como reconhecimento de padrões, tomada de decisão e previsão de eventos. Dentro desse campo, o Machine Learning se destaca por utilizar algoritmos que aprendem padrões a partir de dados, sem necessidade de programação explícita para cada tarefa. O Deep Learning, por sua vez, é uma subárea do Machine Learning que utiliza redes neurais profundas para modelar relações complexas nos dados, alcançando grande sucesso em aplicações como reconhecimento de imagem e processamento de linguagem natural (GÉRON, 2022).

As Redes Neurais Artificiais (ANN do inglês Artificial Neural Network) simulam a funcionalidade do cérebro humano no contexto da inteligência artificial, permitindo que modelos computacionais identifiquem padrões e resolvam problemas complexos que muitas vezes são desafiadores para os humanos compreenderem. Apesar de serem biologicamente inspiradas, não há evidências conclusivas de que as redes neurais representem fielmente o funcionamento do cérebro. No entanto, elas podem ser concebidas como estruturas matemáticas que aprendem a partir dos dados.

As Redes Neurais Convolucionais (CNNs) são um tipo especializado ANNs projetadas para processamento de dados. Elas são compostas por várias camadas interconectadas que extraem automaticamente características relevantes do conjunto de dados. Os principais componentes de uma CNN incluem:

- **Camadas de Convolução**: realizam a operação de convolução sobre a entrada, aplicando filtros que detectam padrões locais, como bordas e texturas.
- **Camadas de Pooling**: reduzem a dimensionalidade dos mapas de características extraídos pelas convoluções, preservando informações essenciais e reduzindo a complexidade computacional. Os métodos mais comuns são o max pooling e o average pooling.
- **Camadas totalmente conectadas**: conectam todas as unidades da camada anterior com todas as unidades da camada seguinte, permitindo que a rede combine características extraídas e produza a classificação final.
- **Funções de ativação**: introduzem não-linearidade no modelo, sendo a ReLU (Rectified Linear Unit) a mais comumente utilizada devido à sua eficiência no treinamento.

Esses componentes trabalham em conjunto para permitir que as CNNs aprendam representações hierárquicas de características, tornando-as altamente eficazes em tarefas como segmentação de imagens, reconhecimento facial, detecção de objetos e previsão de eventos.

3.2 Arquitetura de sub-Redes em CNNs

Neste estudo, as ANNs foram implementadas utilizando uma arquitetura baseada em sub-redes, onde cada sub-rede foi projetada para extrair características específicas dos dados históricos de uma estação. No contexto das sub-redes, a arquitetura começa com uma camada de entrada única para cada estação, responsável por receber os dados históricos correspondentes. As camadas ocultas de cada sub-rede são responsáveis por extrair características específicas da estação, utilizando operações de convolução e pooling para capturar padrões temporais únicos. Essas camadas foram projetadas para aprender características localizadas, como flutuações sazonais e tendências interanuais. A camada final de cada sub-rede produz uma representação refinada de características, que é posteriormente integrada ao modelo global, permitindo aprendizado colaborativo entre todas as estações, preservando, ao mesmo tempo, as assinaturas temporais distintas de cada sub-rede.

Para previsão de séries temporais, a CNN foi implementada para análise. O modelo foi desenvolvido utilizando as bibliotecas TensorFlow e Keras para a construção e treinamento da ANN (CHOLLET, 2019; GÉRON, 2022), e as bibliotecas Pandas e NumPy para pré-processamento e manipulação dos dados. A visualização dos resultados de desempenho foi realizada utilizando Matplotlib, que forneceu representações gráficas. A escolha dessas bibliotecas, conhecidas por sua eficiência e versatilidade, facilitou o design, treinamento e avaliação de modelos complexos de ANNs.

Para validar as topologias das sub-redes, empregamos a técnica estatística de *validação cruzada de k-dobras* (KOHAVI, 1995; SILVA; SPATTI; FLAUZINO, 2020). Este método avalia o desempenho de cada topologia proposta ao ser aplicada a um conjunto de dados distinto daquele usado para ajuste de parâmetros. Na validação cruzada de k-dobras, a amostra é dividida em k partições, onde $(k-1)$ partições são usadas para treinar a rede, e o restante do conjunto de dados é utilizado para teste. Assim, o desempenho de cada topologia candidata é avaliado, e a mais eficiente é selecionada.

Neste estudo, utilizou-se um modelo de aprendizado supervisionado com dados rotulados para prever o nível de água em cinco estações de medição na Bacia do Rio Negro. Essa abordagem permite o treinamento do modelo com base em padrões históricos associados aos valores-alvo fornecidos. No entanto, o modelo também pode ser adaptado para aprendizado não supervisionado, ampliando sua aplicabilidade. Por exemplo, autoencoders podem ser integrados como uma etapa de pré-processamento, onde representações latentes dos dados são aprendidas sem a necessidade de rótulos (PEREIRA et al., 2020; HINTON; SALAKHUTDINOV, 2006). Essas representações

podem, então, ser utilizadas como entrada para a rede principal, possibilitando previsões mesmo na ausência de alvos explícitos. Essa estratégia pode melhorar a detecção de padrões subjacentes e a capacidade de generalização do modelo em cenários com dados incompletos ou não rotulados. A escolha do aprendizado supervisionado baseia-se na necessidade de prever eventos específicos, como cheias e secas, aproveitando padrões identificados em dados históricos confiáveis.

Embora modelos não supervisionados possam revelar padrões latentes, eles não são adequados para prever eventos diretamente associados a rótulos específicos, como é necessário neste caso. Para garantir a generalização, o modelo foi projetado com técnicas de validação e regularização, incorporando dados recentes para prevenir o overfitting. Na arquitetura de sub-redes empregada, utilizou-se dados dos dez anos anteriores aos eventos previstos de cheia e seca. Esse intervalo de tempo mostrou-se suficiente para capturar os padrões necessários durante o processo de ajuste de hiperparâmetros. No entanto, a rede pode ser treinada com um conjunto de dados maior, se disponível, o que pode melhorar ainda mais sua capacidade de generalização e robustez para prever eventos futuros. O uso de autoencoders, por exemplo, é vantajoso, pois eles aprendem a extrair características representativas dos dados ao comprimir informações relevantes em dimensões menores. Essa habilidade os torna particularmente valiosos para detectar padrões complexos e subjacentes, mesmo em conjuntos de dados não rotulados. A subseção a seguir concentra-se exclusivamente nas características arquitetônicas e nos detalhes das sub-redes.

3.3 Sub-redes para extração aprimorada de características

A Rede Neural Convolucional (CNN, do inglês Convolutional Neural Network) desempenham um papel fundamental no avanço dos estudos de previsão em séries temporais, especialmente ao abordar as complexidades inerentes aos fenômenos hidrológicos. Recentemente, as CNNs surgiram como uma ferramenta valiosa para a Amazônia, demonstrando seu potencial na classificação de incêndios florestais na Floresta Amazônica (ELEUTéRIO et al., 2024). Com sua capacidade única de identificar padrões temporais e correlações significativas, as CNNs emergem como uma solução robusta e eficaz para lidar com a variabilidade intrínseca de eventos hidrológicos, incluindo ocorrências extremas em bacias hidrográficas, como a do Rio Negro. Neste trabalho, cada sub-rede foi configurada para explorar os dados de uma estação hidrológica, utilizando uma CNN-1D projetada para aprender características específicas de suas séries temporais. Todas as sub-redes são CNN-1D, mas por simplicidade vamos chamar apenas de CNN. Os dados utilizados foram obtidos de cinco estações hidrológicas, distribuídas na Bacia do Rio Negro. Cada conjunto de dados consiste em séries temporais com colunas representando os níveis de água (H). A estrutura temporal dos dados reflete as dinâmicas sazonais do sistema hidrológico, aspecto essencial para modelar as variações nos níveis de água ao longo do tempo.

Para garantir a integridade e a escalabilidade dos dados, foi realizado um rigoroso processo de pré-processamento. Os dados foram normalizados utilizando o MinMaxScaler da

biblioteca Scikit-learn, que transformou todas as variáveis para um intervalo entre 0 e 1, com um mesmo recorte que corresponde à 10 anos de dados anteriores aos eventos para todas as estações. O MinMaxScaler é um método de normalização que transforma os valores de um conjunto de dados para um intervalo específico. Ele realiza essa transformação reescalando cada valor com base no mínimo e no máximo da distribuição, preservando a relação entre os dados sem distorcer sua estrutura original. Esse processo é essencial para reduzir as disparidades entre os diferentes conjuntos de dados e melhorar a eficiência do treinamento da rede neural. Além disso, os dados foram reorganizados em um formato tridimensional ($amostras \times passos\ temporais \times canais$) , atendendo aos requisitos das camadas convolucionais 1D na rede.

Os dados de cada estação foram tratados como entradas independentes no modelo, com 365 passos temporais que capturam um ano completo de observações diárias. Cada amostra é unidimensional, considerando apenas uma variável de entrada: o nível de água. Assim, a estrutura dos dados para a rede neural segue o formato $(5 \times 365 \times 1)$, onde 5 representa o número de estações, 365 o número de passos temporais, e 1 o número de variáveis em cada passo. Notavelmente, a dimensão ($canais$) pode ser expandida para incluir outras características medidas, como vazão, precipitação, temperatura, sedimentos ou dados de satélites. Contudo, neste estudo, apenas os dados de nível de água foram considerados.

Cada sub-rede segue uma estrutura padrão composta por camadas convolucionais 1D e camadas de pooling. A primeira camada convolucional aplica filtros para extrair características locais dos dados temporais. Em seguida, duas camadas adicionais de convolução e pooling refinam as representações espaciais e temporais dos dados, seguidas de camadas de dropout (com taxa de 0,5) para prevenir overfitting. Finalmente, a saída da sub-rede é achatada (Flatten layer), preparando os dados para integração no modelo global. A configuração das sub-redes variou de acordo com as características específicas de cada estação, como o número de filtros, tamanho do núcleo (kernel size) e taxa de dropout. Para as estações de Caracaraí e Cucuí, foi utilizada uma configuração com 64 filtros e tamanho de núcleo igual a 4, enquanto, para as estações de Moura, Santa Maria do Boiacú e Serrinha, a configuração foi ajustada para 32 filtros e núcleo igual a 4, equilibrando complexidade e capacidade de generalização. A arquitetura do modelo global com suas sub-redes específicas é ilustrada na Figura 8. Alguns termos na figura e aqui no texto são mantidos em inglês por questão de originalidade.

3.3.1 Integração e fusão de dados das estações

Uma característica distinta desta abordagem é a integração de dados de todas as estações em todas as camadas do modelo, permitindo um aprendizado global e robusto. Diferentemente de abordagens convencionais, onde cada estação teria seu próprio modelo independente, esta arquitetura combina as saídas das sub-redes específicas em uma única camada concatenada. Isso garante que os dados de todas as estações influenciem simultaneamente e de forma colaborativa o processo de aprendizado da rede. Esse processo não impede a avalição individual dos resultados

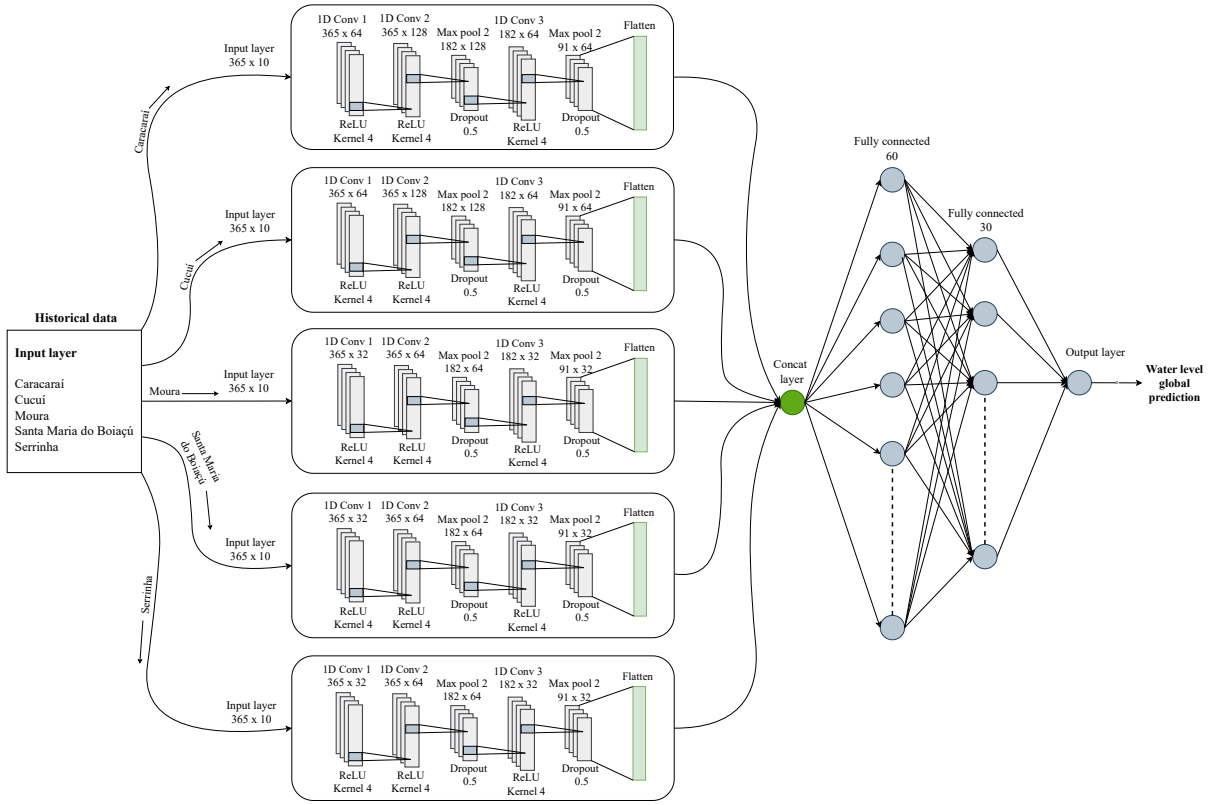


Figura 8 – Diagrama da arquitetura das CNNs ilustrando as sub-redes projetadas para extração de características dos dados de cada estação hidrológica da Bacia do Rio Negro.

pois cada rede é responsável pelo aprendizado das características de cada estação. Esse processo de fusão é crucial, pois permite que o modelo aprenda as interações complexas entre as diferentes características hidrológicas das diversas regiões dentro da bacia. Ao integrar dados de diferentes estações, a rede neural consegue capturar padrões temporais comuns, bem como variabilidades locais influenciadas pelas características geológicas da bacia. Isso melhora a previsão dos níveis de água em áreas da bacia que podem apresentar comportamentos climáticos e estruturas geológicas distintas. Por exemplo, enquanto algumas regiões podem ser mais afetadas por eventos frequentes de precipitação (ver Fig. 7), outras podem experimentar períodos prolongados de seca. O aprendizado conjunto desses padrões aprimora a precisão do modelo e a generalização global.

Embora o modelo aprenda características de todas as estações durante o treinamento, ele pode gerar previsões para uma estação específica durante a inferência. Isso é alcançado fornecendo apenas os dados de entrada correspondentes à estação de interesse, excluindo os dados das outras estações. Essa abordagem garante que a saída do modelo reflita as características temporais únicas da estação selecionada, aproveitando os padrões globais aprendidos para uma maior precisão. Após a fusão das características extraídas pelas sub-redes, o modelo prossegue com uma série de camadas densas totalmente conectadas. A primeira camada densa contém 60 neurônios, seguida por outra camada densa com 30 neurônios, ambas com regularização $L2$ ($\lambda = 0.05$). A regularização $L2$, adiciona a soma dos quadrados dos pesos ($\lambda \sum w^2$) à função de

perda, penalizando pesos altos. Isso reduz o overfitting e melhora a generalização do modelo. A camada de saída consiste em um único neurônio, que gera a previsão global final do nível de água para o evento alvo.

3.3.2 Treinamento e validação do modelo

Para o treinamento do modelo, os dados de todas as estações foram concatenados em um único conjunto de treinamento e validação, garantindo que o modelo fosse exposto a uma ampla variedade de padrões temporais. O conjunto de dados foi dividido em 70% para treinamento e 30% para validação. O modelo foi otimizado utilizando o algoritmo Adam (Adaptive Moment Estimation) com uma taxa de aprendizado de 0,001 e treinado com a função de perda Mean Squared Error (MSE) e a métrica de avaliação Mean Absolute Error (MAE). Essas métricas são melhor explicadas na seção seguinte. Durante o treinamento, o modelo recebeu entradas de todas as estações simultaneamente, permitindo que o processo de aprendizado fosse enriquecido pela diversidade dos dados. A validação foi realizada ao longo de 200 épocas de treinamento, monitorando o desempenho do modelo em termos de MSE e MAE para garantir que ele generalizasse bem e não apresentasse overfitting.

Este modelo não apenas aproveita os dados históricos de cada estação individualmente, mas também os integra de forma coesa para capturar a complexidade do comportamento hidrológico da bacia, fornecendo uma previsão mais precisa e robusta dos níveis de água para o evento alvo. A capacidade de gerar previsões específicas para cada estação a partir de um modelo treinado globalmente oferece uma vantagem significativa, pois combina a precisão local com uma compreensão mais ampla das interações regionais. Isso é essencial para uma gestão hidrológica eficaz e previsões em grandes bacias.

3.4 Avaliação de desempenho das CNNs

Os métodos descritos nesta seção são os modelos estatísticos usados para avaliar a eficiência das CNNs na previsão dos dados.

3.4.1 Correlação de Distância

A Correlação de Distância (DC, do inglês Distance Correlation) é uma medida estatística utilizada para verificar a relação não linear entre variáveis, avaliando a associação entre as distâncias de pares de pontos em cada variável. Ela captura relações não lineares e é robusta a valores discrepantes, tornando-se útil em diversas aplicações estatísticas, como análise de séries temporais, análise de agrupamentos e detecção de padrões em conjuntos de dados complexos. Essa medida pode ser conduzida em espaços multidimensionais, $X \in \mathbb{R}^p$ e $Y \in \mathbb{R}^q$, permitindo a avaliação de dependência entre variáveis em conjuntos de dados de alta dimensão

(SZÉKELY; RIZZO; BAKIROV, 2007; MENDES; BEIMS, 2018; MENDES; SILVA; BEIMS, 2019; MENDES et al., 2022).

A DC analisa a correlação entre os valores observados (dados históricos) e os valores previstos pelas CNNs. Para um conjunto de dados $(X, Y) = \{(x_k, y_k) : k = 1, ..., n\}$ com $n > 2$, x_k representa os dados observados e y_k os valores previstos. Para a amostra X, é definida a matriz

$$A_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \tag{3.1}$$

onde $a_{ij} = |x_i - x_j|_p$ é a distância Euclidiana entre os elementos, $\bar{a}_{i.} = \frac{1}{n}\sum_{j=1}^{n} a_{ij}$ é a média aritmética das linhas, $\bar{a}_{.j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}$ é a média aritmética das colunas e $\bar{a}_{..} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij}$ é a média geral. Similarmente, para a amostra Y, é obtida a matriz B_{ij}.

A covariância de distância para uma amostra conjunta (X, Y) é definida por

$$\sigma(X, Y) = \frac{1}{n}\left(\sum_{i,j=1}^{n} A_{ij}B_{ij}\right)^{1/2}, \tag{3.2}$$

enquanto as variâncias de distância para as amostras X e Y são, respectivamente,

$$\sigma(X) = \frac{1}{n}\left(\sum_{i,j=1}^{n} A_{ij}^2\right)^{1/2} \quad \text{e} \quad \sigma(Y) = \frac{1}{n}\left(\sum_{i,j=1}^{n} B_{ij}^2\right)^{1/2}. \tag{3.3}$$

Portanto, o coeficiente de DC é obtido pela equação

$$\mathrm{DC}(X, Y) = \frac{\sigma(X, Y)}{\sqrt{\sigma(X)}\sqrt{\sigma(Y)}}. \tag{3.4}$$

3.4.2 Erro Quadrático Médio e Erro Médio Absoluto

O Erro Quadrático Médio (MSE, do inglês Mean Squared Error) quantifica a eficácia das previsões ou estimativas do modelo em relação aos valores reais (CHICCO; WARRENS; JURMAN, 2021). O MSE calcula a média dos quadrados das diferenças entre os valores previstos y_k e os valores observados x_k para cada observação no conjunto de dados, conforme a equação

$$\mathrm{MSE} = \frac{1}{n}\sum_{k=1}^{n}(y_k - x_k)^2, \tag{3.5}$$

onde n é o número total de observações, y_k é o valor previsto e x_k é o valor observado. Essa medida atribui maior peso a erros de previsão mais elevados devido às diferenças ao quadrado, tornando-a sensível a valores discrepantes, já que os erros associados a esses pontos são ampliados. Na avaliação do modelo, o objetivo é minimizar o MSE, buscando os parâmetros que resultem nos menores erros médios quadráticos.

Outra medida utilizada para avaliar o desempenho é o Erro Médio Absoluto (MAE, do inglês Mean Absolute Error) (CHICCO; WARRENS; JURMAN, 2021). Essa métrica verifica a

média da diferença absoluta entre os valores previstos pelas CNNs e os valores observados. Para isso, subtrai-se o valor real do previsto, ignora-se o sinal do resultado e calcula-se a média dos erros de todas as amostras, conforme a equação

$$\text{MAE} = \frac{1}{n}\sum_{k=1}^{n}|x_k - y_k|, \tag{3.6}$$

onde n é o número de elementos das amostras, x_k são os dados observados, y_k são os valores previstos pelas CNNs e || indica os valores absolutos. Na Eq. (3.6), percebe-se que a medida não diferencia erros positivos de negativos. O desempenho do modelo melhora à medida que os valores tendem a zero.

3.4.3 Coeficiente de determinação

O coeficiente de determinação, ou R-quadrado (R^2), demonstra a proporção da variância nos valores previstos que pode ser explicada pelas variâncias dos valores observados (CHICCO; WARRENS; JURMAN, 2021). Essa medida pode ser obtida pela equação

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(x_k - y_k)^2}{\sum_{k=1}^{n}(x_k - \bar{x})^2}, \tag{3.7}$$

onde n é o número de elementos, x_k são os dados observados, $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ é a média aritmética, e y_k são os valores previstos. Essa medida varia entre $-\infty$ e 1. Valores próximos de 1 representam um bom ajuste. Por outro lado, quanto mais negativa a medida, pior o ajuste, ou seja, as variâncias nos valores previstos estão menos linearmente relacionadas aos valores observados.

4 Resultados e discussões

Neste capítulo, são apresentados os resultados do estudo, destacando o desempenho das CNNs com arquiteturas baseadas em sub-redes na previsão de cheias e secas extremas na Bacia do Rio Negro. Essas previsões foram realizadas utilizando dados históricos concatenados das estações na rede global. A Tabela 7 destaca os dez eventos mais extremos de cheias e secas registrados nessas estações nos últimos 40 anos (1980 - 2022). Para análise adicional, foi selecionada a cheia extrema de 2021, pois ela aparece consistentemente entre os dez principais eventos em todas as estações. Da mesma forma, a seca extrema de 2015 foi escolhida para previsão, já que também está entre os dez maiores eventos nos registros de todas as estações. Os valores para as cheias de 2021 e as secas de 2015 estão destacados em negrito na tabela.

Tabela 7 – Os dez maiores extremos de cheias e secas nas estações. Os valores estão expressos em metros (m).

	Cucuí		**Serrinha**		**Caracaraí**		**Santa Maria do Boiaçú**		**Moura**	
	Cheia	**Seca**	**Cheia**	**Seca**	**Cheia**	**Seca**	**Cheia**	**Seca**	**Cheia**	**Seca**
1º	22,76 (2018)	12,09 (1997)	10,86 (2019)	**3.34 (2015)**	11,13 (2011)	-0,12 (1984)	13,37 (2011)	**1.56 (2015)**	15,98 (2022)	2,35 (2009)
2º	22,28 (2002)	12,32 (1982)	10,84 (2002)	3,35 (2006)	9,72 (2006)	-0,10 (1997)	13,07 (2022)	1,61 (1997)	**15,93 (2021)**	**2,49 (2015)**
3º	22,18 (1986)	12,38 (2006)	10,77 (1989)	3,39 (1982)	9,62 (2018)	**-0,01 (2015)**	**12,80 (2021)**	2,01 (2002)	15,44 (1989)	2,68 (1997)
4º	**22,14 (2021)**	12,39 (2009)	10,76 (2018)	3,56 (1984)	9,59 (2022)	0,05 (1987)	12,73 (2006)	2,10 (1991)	15,44 (1989)	3,10 (2001)
5º	22,05 (2006)	**12,39 (2015)**	10,74 (1986)	3,56 (2003)	9,59 (2017)	0,06 (1982)	12,56 (2018)	2,17 (1982)	15,08 (2014)	3,25 (1991)
6º	22,00 (2016)	12,40 (1994)	10,73 (2003)	3,66 (2017)	**9,48 (2021)**	0,12 (1986)	12,54 (2002)	2,23 (2009)	15,07 (2019)	3,28 (1992)
7º	21,80 (2010)	12,40 (2002)	**10,68 (2021)**	3,72 (1997)	9,47 (1981)	0,18 (1994)	12,50 (2000)	2,28 (1994)	15,06 (2009)	3,35 (1989)
8º	21,80 (2003)	12,46 (2003)	10,63 (2022)	3,81 (1987)	9,18 (1996)	0,20 (2002)	12,38 (2020)	2,29 (1984)	14,97 (2002)	3,67 (2012)
9º	21,78 (1990)	12,49 (1991)	10,61 (2014)	3,83 (2002)	8,92 (2002)	0,28 (1991)	12,24 (1996)	2,42 (2001)	14,87 (2011)	3,73 (1995)
10º	21,76 (2007)	12,55 (1987)	10,57 (2007)	3,93 (2009)	8,80 (2007)	0,30 (1983)	12,16 (2017)	2,42 (2017)	14,86 (2000)	3,74 (1981)

O ano de 2015 foi caracterizado por um evento El Niño, inicialmente associado com padrões alterados of precipitação na Floresta Amazônica (JIMéNEZ-MUñOZ et al., 2016). Em contraste, 2021 foi marcado por um evento de La Niña, conhecido por seus impactos climáticos significativos, incluindo enchentes extremas na região amazônica (ESPINOZA et al., 2022). Os eventos de 2015 e 2021 foram selecionados como estudos de caso para previsão, embora o modelo possa ser aplicado a qualquer ano. A influência do ENOS (Oscilação Sul-El Niño), que abrange tanto eventos de El Niño quanto de La Niña, é particularmente relevante para esses anos, pois afeta significativamente as dinâmicas hidrológicas da Bacia do Rio Negro. O ENOS refere-se a um fenômeno climático caracterizado por oscilações entre temperaturas da superfície do mar mais altas que a média (El Niño) e condições mais frias que a média (La Niña) no Pacífico equatorial, o que, por sua vez, tem impactos globais nos padrões climáticos e hidrológicos (NOAA, 2024a; NOAA, 2024b).

Na Fig. 9, são destacados os valores máximos alcançados durante as cheias (curva azul), os valores mínimos observados durante as secas (curva laranja) e os valores médios (curva verde). Os dados, como mostrado na Tabela 1 são considerados até o ano de 2022, mas aqui consideramos valores até 2021. Os pontos quadrados pretos indicam a inundação de 2021 e a seca de 2015, ambas analisadas nas previsões da CNN. Pode-se notar que em todos os casos, tanto as máximas quanto médias e ainda as mínimas tem aumentado ao longo desses 40 anos. Todas

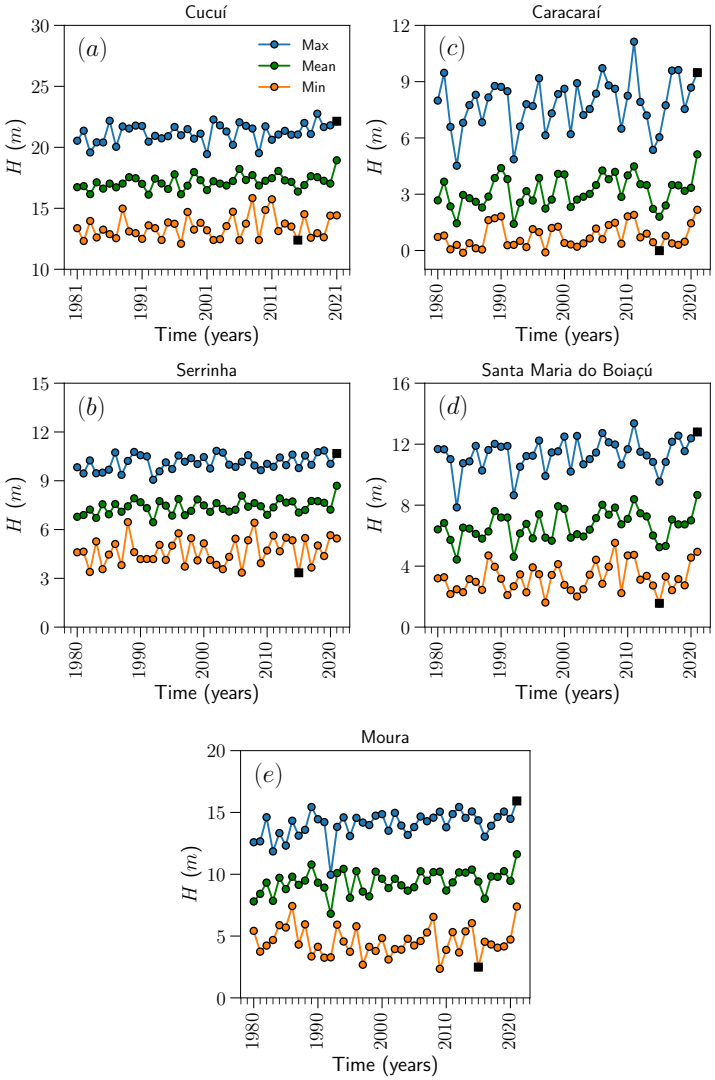


Figura 9 – Máximos anuais, média e mínimos dos níveis de água (H) nas seguintes estações: (a) Cucuí, (b) Serrinha, (c) Caracaraí, (d) Santa Maria do Boiaçú e (e) Moura. Os pontos quadrados pretos destacam os anos selecionados para as previsões de cheia (2021) e seca (2015) extremas.

as estações exibem uma variabilidade significativa nos valores de H. Notavelmente, 2015 foi marcado por uma seca severa, enquanto 2021 presenciou uma cheia significativa. É importante observar que as estações Caracaraí e Santa Maria do Boiaçú, ambas localizadas ao longo do Rio Branco, apresentam alta variabilidade, especialmente nos valores máximos. Esse padrão também é visível nas Fig. 2(c) e 2(d), que apresentam mapas de calor das dinâmicas de nível de água nas estações, demonstrando variações sazonais ao longo do ano. Os anos das secas mostrados na Tabela 7 e na Fig. 9 são considerados como o ano inicial da seca, mesmo que o valor mínimo observado ocorra no ano seguinte. Essa abordagem é adotada porque, em todas as estações da bacia do Rio Negro, o período da seca abrange consistentemente a transição entre dois anos.

4.1 Previsão de cheia extrema

O estudo das cheias é de grande importância devido ao impacto significativo que causam na população da Amazônia. As cheias nessa região geralmente são causadas por chuvas intensas e prolongadas, influenciadas por fenômenos climáticos como o ENOS. Esses eventos resultam em inundações extensas, particularmente em áreas como várzeas e terras baixas, que são mais vulneráveis ao aumento dos níveis dos rios. As cheias afetam moradias, infraestrutura de transporte e acesso a serviços essenciais, além de interromperem atividades agrícolas (MANTOVANI et al., 2024; VASCONCELOS et al., 2022). Prever tais eventos pode ajudar a mitigar seus efeitos, permitindo a implementação de sistemas de alerta precoce e estratégias de resposta mais eficazes.

Assim, a cheia de 2021 foi escolhida para análise utilizando o modelo de arquitetura de sub-redes com CNNs. Na Fig. 10, são apresentados os dados observados dessa cheia em cinco estações da bacia, juntamente com os dados previstos para cada estação, bem como medidas estatísticas robustas avaliando o desempenho do modelo. Notavelmente, as diferenças nos regimes hidrológicos entre as estações são evidentes, com cada uma apresentando variações distintas em diferentes períodos do ano. Essas variações estão associadas à geologia regional das estações, como tipos de rochas e elevação do terreno, padrões de chuva, cobertura do solo e outros fatores ambientais, conforme discutido em seções anteriores.

Na Fig. 10(a), a estação de Cucuí apresenta um padrão complexo de flutuações no nível da água ao longo do ano. Diferentemente de Caracaraí, Cucuí está localizada em uma região com alta precipitação anual (ver Fig. 7) e é predominantemente composta por rochas ígneas, conforme mostrado na Fig. 4(a). A vegetação da região é típica de ecossistemas florestais. Entre janeiro e março, podem ser observadas oscilações comumente chamadas de *repiquetes*, associadas ao período de seca do ano anterior. Um repiquete é uma reversão súbita e temporária no nível de água dos rios, caracterizada por flutuações que podem ocorrer tanto nas fases de enchente quanto de vazante (VALENZUELA et al., 2023). Os repiquetes referem-se a mudanças repentinas e imprevisíveis na direção das flutuações do nível do rio (COOMES et al., 2016). As flutuações nos níveis de água estão associadas a questões de saúde, como a malária, influenciando os padrões sazonais da doença e destacando seus efeitos na incidência (FONSECA et al., 2022).

O período de cheia começa em março, atingindo seu pico em junho com um valor máximo de $22,14\ m$, conforme detalhado na Tabela 7. De julho a novembro, ocorre uma recessão gradual do Rio Negro. No entanto, entre novembro e dezembro, dois repiquetes significativos emergem, complicando ainda mais o comportamento hidrológico da região. A previsão global da CNN (curva vermelha) está intimamente alinhada com os dados observados, capturando com precisão as tendências do repiquete inicial, o período de cheia, a fase de recessão e os dois repiquetes finais. Isso é comprovado pelos baixos valores de erro estatístico, com MSE = 0,04 e MAE = 0,16, e pelos altos indicadores de qualidade de ajuste, $R^2 = 0,99$ e DC = 0,99 (ver Fig. 10(b)). O gráfico de dispersão também segue uma tendência linear, validando ainda mais o desempenho do modelo. Todas as métricas de avaliação de desempenho mostradas na Fig. 10, se encontram na

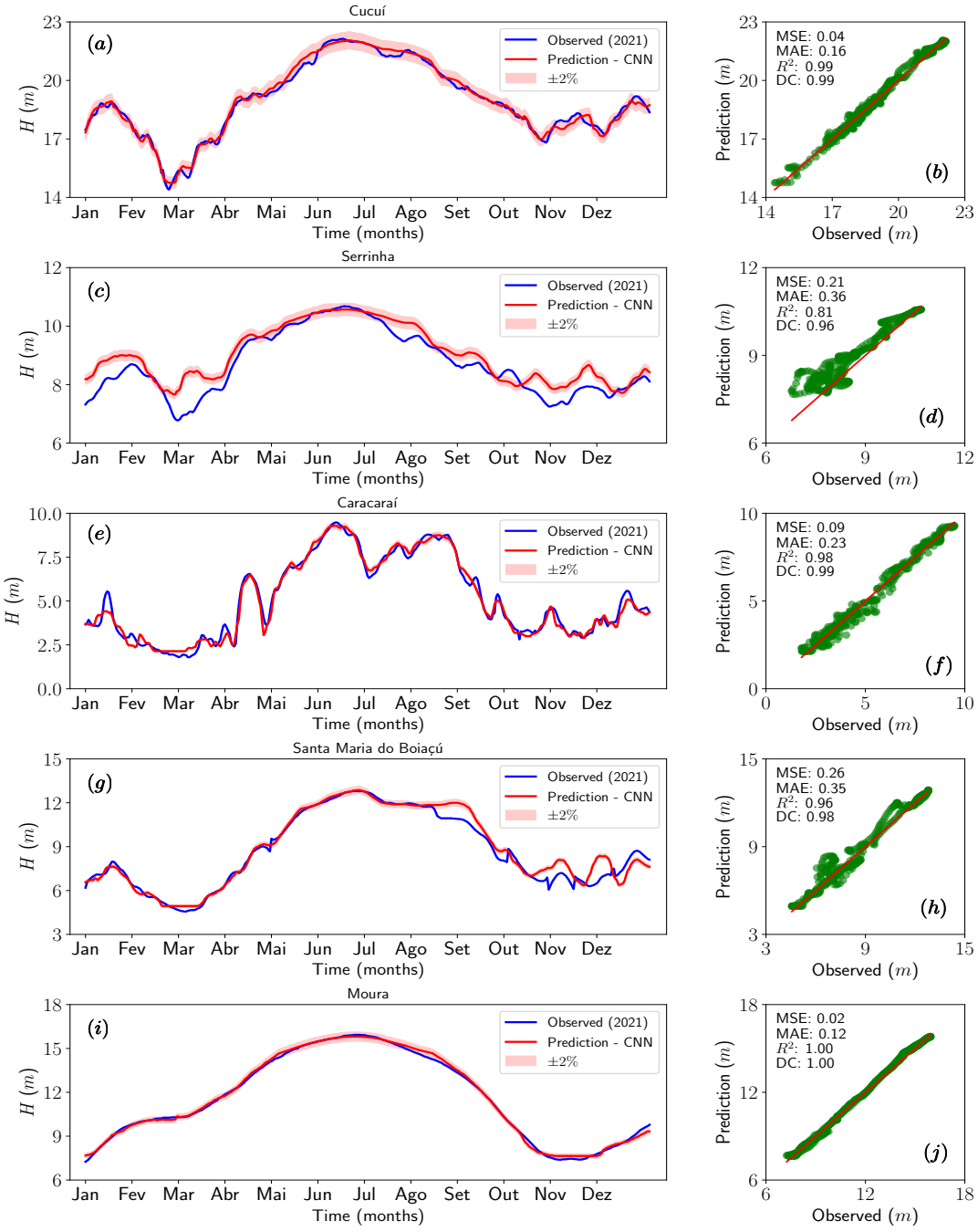


Figura 10 – Nível da água (H) ao longo dos meses para: (a) Cucuí, (c) Serrinha, (e) Caracaraí, (g) Santa Maria do Boiaçú e (i) Moura, durante o período de cheia de 2021, juntamente com seus respectivos gráficos de dispersão de desempenho em (b), (d), (f), (h) e (j). A curva azul representa os dados observados, enquanto a curva vermelha corresponde às previsões do modelo.

Tabela 8.

Na Fig. 10(e), a variação dos níveis de água é apresentada, revelando um comportamento hidrológico complexo na estação de Caracaraí. De janeiro a abril, o rio exibe condições características do período de seca do ano anterior. Um repiquete significativo é observado entre

Tabela 8 – Métricas de desempenho para diferentes estações.

Estações	MSE	MAE	R^2	DC
Cucuí	0,04	0,16	0,99	0,99
Serrinha	0,21	0,36	0,81	0,96
Caracaraí	0,09	0,23	0,98	0,99
Santa Maria do Boiaçú	0,26	0,35	0,96	0,98
Moura	0,02	0,12	1,00	1,00

abril e maio, marcado por uma elevação acentuada nos níveis de água seguida de uma queda igualmente pronunciada. Entre maio e junho, o rio inicia sua elevação sazonal, marcando o início da fase de cheia. No entanto, uma queda inesperada ocorre entre junho e julho, interrompendo temporariamente o aumento do nível de água. Durante esse período, observa-se o pico anual de cheia. Subsequentemente, de julho a agosto, os níveis de água voltam a subir, embora moderadamente. Em setembro, inicia-se a fase de recessão (seca), caracterizada por uma diminuição gradual nos níveis de água. De setembro a dezembro, as águas continuam a baixar, acompanhadas de pelo menos cinco repiquetes adicionais durante essa fase. Essa intrincada alternância de elevações e quedas nos níveis de água destaca o regime hidrológico único e dinâmico da estação.

No entanto, a previsão da CNN captura com precisão notável essas flutuações abruptas. Uma clara semelhança entre os valores observados e previstos é evidente, destacando a capacidade do modelo de aprender e replicar padrões sazonais complexos. Como uma abordagem de aprendizado supervisionado, a CNN demonstra habilidade em reconhecer e se adaptar a essas dinâmicas hidrológicas intrincadas. Na Fig. 10(f), o gráfico de dispersão ilustra um forte alinhamento entre os dados observados e previstos, seguindo uma tendência linear. As métricas estatísticas empregadas para avaliar o desempenho do modelo incluem MSE = 0,09, MAE = 0,23, $R^2 = 0,98$ e DC = 0,99. Esta estação, caracterizada por baixa precipitação anual, encontra-se em uma zona de transição entre formações rochosas ígneas e sedimentares, sendo predominantemente coberta por pastagens e campos. Portanto, tanto para Caracaraí quanto para Cucuí — duas estações localizadas em áreas distintas da bacia, cada uma influenciada por diferentes fatores geológicos e ambientais — a rede demonstra uma capacidade robusta de aprender e generalizar os comportamentos hidrológicos complexos característicos dessas regiões.

A Fig. 10(g) apresenta a estação Santa Maria do Boiaçú, localizada a jusante de Caracaraí no Rio Branco, enquanto a Fig. 10(c) exibe a estação Serrinha, situada a jusante de Cucuí no Rio Negro. Ambas as estações apresentam sinais hidrológicos semelhantes aos observados em suas contrapartes a montante. Em Santa Maria do Boiaçú, os níveis de água flutuam menos ao longo do ano em comparação com Caracaraí, embora repiquetes ainda sejam observados no início e no final do ano, correspondendo ao período de seca. O período de cheia segue um padrão semelhante ao de Caracaraí, mas com flutuações menos pronunciadas. Em Serrinha, o comportamento hidrológico reflete o de Cucuí, com repiquetes significativos ocorrendo no

início do ano (entre janeiro e março) e novamente entre novembro e dezembro. Geologicamente, Serrinha está localizada em uma área predominantemente composta por rochas sedimentares, com precipitação anual moderada e cobertura florestal densa. Em contraste, Santa Maria do Boiaçú, embora também situada sobre rochas sedimentares, recebe maior precipitação anual do que Serrinha.

A transição das rochas ígneas (estações a montante) para formações sedimentares pode contribuir para a atenuação das flutuações nos níveis de água. A previsão da CNN para Santa Maria do Boiaçú apresenta uma forte correspondência com os dados observados, especialmente durante a primeira metade do ano. Na segunda metade, à medida que os níveis de água começam a recuar, algumas oscilações são observadas, mas permanecem próximas aos valores reais. O pico máximo, observado entre junho e julho, com 12,50 m, é previsto com precisão dentro da margem de erro. Para Serrinha, a previsão apresenta ligeiramente mais variação em comparação com os dados observados; no entanto, captura de forma eficaz a tendência anual geral. O valor máximo registrado em Serrinha foi de $10,68\ m$. Conforme mostrado nas Figs. 10(d) e 10(h), as métricas para ambas as estações são bastante semelhantes, com exceção do R^2, que é ligeiramente menor para Serrinha.

A Fig. 10(i) apresenta os dados hidrológicos para a estação Moura, localizada imediatamente a jusante da confluência dos rios Branco e Negro. A confluência de rios refere-se ao ponto onde dois ou mais rios se encontram, marcado pelas linhas de fluxo convergentes e pela potencial mistura de fluxos. Isso geralmente resulta em interações complexas envolvendo o fluxo de água, o transporte de sedimentos e dinâmicas ecológicas (Dos Santos; STEVAUX; SZUPIANY, 2022; GUALTIERI et al., 2018). A confluência dos rios Branco e Negro tem sido amplamente estudada, particularmente no que se refere à morfologia do leito do rio e aos processos hidrológicos na região (MARINHO et al., 2022). A estação Moura serve como um ponto-chave onde os sinais hidrológicos de ambos os sistemas fluviais convergem, refletindo a interação entre as dinâmicas desses dois grandes rios. As influências combinadas desses rios, cada um com regimes de fluxo e variações sazonais distintos, contribuem para uma assinatura hidrológica única que é fundamental para entender o comportamento da bacia hidrográfica em um contexto mais amplo.

O evento de cheia de 2021 na estação Moura apresentou uma curva hidrológica notavelmente suave, uma característica que aprimora a capacidade das sub-redes CNN de capturar suas dinâmicas. O pico observado alcançou $15,93\ m$, tornando-se o segundo maior evento de cheia registrado na estação. As previsões do modelo alinham-se estreitamente com os dados observados, como mostrado pelas métricas de desempenho na Fig. 10(j), que reportam valores excepcionais como $R^2 = 1$ e $DC = 1$. Geologicamente, a estação Moura está situada na interface de transição entre rochas sedimentares e metamórficas, em uma região de baixa elevação topográfica, níveis moderados de precipitação e cobertura florestal. Essas características geológicas e climáticas, juntamente com sua localização a jusante da confluência dos rios Branco e Negro, influenciam significativamente o comportamento hidrológico da estação, contribuindo para a estabilidade e

previsibilidade de seus padrões de cheia.

4.2 Previsão de secas extremas

Esta seção examina o desempenho da CNN na previsão da seca de 2015, um ano significativamente impactado por um forte evento de El Niño e classificado entre as dez secas mais severas nas cinco estações, conforme apresentado na Tabela 7. Como mencionado anteriormente, os anos de El Niño estão associados a anomalias climáticas significativas na região Amazônica, incluindo períodos prolongados de seca, precipitação reduzida e temperaturas elevadas (JIMéNEZ-MUñOZ et al., 2016). As secas na Amazônia têm consequências profundas, incluindo diminuição da densidade de árvores, alteração da composição florística, aumento da mortalidade de árvores (PRESTES et al., 2024), maior risco de incêndios florestais (dos Reis et al., 2021), redução da navegabilidade dos rios e isolamento de comunidades (LIMA et al., 2024). Compreender e prever esses eventos é cada vez mais essencial, pois as secas na região se tornaram mais frequentes, provavelmente devido a uma combinação de variabilidade natural e mudanças climáticas e antropogênicas. Desenvolver modelos preditivos confiáveis, como as CNNs, é vital para mitigar os impactos desses eventos extremos, melhorar a gestão dos recursos hídricos e proteger a estabilidade ecológica e socioeconômica da Amazônia.

Na Fig. 11, a seca de 2015 é analisada em cinco estações. Para todas as estações, adota-se uma abordagem que considera o segundo semestre de um ano, durante o qual a seca começa, seguido pelo primeiro semestre do ano subsequente, quando a seca termina. Isso cria um ciclo contínuo de 365 dias cobrindo o período de 2015/2016. Esse ciclo ininterrupto reflete efetivamente a dinâmica da seca, com o segundo semestre de um ano se ligando ao primeiro semestre do próximo, formando uma sequência lógica e contínua. Embora o valor mínimo medido possa ocorrer no ano seguinte, ele está associado ao início da seca do ano anterior, garantindo que a análise capture com precisão a evolução do fenômeno ao longo do tempo. Essa configuração também é aplicada para o treinamento e validação da CNN. Dez anos de dados usados para o treinamento e validação seguem a mesma metodologia, com o segundo semestre de um ano seguido pelo primeiro semestre do próximo, formando ciclos contínuos de 365 dias. Essa abordagem permite que a CNN aprenda o padrão da seca ao longo do tempo, capturando efetivamente tanto a dinâmica temporal quanto os efeitos sazonais característicos desse fenômeno nas cinco estações.

A Fig. 11(a) ilustra a seca de 2015 na estação Cucuí. Embora apresente menos oscilações do que Caracaraí, o comportamento hidrológico ainda demonstra complexidade significativa. A curva prevista pela CNN se alinha de perto com a tendência observada ao longo de todo o ciclo da seca, capturando com precisão a queda acentuada de julho a outubro, a descida gradual que se estende por cerca de cinco meses entre outubro e março, e o subsequente aumento que leva ao período de cheia de 2016. Isso demonstra a capacidade do modelo de representar efetivamente

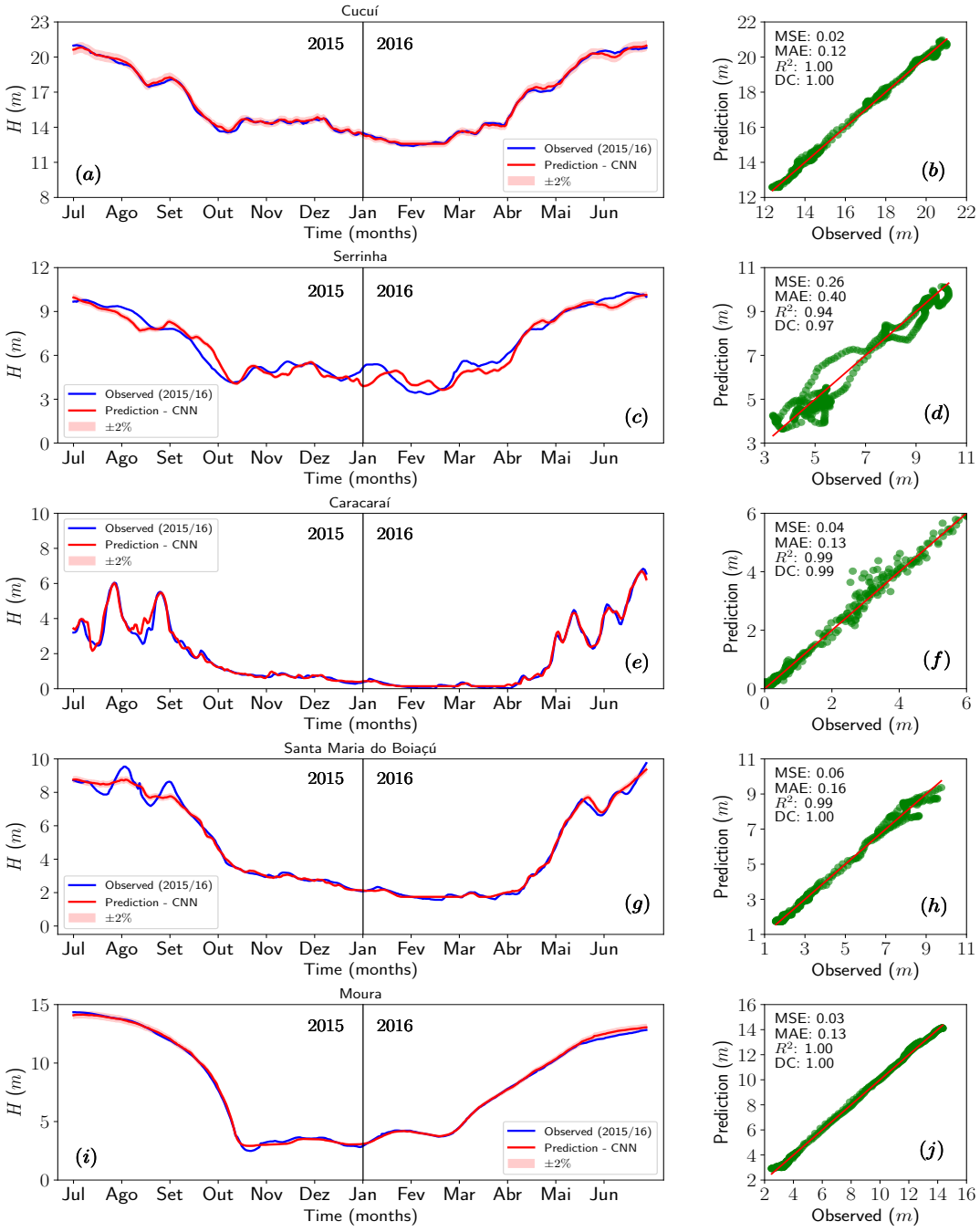


Figura 11 – Nível da água (H) ao longo dos meses para: (a) Cucuí, (c) Serrinha, (e) Caracaraí, (g) Santa Maria do Boiaçú, (i) Moura, durante o período de seca de 2015/2016, e seus respectivos gráficos de dispersão de desempenho em (b), (d), (f), (h) e (j). A curva azul representa os dados observados, enquanto a curva vermelha corresponde às previsões.

as flutuações características da dinâmica da seca nesta estação. As métricas de desempenho, apresentadas na Fig. 11(b), validam ainda mais a precisão do modelo. Notavelmente, o nível mínimo de água registrado durante esta seca foi de $12,39\ m$, marcando-o como o quinto mais baixo. Todas as métricas de avaliação de desempenho mostradas na Fig. 11, se encontram na Tabela 9.

Tabela 9 – Métricas de desempenho para diferentes estações.

Estações	MSE	MAE	R^2	DC
Cucuí	0,02	0,12	1,00	1,00
Serrinha	0,26	0,40	0,94	0,97
Caracaraí	0,04	0,13	0,99	0,99
Santa Maria do Boiaçú	0,06	0,16	0,99	1,00
Moura	0,03	0,13	1,00	1,00

Na Fig. 11(e), Caracaraí revela um padrão hidrológico complexo. De julho a setembro, três repiquetes significativos são observados, seguidos por um declínio prolongado e constante no fluxo do Rio Branco, que dura aproximadamente sete meses. Esse declínio transita para uma fase de recuperação iniciada em abril, marcada por um aumento gradual e três repiquetes adicionais entre maio e junho de 2016. A curva prevista pela CNN captura efetivamente essas flutuações intricadas, refletindo de perto os níveis observados. Esse forte alinhamento é evidenciado ainda mais no gráfico de dispersão, onde os pontos de dados seguem de perto uma tendência linear. Além disso, o desempenho do modelo é apoiado por excelentes métricas, incluindo baixos valores de erro (MSE = 0.04, MAE = 0.13) e altos indicadores de correlação ($R^2 = 0,99$, $DC = 0,99$), conforme mostrado na Fig. 11(f). O nível mínimo de água observado durante esta seca em Caracaraí foi registrado em $-0,01\ m$, o terceiro valor mais baixo, conforme indicado na Tabela 7.

As Figs. 11(g) e 11(c) mostram as secas de 2015 nas estações Santa Maria do Boiaçú e Serrinha, respectivamente, revelando um padrão hidrológico consistente com as estações a montante Caracaraí e Cucuí. Em Santa Maria do Boiaçú, a CNN não conseguiu prever os dois repiquetes observados entre julho e setembro. No entanto, demonstrou um alto grau de precisão ao capturar a curva geral de recessão, incluindo o nível mínimo de água observado de $1,56\ m$, que representa a seca mais severa registrada nesta estação. Para a estação Serrinha, as previsões da CNN exibiram mais oscilações do que os dados observados, mas ainda seguiram a tendência geral do ciclo de seca. Os gráficos de dispersão e as métricas de desempenho correspondentes, mostrados nas Figs. 11(d) e 11(h), indicam um desempenho ligeiramente melhor para Santa Maria do Boiaçú. Esses resultados sugerem que o desempenho da CNN foi comparativamente mais fraco para a estação Serrinha durante tanto a cheia de 2021 quanto a seca de 2015. Essa discrepância pode ser atribuída à complexidade dos dados de treinamento e validação, que provavelmente representaram desafios adicionais para o modelo durante o processo de aprendizado, além dos fatores geológicos, ambientais e climáticos previamente mencionados. E isto será investigado em trabalhos futuros.

A Fig. 11(i) apresenta a seca de 2015 na estação Moura, que se distingue das outras estações por exibir um comportamento hidrológico caracterizado por transições sazonais mais suaves e menos oscilações. Esse padrão consistente permite previsões mais precisas pela CNN, que captura efetivamente a tendência observada ao longo de todo o ciclo da seca. A curva predita segue de perto tanto a queda do fluxo quanto a recuperação gradual que leva ao aumento das

águas. As métricas de desempenho, mostradas na Fig. 11(j), validam o forte desempenho do modelo, com valores de erro baixos e altas taxas de correlação, indicando que a previsão está intimamente alinhada com os dados observados. Isso demonstra a capacidade da CNN de replicar padrões hidrológicos mais suaves e contínuos, como observado em Moura. A boa generalização do modelo nesta estação pode ser atribuída à consistência e regularidade dos dados temporais, o que permitiu que o modelo se adaptasse bem aos padrões sazonais estáveis observados nesta região, uma característica diretamente relacionada à posição geográfica de Moura dentro da bacia em relação às outras estações.

4.3 Associação dos erros com fatores geoambientais

A partir dos erros associados aos dados previstos pelas CNNs como MSE e MAE em relação aos dados observados, faremos agora uma associação com o intuito de tentar verificar alguma relação entre as porcentagens de tipos de rochas, precipitação, cobertura do solo e tipos de solo nas diferentes regiões da Bacia do Rio Negro mostradas na Fig. 5. Para classificação dos tipos de solo, usamos o Sistema Brasileiro de Classificação de Solos (SiBCS) (EMBRAPA, 2025), e o documento internacional World Reference Base for Soil Resources (WRB) (ISRIC, 2025b). O mapa de solos da Bacia do Rio Negro que é apresentado aqui é fornecido pela Soil and Terrain Database for Latin America and the Caribbean (SOTERLAC) (ISRIC, 2025a). E o padrão utilizado de classificação é o WRB. Na Fig. 12 é mostrado os tipos de solo pela classificação WRB, mas faremos a correspondência de acordo com o SiBCS, classificação brasileira, como pode ser visto na Tabela 10.

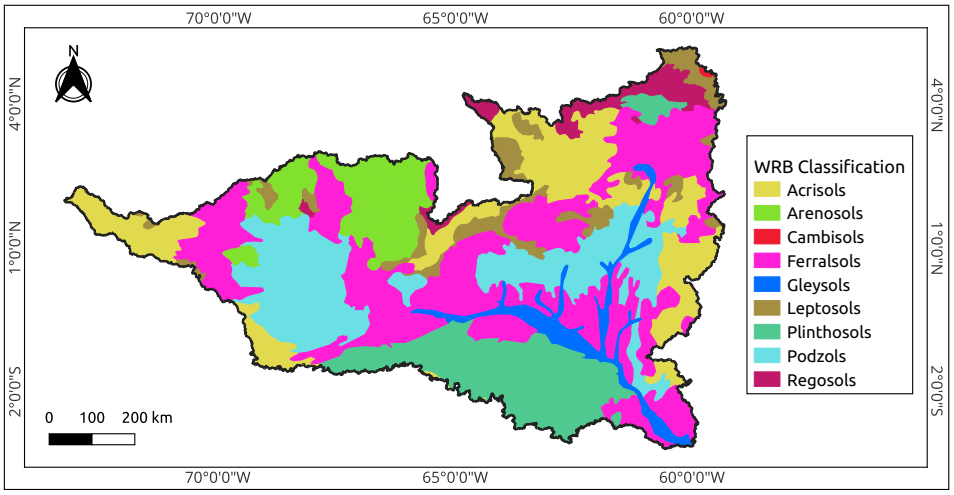


Figura 12 – Tipos de solo na Bacia do Rio Negro segundo a classificação WRB.

Na Tabela 11, apresentamos os valores das porcentagens dos erros MSE e MAE para a cheia de 2021 e a seca de 2015, além das porcentagens dos tipos de rochas, precipitação, cobertura do solo e tipos de solo para as diferentes regiões a montante de cada estação da bacia, conforme definidas na Fig. 5. Nesta tabela, para cada valor de erro, estão associadas apenas as duas maiores

Tabela 10 – Correspondência de classes de solo entre WRB e SiBCS.

WRB	**SiBCS**
Acrisols	Argissolos
Arenosols	Neossolos Quartzarênicos
Cambisols	Cambissolos
Ferralsols	Latossolos
Gleysols	Gleissolos
Leptosols	Neossolos Litólicos
Plinthosols	Plintossolos
Podzols	Espodossolos
Regosols	Neossolos Regolíticos

porcentagens de contribuições dos fatores geoambientais nessas regiões. Essa abordagem permite identificar os fatores dominantes que influenciam os erros observados, facilitando a análise das relações entre características ambientais e as variações hidrológicas. Dessa forma, é possível compreender melhor os impactos desses fatores na dinâmica da bacia e nas discrepâncias entre os cenários analisados.

Tabela 11 – MSE e MAE e porcentagem dos fatores geoambientais para as diferentes regiões da bacia.

Região	Cheia de 2021		Seca de 2015		Precipitação	Rochas		Cobertura do Solo		Tipos de Solo	
	MSE	MAE	MSE	MAE	(%)	Tipo	(%)	Tipo	(%)	Solo	(%)
A (Cucuí)	0,04	0,16	0,02	0,12	11,74	Ígnea	55,93	Formação Florestal	91,74	Arenosols	40,02
						Sedimentar	35,49	Formação de Savana	4,45	Ferralsols	26,10
B (Serrinha)	0,21	0,36	0,26	0,40	22,45	Sedimentar	48,49	Formação Florestal	94,78	Ferralsols	30,48
						Ígnea	47,51	Floresta Inundável	2,25	Podzols	23,53
C (Caracaraí)	0,09	0,23	0,04	0,13	7,15	Ígnea	71,69	Formação Florestal	65,21	Acrisols	32,35
						Sedimentar	21,68	Campo	28,34	Ferralsols	30,54
D (Santa Maria do Boiaçú)	0,26	0,35	0,06	0,16	14,06	Ígnea	57,37	Formação Florestal	74,24	Ferralsols	32,47
						Sedimentar	36,73	Campo	19,81	Acrisols	25,55
E (Moura)	0,02	0,12	0,03	0,13	44,60	Ígnea	55,92	Formação Florestal	85,60	Ferralsols	29,56
						Sedimentar	37,46	Campo	8,92	Podzols	18,14

Na Fig.13, apresentam-se os spider maps (ou mapas de teia de aranha, também conhecidos como mapas de radar) que associam os erros MSE e MAE para a cheia de 2021 (ver Fig. 10) e para a seca de 2015 (ver Fig. 11) com as porcentagens dos fatores geoambientais predominantes nas diferentes regiões da bacia. Na Fig. 13(a), observa-se o spider map para a região a montante da estação de Cucuí. Nota-se que a maior predominância (representada pela região azul) está associada à combinação entre a Formação Florestal, Precipitação, Rochas Ígneas e Arenosols (Neossolos Quartzarênicos). Como segunda contribuição significativa (região laranja), destaca-se a relação entre a Formação de Savana, Precipitação, Rochas Sedimentares e Ferralsols (Latossolos).

De Cucuí para Serrinha (Fig. 13(b)), observam-se algumas mudanças significativas. Há um aumento na presença de Ferralsols (Latossolos) e um crescimento na proporção de Podzols

(Espodossolos). Segundo o SiBCS (EMBRAPA, 2025), os Latossolos são solos altamente intemperizados, formados a partir de intensas transformações no material de origem. Já os Espodossolos são solos predominantemente arenosos, com baixa fertilidade devido à reduzida reserva de nutrientes, desenvolvendo-se principalmente em materiais arenoquartzosos sob condições de umidade elevada. Além disso, nota-se um equilíbrio entre Rochas Sedimentares e Rochas Ígneas, bem como uma substituição da dominância da Formação de Savana pela Floresta Inundável. Essas mudanças podem estar associadas aos valores elevados de MSE e MAE observados nas previsões das CNNs para Serrinha.

Na Fig. 13(c), referente à região a montante da estação de Caracaraí, observa-se a dominância dos seguintes fatores geoambientais (região azul): Formação Florestal, baixa precipitação, predominância de Rochas Ígneas e Acrisols (Argissolos). De acordo com o SiBCS, os Argissolos são caracterizados por uma camada de argila mais densa e compacta. Como segunda maior influência (região laranja), nota-se a combinação de baixo percentual de Rochas Sedimentares, baixa precipitação, presença de Campos e Ferralsols (Latossolos), que representa praticamente o mesmo percentual do tipo de solo da região de maior dominância. Isso indica que a composição do solo predominante da região azul e laranja é bastante semelhante, com diferenças sutis.

De Caracaraí para Santa Maria do Boiaçú (Fig. 13(d)), observa-se uma redução na presença de Rochas Ígneas e um aumento de Rochas Sedimentares. Há também um crescimento da Formação Florestal, uma diminuição das áreas de Campos e uma leve redução na ocorrência de Acrisols (Argissolos). Além disso, verifica-se um aumento na porcentagem de precipitação. Essas mudanças podem estar relacionadas aos elevados valores de MSE e MAE nas previsões para a cheia, enquanto para a seca, os erros foram menores, em contraste com o observado em Serrinha.

Na Fig. 13(e) (região montante à estação de Moura), observa-se que a maior dominância (região azul) ocorre na combinação de Formação Florestal, alta precipitação, Rochas Ígneas e Ferralsols (Latossolos). Em seguida, na região laranja, destacam-se a alta precipitação, Rochas Sedimentares, Campos e Podzols (Espodossolos). Nota-se aqui uma combinação de geofatores semelhantes aos das estações de Serrinha e Santa Maria do Boiaçú juntas. Os erros MSE e MAE para Moura foram baixos, como já observado anteriormente.

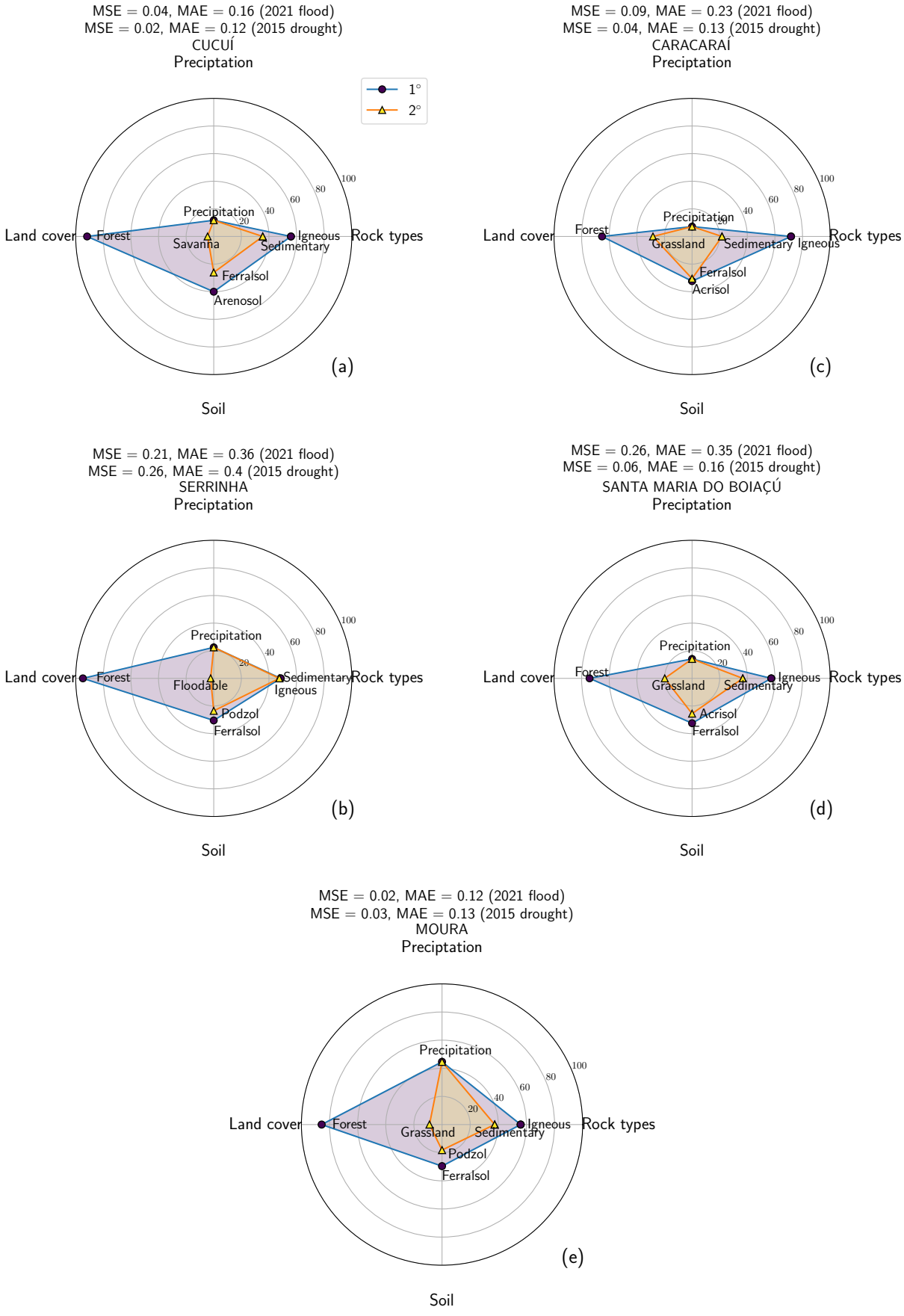


Figura 13 – Associação das porcentagens dos fatores geoambientais das regiões de contribuição das áreas a montante de cada estação da bacia.

5 Conclusões

Este estudo demonstra a aplicação bem-sucedida de Redes Neurais Convolucionais (CNNs) na previsão dos comportamentos hidrológicos durante eventos de seca e cheia extremas na Bacia do Rio Negro, com foco na seca de 2015 e na cheia de 2021. O modelo capturou de maneira eficaz as flutuações dinâmicas nos níveis da água, incluindo quedas acentuadas, recuperações graduais e aumentos durante as enchentes, demonstrando sua capacidade de replicar padrões temporais complexos. Os baixos valores de erro (MSE e MAE) e os altos coeficientes de correlação (R^2 e DC) observados nas métricas de desempenho validam a robustez do modelo.

Estações como Caracaraí e Cucuí apresentaram flutuações pronunciadas, que a CNN rastreou com precisão. No entanto, pequenas discrepâncias foram observadas durante a previsão de pequenos picos nas estações como Santa Maria do Boiaçú e Serrinha. O desempenho do modelo pode ser aprimorado com o aumento da quantidade de dados (frequência de medição) de treinamento e validação, o que proporcionará uma representação mais abrangente da dinâmica hidrológica. Além disso, o ajuste fino dos hiperparâmetros do modelo pode ajudar a melhorar sua precisão, permitindo que ele capture melhor as complexidades dos dados. Para a estação Moura, o modelo previu muito bem o comportamento real, principalmente devido à estabilidade dos padrões sazonais.

A geologia, o uso do solo e os padrões de precipitação são fatores-chave que influenciam a variação nos níveis da água dos rios, pois afetam o escoamento, a infiltração e a retenção de água dentro da bacia. Essas variáveis ambientais são determinantes para entender o comportamento hidrológico regional, pois áreas com diferentes composições geológicas, tipos de solo e cobertura vegetal experimentam padrões variados de fluxo de água e respostas a secas. A inclusão de fatores adicionais, como mudanças no uso do solo, urbanização e fenômenos climáticos, é essencial para entender a complexidade da dinâmica hidrológica. Incorporar essas variáveis proporcionará uma compreensão mais abrangente dos padrões de fluxo de água regionais.

O sucesso desta abordagem destaca o potencial das CNNs para aprimorar as previsões de secas e cheias e o gerenciamento dos recursos hídricos, particularmente em regiões com dinâmicas sazonais complexas. Embora desafios persistam em determinadas estações, a capacidade do modelo de generalizar através de padrões hidrológicos diversos aponta para sua aplicabilidade mais ampla em previsões ambientais. Melhorias futuras podem envolver a incorporação de variáveis adicionais, como sedimentos, descarga, parâmetros físico-químicos, perfilagem Doppler e radiometria, juntamente com dados de satélite. Essas adições poderiam aprimorar a capacidade do modelo de prever eventos mais localizados e extremos, melhorando, assim, as previsões em tempo real e as estratégias de gestão para áreas propensas a secas e cheias.

Além disso, a transição de aprendizado supervisionado para não supervisionado com

autoencoders permitirá que o modelo capture padrões mais complexos e não lineares nos dados. O modelo CNN pode servir como uma ferramenta adicional para auxiliar os modelos existentes utilizados pelas autoridades de monitoramento, em vez de ser visto como uma solução completa. Dada a complexidade de monitorar não apenas a Bacia do Rio Negro, mas também toda a região Amazônica, o modelo tem o potencial de fornecer suporte valioso. No entanto, para que o modelo atinja desempenho ideal nas previsões, ele requer o ajuste fino dos hiperparâmetros e dados históricos confiáveis. Com esses ajustes, ele pode se tornar uma ferramenta promissora para auxiliar as autoridades na gestão hidrológica e na tomada de decisões nesse ambiente complexo e dinâmico. Expandir as sub-redes para incluir outras subbacias da bacia amazônica proporcionará uma compreensão mais abrangente dos comportamentos hidrológicos regionais e melhorará o gerenciamento dos recursos hídricos, oferecendo insights mais profundos para previsões ambientais em larga escala.

O uso de sub-redes em modelos de aprendizado profundo, como as CNNs, é promissor, mas futuras aplicações podem se beneficiar da integração de arquiteturas como Redes Neurais Recorrentes (RNN, do inglês Recurrent Neural Network), Memória de Longo Curto Prazo (LSTM, do inglês Long Short-Term Memory), Redes Generativas Adversariais (GANs, do inglês Generative Adversarial Networks) e modelos híbridos. As RNNs e LSTMs, com sua capacidade de capturar dependências temporais, podem melhorar a previsão de séries temporais hidrológicas, enquanto as GANs podem gerar dados sintéticos para enriquecer os modelos. Modelos híbridos, que combinam diferentes arquiteturas, podem integrar as forças específicas de cada abordagem, aprimorando as previsões e o gerenciamento de recursos hídricos, especialmente em regiões como a Amazônia, onde o monitoramento é fundamental.

6 Considerações finais

A aplicação de Redes Neurais Convolucionais (CNNs) na previsão de eventos hidrológicos extremos na Bacia do Rio Negro representa uma contribuição significativa tanto para a ciência quanto para a formulação de políticas públicas voltadas à gestão de recursos hídricos. O modelo desenvolvido neste estudo demonstra grande potencial para aprimorar o monitoramento ambiental em tempo real, oferecendo suporte técnico e preditivo às autoridades responsáveis por ações emergenciais durante secas e cheias. Ao fornecer previsões mais precisas e detalhadas sobre o comportamento dos níveis dos rios, especialmente em áreas críticas, essa abordagem pode subsidiar estratégias mais eficazes de prevenção e resposta a desastres naturais, além de apoiar decisões relacionadas ao uso do solo, infraestrutura e planejamento urbano. A integração futura com variáveis ambientais adicionais e outras arquiteturas de aprendizado profundo pode fortalecer ainda mais a capacidade do modelo de antecipar cenários hidrológicos complexos. Dessa forma, o trabalho não apenas amplia o conhecimento técnico sobre a dinâmica fluvial da região amazônica, como também oferece uma base concreta para o desenvolvimento de políticas públicas orientadas por dados, com foco na sustentabilidade, na segurança hídrica e na resiliência das comunidades locais.

Referências

AHMAD, M. et al. Development of machine learning flood model using artificial neural network (ann) at var river. *Liquids*, v. 2, p. 147–160, 2022. ISSN 2673-8015. Nenhuma citação no texto.

ANA. *Agência Nacional de Águas e Saneamento Básico*. 2024. Accessed: December 11, 2024. Disponível em: <https://www.gov.br/ana/pt-br>. Nenhuma citação no texto.

BRUGNAGO, E. L.; FELICIO, C. C.; BEIMS, M. W. Forecasting the duration of three connected wings in a generalized lorenz model. *International Journal of Bifurcation and Chaos*, v. 32, p. 2230031, 2022. Nenhuma citação no texto.

BRUGNAGO, E. L.; GALLAS, J. A. C.; BEIMS, M. W. Machine learning, alignment of covariant lyapunov vectors, and predictability in rikitake's geomagnetic dynamo model. *Chaos*, v. 30, p. 083106, 2020. Nenhuma citação no texto.

BRUGNAGO, E. L.; GALLAS, J. A. C.; BEIMS, M. W. Predicting regime changes and durations in lorenz's atmospheric convection model. *Chaos*, v. 30, p. 103109, 2020. Nenhuma citação no texto.

BRUGNAGO, E. L. et al. Classification strategies in machine learning techniques predicting regime changes and durations in the lorenz system. *Chaos*, v. 30, p. 053101, 2020. Nenhuma citação no texto.

CHEBII, S. J.; MUKOLWE, M. M.; ONG'OR, B. I. River flow modelling for flood prediction using artificial neural network in ungauged perkerra catchment, baringo county, kenya. *Water Practice and Technology*, v. 17, p. 914–929, 2022. ISSN 1751-231X. Nenhuma citação no texto.

CHEVUTURI, A. et al. Forecasting annual maximum water level for the negro river at manaus. *Climate Resilience and Sustainability*, v. 1, n. 1, p. e18, 2022. Nenhuma citação no texto.

CHEVUTURI, A. et al. Forecasting annual maximum water level for the negro river at manaus using dynamical seasonal predictions. *Climate Services*, v. 30, p. 100342, 2023. Nenhuma citação no texto.

CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Comput Sci*, v. 7, p. e623, 2021. ISSN 2376-5992. Nenhuma citação no texto.

CHIRPS. *CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations*. 2024. Accessed: December 11, 2024. Disponível em: <https://www.chc.ucsb.edu/data/chirps>. Nenhuma citação no texto.

CHOLLET, F. *Deep Learning with python*. 1. ed. [S.l.]: Manning Publications Co, 2019. ISBN 9781617294433. Nenhuma citação no texto.

CHOWDHURY, M.; ALOUANI, A.; HOSSAIN, F. Comparison of ordinary kriging and artificial neural network for spatial mapping of arsenic contamination of groundwater. *Stochastic Environmental Research and Risk Assessment*, v. 24, p. 1–7, 2010. ISSN 1436-3259. Nenhuma citação no texto.

COOMES, O. T. et al. Amazon river flow regime and flood recessional agriculture: Flood stage reversals and risk of annual crop loss. *Journal of Hydrology*, v. 539, p. 214–222, 2016. ISSN 0022-1694. Nenhuma citação no texto.

DIBIKE, Y.; SOLOMATINE, D. River flow forecasting using artificial neural networks. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, v. 26, p. 1–7, 2001. ISSN 1464-1909. Nenhuma citação no texto.

dos Reis, M. et al. Forest fires and deforestation in the central amazon: Effects of landscape and climate on spatial and temporal dynamics. *Journal of Environmental Management*, v. 288, p. 112310, 2021. ISSN 0301-4797. Nenhuma citação no texto.

Dos Santos, V.; STEVAUX, J.; SZUPIANY, R. Confluence analysis at basin scale in a tropical bedrock river – the ivaí river, southern brazil. *Journal of South American Earth Sciences*, v. 116, p. 103877, 2022. ISSN 0895-9811. Nenhuma citação no texto.

DUBROEUCQ, D.; VOLKOFF, B. From oxisols to spodosols and histosols: evolution of the soil mantles in the rio negro basin (amazonia). *CATENA*, v. 32, n. 3, p. 245–280, 1998. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0341816298000459>. Nenhuma citação no texto.

ELEUTéRIO, C. L. et al. Identifying wildfires with convolutional neural networks and remote sensing: application to amazon rainforest. *International Journal of Remote Sensing*, Taylor & Francis, v. 0, n. 0, p. 1–22, 2024. Nenhuma citação no texto.

EMBRAPA. *Sistema Brasileiro de Classificação de Solos (SiBCS)*. 2025. Acesso em: 2 mar. 2025. Disponível em: <https://www.embrapa.br/solos/sibcs>. Nenhuma citação no texto.

ESPINOZA, J.-C. et al. The new historical flood of 2021 in the amazon river compared to major floods of the 21st century: Atmospheric features in the context of the intensification of floods. *Weather and Climate Extremes*, v. 35, p. 100406, 2022. ISSN 2212-0947. Nenhuma citação no texto.

FILIZOLA, N.; GUYOT, J. L. Suspended sediment yields in the amazon basin: an assessment using the brazilian national data set. *Hydrological Processes*, v. 23, n. 22, p. 3207–3215, 2009. Nenhuma citação no texto.

FILIZOLA, N.; GUYOT, J.-L. O fluxo de sedimentos em suspensão nos rios da amazônia brasileira. *Revista Brasileira de Geociencias*, v. 41, p. 566–576, 12 2011. Disponível em: <https://www.documentation.ird.fr/hor/fdi:010070669>. Nenhuma citação no texto.

FONSECA, F. et al. Hydrological scenarios and malaria incidence in the amazonian context. *Water*, v. 14, n. 8, 2022. ISSN 2073-4441. Nenhuma citação no texto.

FRAPPART, F. et al. Floodplain water storage in the negro river basin estimated from microwave remote sensing of inundation area and water levels. *Remote Sensing of Environment*, v. 99, n. 4, p. 387–399, 2005. ISSN 0034-4257. Nenhuma citação no texto.

GARCIA, B. N.; LIBONATI, R.; NUNES, A. M. B. Extreme drought events over the amazon basin: The perspective from the reconstruction of south american hydroclimate. *Water*, v. 10, n. 11, 2018. Nenhuma citação no texto.

GeoSGB. *GeoSGB - Sistema de Informação Geográfica do Serviço Geológico do Brasil*. 2024. Accessed: December 11, 2024. Disponível em: <https://geosgb.sgb.gov.br/>. Nenhuma citação no texto.

GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 3. ed. [S.l.]: O'Reilly Media, 2022. ISBN 9781098125974; 1098125975. Nenhuma citação no texto.

GETIRANA, A. C. V. *Contribuições da altimetria espacial à modelagem hidrológica de grandes bacias na amazônia*. Tese (Doutorado) — Universidade Federal do Rio Janeiro, Rio de Janeiro, RJ, 2009. Disponível em: <http://www.coc.ufrj.br/pt/teses-de-doutorado/153-2009/1192-augusto-cesar-vieira-getirana>. Nenhuma citação no texto.

GRANATO-SOUZA, D.; STAHLE, D. W. Drought and flood extremes on the amazon river and in northeast brazil, 1790–1900. *Journal of Climate*, American Meteorological Society, v. 36, n. 20, p. 7213 – 7229, 2023. Nenhuma citação no texto.

GUALTIERI, C. et al. A field study of the confluence between negro and solimões rivers. part 1: Hydrodynamics and sediment transport. *Comptes Rendus Geoscience*, v. 350, n. 1, p. 31–42, 2018. ISSN 1631-0713. Nenhuma citação no texto.

GUYOT, J.-L. et al. Balanço hídrico da bacia do rio negro. *Anais do X Simpósio Brasileiro de Recursos Hídricos, Gramado, 11p*, Gramado, 11p, 1993. Nenhuma citação no texto.

HA, S.; LIU, D.; MU, L. Prediction of yangtze river streamflow based on deep learning neural network with el niño–southern oscillation. *Scientific Reports*, v. 2, p. 11738, 2021. ISSN 2045-2322. Nenhuma citação no texto.

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. *Science*, v. 313, n. 5786, p. 504–507, 2006. Nenhuma citação no texto.

HOLANDA, J. L. R.; MARMOS, J. L.; MAIA, M. A. M. Geodiversidade do estado de roraima. *CPRM*, p. 252, 2014. Disponível em: <https://rigeo.sgb.gov.br/handle/doc/16775>. Nenhuma citação no texto.

HYBAM. *SO-HYBAM Amazon Basin Water Resources Observation Service*. 2024. Accessed: December 11, 2024. Disponível em: <https://hybam.obs-mip.fr/pt/website-under-development-4/>. Nenhuma citação no texto.

ISRIC. *Soil Data - Geonetwork Record*. 2025. Acesso em: 2 mar. 2025. Disponível em: <https://data.isric.org/geonetwork/srv/api/records/436bd4b0-7ffc-4272-be57-686b7d7eea7d>. Nenhuma citação no texto.

ISRIC. *World Reference Base for Soil Resources (WRB)*. 2025. Acesso em: 2 mar. 2025. Disponível em: <https://www.isric.org/explore/wrb>. Nenhuma citação no texto.

JIMéNEZ-MUñOZ, J. C. et al. Record-breaking warming and extreme drought in the amazon rainforest during the course of el niño 2015–2016. *Scientific Reports*, v. 6, n. 1, p. 33130, 2016. ISSN 2045-2322. Nenhuma citação no texto.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143. ISBN 1558603638. Nenhuma citação no texto.

LARAQUE, A.; GUYOT, J. L.; FILIZOLA, N. Mixing processes in the amazon river at the confluences of the negro and solimões rivers, encontro das Águas, manaus, brazil. *Hydrological Processes*, v. 23, n. 22, p. 3131–3140, 2009. Nenhuma citação no texto.

LATRUBESSE, E. M.; FRANZINELLI, E. The late quaternary evolution of the negro river, amazon, brazil: Implications for island and floodplain formation in large anabranching tropical systems. *Geomorphology*, v. 70, p. 372–397, 2005. Nenhuma citação no texto.

LIMA, L. Santos de et al. Severe droughts reduce river navigability and isolate communities in the brazilian amazon. *Communications Earth & Environment*, v. 5, n. 1, p. 370, 2024. ISSN 2662-4435. Nenhuma citação no texto.

LINEROS, M. L. et al. Optimized design of neural networks for a river water level prediction system. *Sensors*, v. 21, p. 6504, 2021. ISSN 1424-8220. Nenhuma citação no texto.

MACIEL, J. S. C. et al. Analysing the flood warning of negro river in manaus. *International Journal of Environmental Impacts: Management, Mitigation and Recovery*, v. 5, p. 51–64, 2022. Nenhuma citação no texto.

MAIA, M. A. M. Geodiversidade do estado do amazonas. *CPRM*, p. 275, 2010. Disponível em: <https://rigeo.sgb.gov.br/handle/doc/16624>. Nenhuma citação no texto.

MANTOVANI, J. et al. Flood risk mapping during the extreme february 2021 flood in the juruá river, western brazilian amazonia, state of acre. *Sustainability*, v. 16, n. 7, 2024. ISSN 2071-1050. Nenhuma citação no texto.

MapBiomas. *MapBiomas*. 2024. Accessed: 2024-12-25. Disponível em: <https://brasil.mapbiomas.org/>. Nenhuma citação no texto.

MARINHO, R. R. *Integração de dados de campo e sensoriamento remoto no estudo do fluxo de água e matéria no arquipélago de anavilhanas, Rio Negro – Amazonas, Brasil.* Tese (Doutorado) — Instituto Nacional de Pesquisas da Amazônia (INPA), 2019. Disponível em: <https://repositorio.inpa.gov.br/bitstream/1/12984/1/Tese_Rogerio_Ribeiro_Marinho.pdf>. Nenhuma citação no texto.

MARINHO, R. R.; FILIZOLA, N. P.; CREMON, H. Analysis of suspended sediment in the anavilhanas archipelago, rio negro, amazon basin. *Water*, v. 12, n. 4, 2020. Nenhuma citação no texto.

MARINHO, R. R. et al. Riverbed morphology and hydrodynamics in the confluence of complex mega rivers - a study in the branco and negro rivers, amazon basin. *Journal of South American Earth Sciences*, v. 118, p. 103969, 2022. ISSN 0895-9811. Nenhuma citação no texto.

MARINHO, R. R.; ZANIN, P. R.; FILIZOLA, N. P. The negro river in the anavilhanas archipelago: Streamflow and geomorphology of a complex anabranching system in the amazon. *Earth Surface Processes and Landforms*, v. 47, n. 4, p. 1108–1123, 2022. Nenhuma citação no texto.

MARTINEZ, J.-M.; Le Toan, T. Mapping of flood dynamics and spatial distribution of vegetation in the amazon floodplain using multitemporal sar data. *Remote Sensing of Environment*, v. 108, n. 3, p. 209–223, 2007. ISSN 0034-4257. Nenhuma citação no texto.

MASLOV, A. Glaciogenic and related sedimentary rocks: Main lithochemical features. communication 1. late archean and proterozoic. *Lithology and Mineral Resources*, v. 45, p. 377–397, 2010. Nenhuma citação no texto.

MEADE, R. H. et al. Backwater effects in the amazon river basin of brazil. *Environmental Geology and Water Sciences*, v. 18, n. 2, p. 105–114, 1991. ISSN 1432-0495. Nenhuma citação no texto.

MENDES, C. F. O.; BEIMS, M. W. Distance correlation detecting lyapunov instabilities, noise-induced escape times and mixing. *Physica A: Statistical Mechanics and its Applications*, v. 512, p. 721–730, 2018. Nenhuma citação no texto.

MENDES, C. F. O. et al. Detecting correlations in desynchronized chaotic chimera states. *International Journal of Bifurcation and Chaos*, v. 32, n. 05, p. 2230012, 2022. Nenhuma citação no texto.

MENDES, C. F. O.; SILVA, R. M. da; BEIMS, M. W. Decay of the distance autocorrelation and lyapunov exponents. *Phys. Rev. E*, v. 99, p. 062206, 2019. Nenhuma citação no texto.

MENDES, T. A. A. et al. Geological knowledge advances on the alto rio negro region, northwestern amazonian craton, brazil: a review. *Journal of the Geological Survey of Brazil*, v. 4, n. 3, p. 209–222, 2021. Nenhuma citação no texto.

MOLINIER, M. et al. Les régimes hydrologiques de l'Amazone et de ses affluents. In: *L'hydrologie tropicale : géosciences et outil pour le développement : mélanges à la mémoire de Jean Rodier*. Wallingford: AISH, 1996, (Publication - AISH). p. 209–222. ISBN 0-947571-99-X. Disponível em: <https://www.documentation.ird.fr/hor/fdi:010009811>. Nenhuma citação no texto.

NOAA. *El Niño and La Niña Events: Past Events*. 2024. Accessed: December 16, 2024. Disponível em: <https://psl.noaa.gov/enso/past_events.html>. Nenhuma citação no texto.

NOAA. *El Niño-Southern Oscillation (ENSO)*. 2024. Accessed: December 20, 2024. Disponível em: <https://psl.noaa.gov/enso/>. Nenhuma citação no texto.

OpenTopography. *OpenTopography - High-Resolution Topography Data and Tools*. 2024. Accessed: December 11, 2024. Disponível em: <https://opentopography.org/>. Nenhuma citação no texto.

PEREIRA, R. C. et al. Reviewing autoencoders for missing data imputation: Technical trends, applications and outcomes. *Journal of Artificial Intelligence Research*, v. 69, p. 1255–1285, 12 2020. Nenhuma citação no texto.

PRESTES, N. C. et al. Impact of the extreme 2015-16 el niño climate event on forest and savanna tree species of the amazonia-cerrado transition. *Flora*, v. 319, p. 152597, 2024. ISSN 0367-2530. Nenhuma citação no texto.

QUESADA, C. A. et al. Soils of amazonia with particular reference to the rainfor sites. *Biogeosciences*, v. 8, n. 6, p. 1415–1440, 2011. Nenhuma citação no texto.

REIS, M. M. T. et al. Socioeconomic aspects and profile of fishing according to fishers of commercial edible fish in the municipality of barcelos, middle negro river, amazonas, brazil. *Brazilian journal of biology*, Instituto Internacional de Ecologia, v. 82, p. 1–14, 2022. Nenhuma citação no texto.

SGB. *Serviço Geológico do Brasil (SGB-CPRM)*. 2024. Accessed: December 11, 2024. Disponível em: <https://www.sgb.gov.br/>. Nenhuma citação no texto.

SHARMA, P.; SINGH, S.; SHARMA, S. D. Artificial neural network approach for hydrologic river flow time series forecasting. *Agricultural Research*, v. 11, p. 465, 2022. ISSN 2249-7218. Nenhuma citação no texto.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas*. 1. ed. São Paulo: Artliber editora Ltda, 2020. ISBN 9788588098879. Nenhuma citação no texto.

SIQUEIRA, L. F.; FILIZOLA, N. Estudo hidrológico do efeito de barramento hidráulico no rio tarumã-açú, manaus-am. *Revista Brasileira de Geomorfologia*, v. 22, n. 2, abr. 2021. Nenhuma citação no texto.

SIT, M. et al. A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, v. 82, p. 2635–2670, 2020. ISSN 0273-1223. Nenhuma citação no texto.

SZÉKELY, G. J.; RIZZO, M. L.; BAKIROV, N. K. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, The Institute of Mathematical Statistics, v. 35, p. 2769–2794, 2007. Nenhuma citação no texto.

VALENZUELA, J. et al. Flooding risk of cropland areas by repiquetes in the western amazon basin: A case study of peruvian tamshiyacu city. *Journal of Hydrology: Regional Studies*, v. 47, p. 101428, 2023. ISSN 2214-5818. Nenhuma citação no texto.

VASCONCELOS, M. A. de et al. Climate change and its impact on the agricultural calendar of riverine farmers in médio juruá, amazonas state, brazil. *Atmosphere*, v. 13, n. 12, 2022. ISSN 2073-4433. Nenhuma citação no texto.

Apêndices

Artigo 1 - Multi-station data integration for flood and drought prediction in the Negro River basin using convolutional neural networks

Artigo submetido (20/01/2025) para publicação na revista internacional indexada **Hydrological Processes**:

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Multi-station data integration for flood and drought prediction in the Negro River basin using convolutional neural networks

Cíntia L. Eleutério[1] | **Carlos F. O. Mendes**[2] | **Marcus W. Beims**[3] | **Mircea Galiceanu**[2] | **Naziano P. Filizola Júnior**[1]

[1]Departamento de Geociências, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brazil

[2]Departamento de Física, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brazil

[3]Departamento de Física, Universidade Federal do Paraná, Curitiba, 81531-980, PR, Brazil

Correspondence
Corresponding author Cíntia L. Eleutério.
Email: cintia.e.fisica@gmail.com

Abstract

We apply convolutional neural networks with subnet-based architectures to predict extreme floods and droughts in the Negro River basin, located in the northern region of South America. The basin is characterized by diverse geological formations, including a variety of rock types that influence water infiltration and flow patterns, alongside climatic factors that shape precipitation dynamics. Additionally, the basin's hydrodynamics, driven by interactions between its geological formations and water flow, as well as its topography, which govern the direction and speed of runoff, play crucial roles in its hydrological behavior. Rainfall regime and seasonality further dictate the temporal distribution of water levels. Land cover types also play a significant role in altering infiltration rates and surface runoff. By analyzing hydrological data from five stations, namely Cucuí, Serrinha, Caracaraí, Santa Maria do Boiaçú, and Moura, this research aims to model and forecast extreme hydrological events. The application of deep learning approach underscores the importance of integrating multi-station data to capture localized variations and regional differences. These findings highlight the potential of artificial intelligence as a complementary tool to existing models used by competent authorities responsible for monitoring, aiding in mitigating the impacts of extreme floods and droughts. While not a standalone solution, the model provides valuable insights and emphasizes the need for further improvements, particularly through hyperparameter tuning, to enhance its reliability and accuracy.

KEYWORDS
Hydrological forecasting, convolutional neural network, geological and climatic influences

1 | INTRODUCTION

The Negro River, located in the northwestern Amazon, is the sixth-largest river globally in terms of water discharge and is the tributary which carries the largest discharge to the Amazon River (Frappart et al., 2005). It is characterized by a complex multi-channel system and a distinctive geological structure (Marinho et al., 2022). Serving as the primary watercourse of the basin, holds particular significance, especially as it merges with the Solimões River to form the mighty Amazon River (Laraque, Guyot, & Filizola, 2009). The fluvial dynamics of the Negro River basin are exceptionally intricate, prompting extensive research efforts aimed at unraveling its complexities and enhancing our understanding of this basin (Marinho, Zanin, & Filizola Junior, 2022; Marinho, Filizola Junior, & Cremon, 2020). The basin not only has ecological significance but also holds substantial socioeconomic importance in the region (Reis et al., 2022). Modeling the Negro River has been instrumental in comprehending its complex dynamics, focusing on preventing disasters resulting from extreme floods and droughts that have become more pronounced in recent years (Maciel et al., 2022; Chevuturi et al., 2023, 2022).

The Negro River basin is a sub-basin of the Amazon basin, recognized as one of the largest tropical basins in the world. The Amazon basin covers approximately 6.1 million km^2, representing about 5% of the planet's continental landmass. This vast hydrological system discharges around 6.6×10^{12} m^3 of freshwater into the ocean annually, accounting for 16% to 20% of the global continental water reserves (Molinier et al., 1996). Its climatic and topographic diversity is remarkable. The western

Artigo 2 - Artificial neural networks tackling extreme hydrological challenges: application to the Negro River

Artigo submetido (04/02/2025) para publicação na revista internacional indexada **Hydrological Sciences Journal**:

ARTICLE TEMPLATE

Artificial neural networks tackling extreme hydrological challenges: application to the Negro River

Cíntia L. Eleutério[a], Naziano P. Filizola[a], Carlos F. O. Mendes[b], Mircea Galiceanu[b], Marcus W. Beims[c]

[a]Departamento de Geociências, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brazil; [b]Departamento de Física, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brazil; [c]Departamento de Física, Universidade Federal do Paraná, Curitiba, 81531-980, PR, Brazil

ARTICLE HISTORY
Compiled February 4, 2025

Abstract
We investigate the use of artificial neural networks in predicting extreme floods and droughts in the Negro River basin, situated in northern South America. By utilizing historical data from two key monitoring stations, Moura and Cucuí, the research delves into the intricate dynamics of water levels characteristic of the region. To achieve accurate forecasting of extreme hydrological events, three deep learning models were applied, showcasing the growing role of these advanced techniques in water resource management. The analysis shows the critical importance of historical records from these stations in improving our understanding of hydrological behavior. Despite notable differences in water-level patterns between Moura and Cucuí, the models demonstrate reliable predictive capabilities when their hyperparameters were properly tuned. Our findings highlight the potential for artificial neural networks to support decision-makers in anticipating and mitigating the impacts of extreme floods and droughts, complementing existing methodologies and enhancing predictive capabilities.

KEYWORDS
Hydrological Forecasting, Artificial Neural Networks, Extreme Hydrological Phenomena.

1. Introduction

The Negro River basin is a vast and intricate hydrological network within the greater Amazon basin. Encompassing several South American countries, the majority of its area lies within Brazil. Characterized by a unique geological framework and intricate multi-channel system, the basin's hydrological significance is extensive. Its fluvial dynamics are remarkably complex, prompting ongoing research to better understand its intricate processes (Marinho, Zanin, and Filizola Junior 2022; Marinho, Filizola Junior, and Cremon 2020; Marinho et al. 2022). Among its key features, the Negro River stands out as the basin's primary watercourse, gaining particular importance at its confluence with the Solimões River near the city of Manaus, where together they form the Amazon River (Laraque, Guyot, and Filizola 2009). The basin is a vital socioeconomic asset,

CONTACT Cíntia L. Eleutério. Email: cintia.e.fisica@gmail.com

supporting regional economies and sustaining livelihoods (Reis et al. 2022). Recent advances in hydrological modeling have provided key insights into managing the basin's dynamic challenges, especially nowadays as extreme floods and droughts become increasingly frequent. These developments are crucial in mitigating the risks associated with such events, which have intensified in recent years, particularly in hydrographic basins, underscoring the necessity for understanding these events (Maciel et al. 2022; Chevuturi et al. 2023, 2022; Granato-Souza and Stahle 2023; Garcia, Libonati, and Nunes 2018).

The ability to predict future outcomes based on historical and current data is essential across various research areas. There are known applications in anticipating monitoring climate variations, extreme or rare phenomena, and severe natural events like floods and droughts. Real-world problems often involve highly complex systems that present significant analytical challenges. In light of ongoing global climate shifts, which are likely influenced by a combination of natural variability and human activities, the occurrence of extreme events has risen noticeably. To address this, the present study employs Artificial Neural Networks (ANNs), a tool of growing scientific importance that has been widely explored in numerous fields. These ANNs excel at recognizing patterns and making predictions, particularly when provided with extensive datasets for training.

ANNs have found widespread use in a variety of scientific areas. In healthcare, they have been applied to early detection of breast cancer (Cireşan et al. 2013), diagnosis of COVID-19 symptoms using chest radiographs (Wang, Lin, and Wong 2020), categorizing lung disease images (Li et al. 2014), and assessing heart disease risk (Yan et al. 2006). In the field of chemistry, ANNs play a crucial role in the identification of polymeric compounds (Steiner et al. 2011) and the prediction of water quality levels (Han, Chen, and Qiao 2011). In biology, ANNs have been utilized to classify bat echolocation calls (Armitage and Ober 2010) and distinguish mouse species based on vocal patterns (Tian and Shang 2006). In physics, ANNs have been used to areas such as high energy physics (Karagiorgi and Shih 2022), quantum physics (Kessler and Kühne 2021; Wu and Chiribella 2024) or material sciences (Kulichenko 2021), to name only a few. In the realm of ecology, they are used to examine the impact of climate change on tree growth patterns (Lek and Guégan 2012). Moreover, in climatology, ANNs have been pivotal in developing robust predictive models for weather forecasting (Abhishek et al. 2012). Furthermore, these networks have been applied to identify and classify wildfires in the Amazon Rainforest (Eleutério et al. 2024). ANNs have become a crucial tool in hydrology, significantly improving the management and monitoring of water resources (Sit et al. 2020). They have been successfully applied to spatial mapping in detecting arsenic contamination in groundwater (Chowdhury, Alouani, and Hossain 2010). Additionally, ANNs are widely used for predicting extreme hydrological events, such as streamflow in the Yangtze River (Ha, Liu, and Mu 2021), flood forecasting in Kenya's Perkerra River (Chebii, Mukolwe, and Ong'or 2022), water level prediction in Spain's Carrión River (Lineros et al. 2021). Other notable applications include flow forecasting on India's Sot River (Sharma, Singh, and Sharma 2022) and flood modeling on France's Var River (Ahmad et al. 2022), highlighting the broad impact of ANNs on water resource management.

Temporal prediction, particularly with respect to the dynamics of water levels in the Negro River basin and phenomena such as floods and droughts throughout the year, presents significant complexity. To address this, we employ three deep learning models (Chollet 2019; Géron 2022): Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and 1D Convolutional Neural Network (1D-CNN). The model is trained

using historical data from two fluviometric stations, Moura and Cucuí, which are vital to understanding the hydrological behavior of the basin. Time series forecasting has traditionally been used to identify regime changes in chaotic systems, such as changes in the weather season modeled by Lorenz's atmospheric convection model (Brugnago et al. 2020; Brugnago, Gallas, and Beims 2020b) and its generalization (Brugnago, Felicio, and Beims 2022), as well as to predict Earth's magnetic field reversals (Brugnago, Gallas, and Beims 2020a).

The paper is structured as follows. Section 2 presents the data acquisition process and the analyzed area. In Section 3, we provide an overview of the three ANN models used in this study, along with a brief description of the statistical methods employed to evaluate the network's performance. Section 4 is devoted to presenting the results, while Section 5 discusses the conclusions.

2. Study area

2.1. *Data and geographic positions of the analyzed fluviometric stations*

The hydrological data analyzed in this study relate to two key fluviometric stations within the Negro River basin: Cucuí, situated upstream, and Moura, located downstream on the Negro River. The Moura station is positioned at latitude 1°27'24"S and longitude 61°38'4"W, while the Cucuí station is located at latitude 1°12'55"N and longitude 66°51'9"W. These geographic locations are illustrated in Fig. 1. The data are from the Brazilian hydrometric network, managed by the National Water and Basic Sanitation Agency (ANA) (ANA 2024). The available data for the Cucuí station include measurements of water level, sediment, discharge, water quality, and other hydrological parameters. In contrast, the Moura station provides only water level data. In this study, we focus specifically on water level data. Both stations are managed by the ANA and operated by the Geological Survey of Brazil (SGB) (SGB 2024). Strategically located within the Negro River basin, Moura and Cucuí stations provide essential hydrological data. These datasets play a critical role in advancing the understanding of basin dynamics and contribute significantly to the monitoring and management of water resources in the Amazon Basin.

Table 1. Information about the stations.

Station	Data period	Drainage area (km^2)	River	Code
Moura	1980 - 2021	593,000	Negro	14840000
Cucuí	1981 - 2021	74,300	Negro	14110000

Further details about the stations, including the data availability period, drainage area, and the specific rivers on which they are located, are summarized in Table 1. The table presents the data periods selected for this study. The codes for the fluviometric stations displayed in the table are provided by ANA. It is noteworthy that the hidrological data from Cucuí begin in 1981. This is because the records made available by the ANA only start from that year. As a result, the dataset for Cucuí includes one year less compared to Moura. However, this difference does not affect the analyses conducted in this study.

The contrasting water level dynamics (H) of the Negro River at the two stations are illustrated in Fig. 2. In Fig. 2(a), the dynamic surface of H for the Moura station is presented, covering the data period listed in Table 1. A relatively steady pattern can

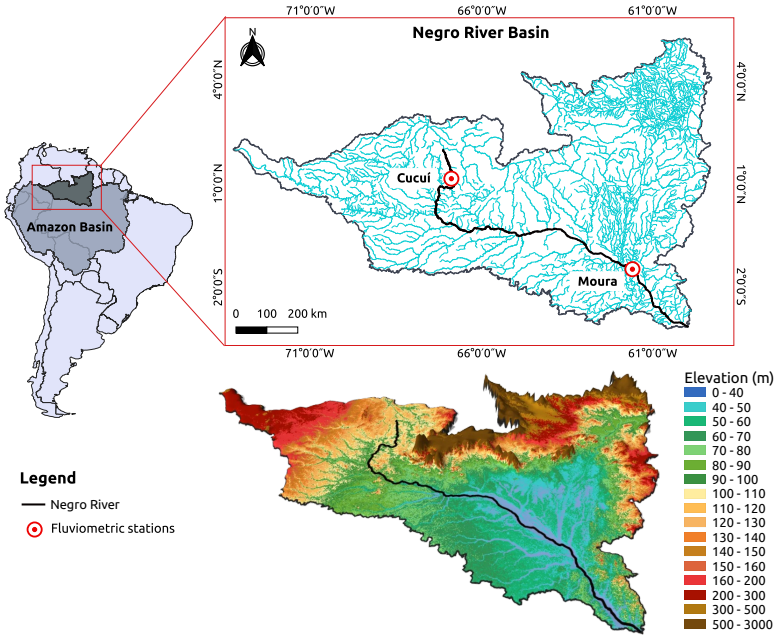


Figure 1. A comprehensive depiction of the Negro River basin, showcasing the locations of the monitoring stations. The map emphasizes key features within the basin, particularly the Negro River itself. Additionally, it illustrates the position of the Negro River basin as a sub-basin within the larger Amazon Basin. The figure also highlights the terrain elevation across the basin, with different colors to indicate varying elevation levels.

be observed over the years. In contrast, Fig. 2(b) shows the dynamic surface for the Cucuí station, where the H values reveal a distinct pattern, suggesting a more complex hydrological behavior compared to Moura. The blue region in the center indicates periods of higher values, corresponding to flood events, while the orange areas represent drought periods. The differences in the dynamics between the two stations emphasize the need to understand these variations for accurate forecasting.

The irregular patterns at Cucuí may be linked to its geographical location within the basin, influenced by geological factors and land cover, which differ from those of the Moura region, as well as the higher annual precipitation in the Upper Negro River region, which receives more rainfall than the Moura region. Furthermore, the presence of numerous tributaries both upstream and downstream of the stations exerts significant influence on the hydrological dynamics, directly impacted by hydraulic effects. The hydraulic effect occurs when a river with higher discharge impounds the flow of a smaller tributary, causing water to accumulate in the tributary's basin and altering the flow velocity and water levels in both rivers (Meade et al. 1991; Siqueira and Filizola 2021; Marinho, Filizola Junior, and Cremon 2020).

The observations at Moura exhibit a much more stable pattern, influenced by geological, land cover, and climatic factors that differ from those of the Cucuí region. It is also important to highlight that Moura station is influenced by the Branco River, which flows to the north of the basin. The confluence of the Negro and Branco Rivers

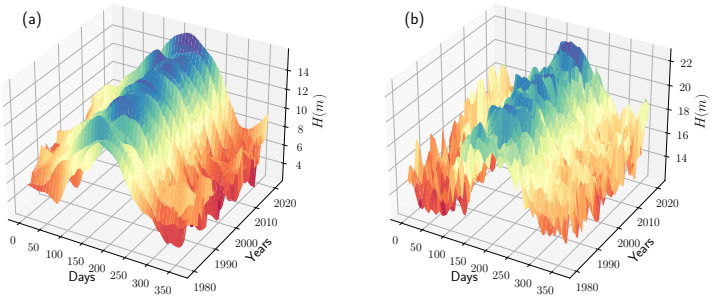


Figure 2. Surface dynamics of water level of Negro River at: (a) Moura and (b) Cucuí stations.

determines the observed hydrological behavior at Moura (Marinho et al. 2022). Additionally, factors such as geomorphology, sedimentation processes, variations in water temperature, interactions with groundwater aquifers, extreme climatic events, and various anthropogenic influences also play a significant role in shaping the hydrological dynamics of the basin. Future studies will incorporate these aspects to provide a more integrated and comprehensive analysis of the system.

In Fig. 3(a), the maximum (flood) and minimum (drought) water levels of the Negro River observed in Manaus are shown, covering the period from 1903 to 2023. The lime-colored points on the blue curve represent six significant floods selected for analysis, which are compared with data from the Moura and Cucuí stations to develop AI-based forecasts. Predictions were made using data exclusively from these two stations, strategically located at different points in the basin (see Fig. 1). The blue curve with circles illustrates flood levels, with the smallest recorded flood in 1926 at 21.77m and the largest in 2021, reaching 30.02m. The dashed blue line represents the average flood height of 27.94 m. The red curve with diamonds represents drought levels, highlighting the most severe drought in 2023, with a water level of 12.70 m. The average drought level is 17.66 m. It is important to note that flood and drought events occur during different periods of the year.

Figures 3(b) and 3(c) display data for floods and droughts at the Moura and Cucuí stations, starting in 1980 and 1981, respectively. The six extreme flood events observed in Moura correspond to those recorded in Manaus. However, except for 2021, all major floods at Cucuí occurred in years that differ from those recorded in Moura and Manaus. Consequently, the analysis for Moura and Cucuí focuses exclusively on floods that also occurred in Manaus. For drought analysis, the 2009 event was examined for Moura and Cucuí. This year did not present an extreme drought in Manaus. However, since an extreme drought was observed at both stations, this year was chosen for ANN predictions.

Figure 4 illustrates the observed water level at the two stations over the years, spanning from 1981 to 2021 for Cucuí and from 1980 to 2021 for Moura. The shaded regions represent the range of measurements throughout the considered period. In Fig. 4(a), the variation of H is shown as a function of the months, with values ranging from 12 to 23 meters. Finally, Fig. 4(b) presents the variation in H exclusively for the Moura station, where the water level fluctuates approximately between 2 and 15 meters over time. Notably, the observations at Cucuí exhibit significant variability, while those

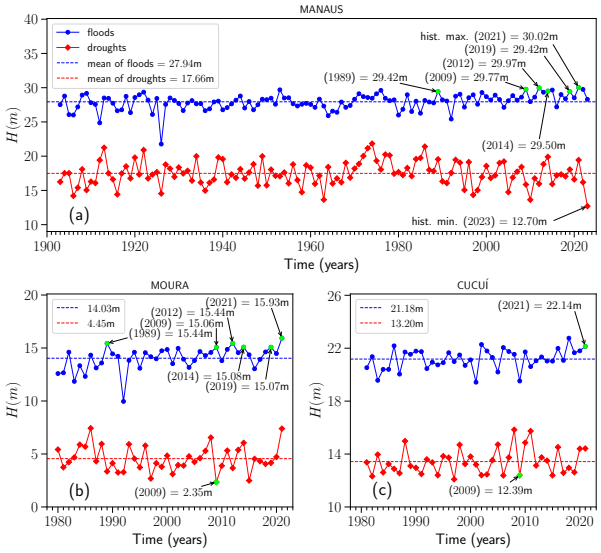


Figure 3. Annual maximum and minimum water levels (H) of Negro River recorded (a) in Manaus (1903–2023), (b) at the Moura station (1980–2021), and (c) at the Cucuí station (1981–2021). The lime-colored markers on the curves highlight the years of significant floods and droughts selected for analysis using ANNs.

at Moura appear much more stable throughout the considered period.

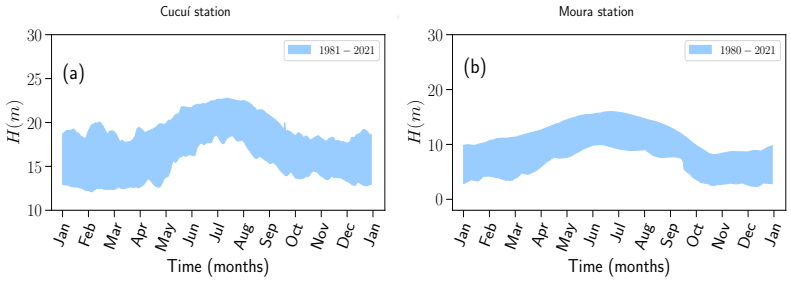


Figure 4. Water level for: (a) Cucuí and (b) Moura stations. The shaded region indicates the range of variation observed during the measurement periods at the respective stations.

2.2. *Hydrology, geological formations and precipitation pattern*

The Negro River basin spans 12% of the Amazon Basin, covering an area of $712,000\ km^2$, with 82% of it located in Brazil. The river flows for about $900\ km$ to Manaus and is joined by major tributaries such as the Branco River, which drains an area of $189,925\ km^2$. During flood seasons, vast plains become inundated, influencing the hydrological cycle of the basin (Getirana 2009). Flooding typically peaks from May to July, marking the most critical period of the hydrological year (Guyot

et al. 1993). The basin is primarily covered by tropical forests (89.3%) and savannahs (8.5%) (Martinez and Le Toan 2007), with significant water bodies such as rivers, lakes, and floodplains. Around 6.4% of the basin's surface is permanently waterlogged, while 11.5% experiences seasonal flooding. The upper Negro River flows through a region in Colombia, composed of sedimentary plains and savannahs, underlaid by Pre-Cambrian rocks from the Guiana Shield. The terrain here is mostly flat, ranging from 60 to 160 meters in elevation, but isolated inselbergs rise up to 700 meters (Latrubesse and Franzinelli 2005).

Topographic variations in the basin are shown in Fig. 1, with northern regions reaching elevations between 500 and 3000 meters. Moving westward, the land gradually lowers to between 100 and 300 meters. The central and southern areas are mostly flat, with elevations as low as 100 meters. These differences in elevation play a crucial role in shaping the hydrological dynamics of the basin, influencing flow patterns, water levels, and interactions between the river and its landscape, particularly during extreme flooding or drought events. The basin spans both the Guiana Shield and the Amazonian plain, influencing its unique hydrological characteristics. Notably, water volume fluctuations are more pronounced at the confluence of the main river and its tributaries, with widespread flooding during the wet season (Martinez and Le Toan 2007; Frappart et al. 2005).

Figure 5 illustrates two regions, Region A, which includes the Cucuí station, and Region B, which includes the Moura station. These regions highlight the different geological formations specific to these areas within the Negro River basin. The data shown in the figure are sourced from the SGB (GeoSGB 2024), and it is important to emphasize that these data represent only the Brazilian portion of the basin, as the Negro River basin extends into other South American countries. In this figure, each type of geological formation in both regions is distinctly highlighted using different colors, providing a clear visual distinction between the various formations in each area of the basin. The objective is to analyze the importance of these geological formations in the distinct hydrological behaviors observed at these stations.

The region A, shown in Fig. 5, which includes the Cucuí station within the Negro River basin, exhibits a diverse range of geological formations. The Içá Formation, the largest in extent, covers approximately 41.7% (see Table 2) of the area and is predominant in the region, likely influencing the hydrological characteristics of the basin. Sandy deposits in basement areas, representing 17% of the area, and the Tarsira and Santa Izabel do Rio Negro lithofacies, covering 12.1% and 9.4%, respectively, also contribute to the geological pattern of region A. These geological formations may have a direct impact on the hydrological behavior observed at the Cucuí station, which exhibits a more unstable dynamic, reflecting the interaction between the rocks and local water systems. Formations like Içá, with a large presence of sediments and more permeable rocks, may facilitate water infiltration, while lithofacies and alluvial deposits may influence the slow storage and release of water, contributing to the variable flow patterns observed. Additionally, other geological units, such as the Rio Içana Intrusive Suite and the Serra da Neblina Formation, with smaller but significant areas, also affect the hydrological behavior of the region, influencing both water quality and quantity in the basin. The diversity of these formations suggests a complex interaction between geology and hydrological phenomena in the Negro River basin, as reflected in the variation of hydrological conditions at the Cucuí station.

The hydrological behavior at the Moura station, located in region B of the Negro River basin, is notably more stable when compared to the Cucuí station in region A. This difference can be partially attributed to the dominant geological formations

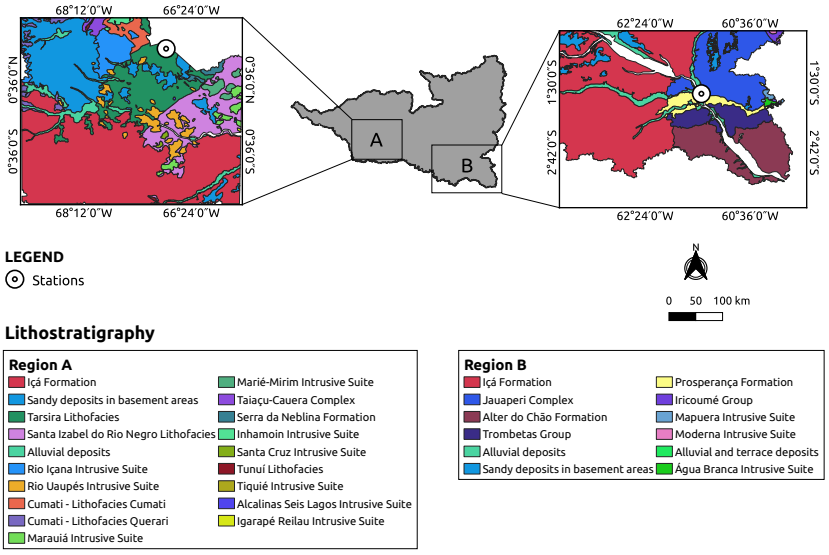


Figure 5. Lithostratigraphic units in the highlighted regions A and B of the Negro River basin, encompassing various formations, intrusive suites, lithofacies, and sedimentary deposits.

Table 2. Lithostratigraphic units with their respective areas and percentage representation in the region A.

Units	Area (km^2)	Percentage (%)
Içá Formation	53,377.464	41.651
Sandy deposits in basement areas	21,843.461	17.045
Tarsira Lithofacies	15466.021	12.068
Santa Izabel do Rio Negro Lithofacies	12,108.700	9.449
Alluvial deposits	4,826.997	3.767
Rio Içana Intrusive Suite	4,685.492	3.656
Rio Uaupés Intrusive Suite	4,311.253	3.364
Cumati-Lithofacies Cumati	2,540.432	1.982
Cumati-Lithofacies Querari	2,366.789	1.847
Marauiá Intrusive Suite	1,599.492	1.248
Marié-Mirim Intrusive Suite	1,446.445	1.129
Taiaçu-Cauera Complex	1,377.960	1.075
Serra da Neblina Formation	856.556	0.668
Inhamoin Intrusive Suite	836.346	0.653
Santa Cruz Intrusive Suite	240.456	0.188
Tunuí, Litofáceis	161.193	0.126
Tiquié Intrusive Suite	56.155	0.044
Alcalinas Seis Lagos Intrusive Suite	21.550	0.017
Igarapé Reilau Intrusive Suite	31.618	0.025

in the region. The Içá Formation, which covers almost 48.4% of the area, as shown in Table 3, plays a significant role in the hydrological stability by contributing to water infiltration and storage. The Jauaperi Complex, occupying 15.6% of the area, and the Alter do Chão Formation, covering 14.7%, both consist of more compact and less permeable rocks, potentially leading to slower water movement and reducing fluctuations in the hydrological regime. The Trombetas Group, representing 6.2% of the region, could further stabilize water flows, as it is likely to influence the retention and gradual release of water. Additionally, the Alluvial deposits, which cover 5.7% of the area, may contribute to slow water release, further supporting a more consistent

hydrological response. In contrast to region A, these geological formations in region B contribute to the more stable hydrological dynamics observed at the Moura station, helping to reduce the variability in water levels and flow rates. For a detailed analysis of the geology of the Negro River basin, the references should be consulted (Mendes et al. 2021; Maia 2010; Holanda, Marmos, and Maia 2014).

Table 3. Lithostratigraphic units with their respective areas and percentage representation in region B.

Units	Area (km^2)	Percentage (%)
Içá Formation	46,588.314	48.350
Jauaperi Complex	15,022.739	15.591
Alter do Chão Formation	14,190.283	14.727
Trombetas Group	5,993.468	6.220
Alluvial deposits	5,492.448	5.700
Sandy deposits in basement areas	4,499.729	4.670
Prosperança Formation	3,305.132	3.430
Iricoumé Group	517.279	0.537
Mapuera Intrusive Suite	320.499	0.333
Água Branca Intrusive Suite	225.413	0.234
Moderna Intrusive Suite	159.177	0.165
Alluvial and terrace deposits	42.372	0.044

Guyot et al. (1993) states that the precipitation patterns in the Negro River basin are divided as follows: in the region extending up to Serrinha, rainfall varies between 2,000 and 3,600 mm/year, with 30% of the precipitation occurring from May to July. In the region covering the Branco River basin up to Caracaraí, precipitation ranges from 1,400 to 2,300 mm/year. Finally, in the region spanning downstream of the Negro and Branco River basins up to the city of Manaus, 40% of the annual precipitation occurs between March and May. The following Figs. 6 and 7 illustrate these observations for the Moura and Cucuí stations.

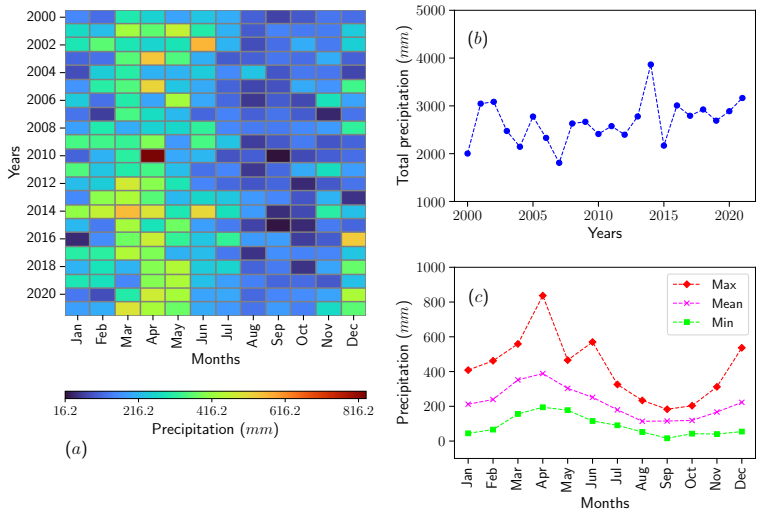


Figure 6. (a) Heatmap showing the precipitation by month in Moura station between the years 2000 and 2021, depicted using a colour palette. (b) Total precipitation in Moura station over the years. (c) Maximum, mean, and minimum precipitation values by month.

The precipitation data are provided by (ANA 2024), with codes being 161002 (Moura

station) and 8166000 (Cucuí station). Figure 6 shows the precipitation pattern at the Moura station between the years 2000 and 2021. Figure 6(a) presents the heatmap of precipitation intensities throughout the months of the year. The color palette indicates different intensity levels. It is observed that from January to February, moderate precipitation occurs. Between March and April, the highest rainfall intensities are recorded, with the peak observed in April 2010, reaching 837.1 *mm*. From June to July, rainfall begins to decrease, and between August and October, the lowest precipitation levels are observed at this station. The lowest recorded value was also in 2010, with only 16.2 *mm* in September. From November to December, rainfall begins to increase again.

Figure 6(b) shows the total annual precipitation from 2000 to 2021, represented by the dashed blue curve with circles. A significant fluctuation in annual precipitation is observed, ranging between 2,000 and 3,000 *mm*/year. The maximum value was recorded in 2014, reaching 3,865.3 *mm*, followed by 2021, with a total of 3,164.7 *mm*. Figure 6(c) presents a comparison of maximum, mean, and minimum recorded precipitation values, represented by the dashed red curve with diamonds, the dashed magenta curve with x symbols, and the dashed lime curve with squares, respectively. A well-defined precipitation pattern at the Moura station is evident. Between March and May, the highest precipitation levels occur, with the maximum recorded in April 2010. Conversely, between August and October, the lowest values are observed. This confirms that the Moura station exhibits a distinct precipitation pattern, which clearly contributes to the observed stable hydrological behavior.

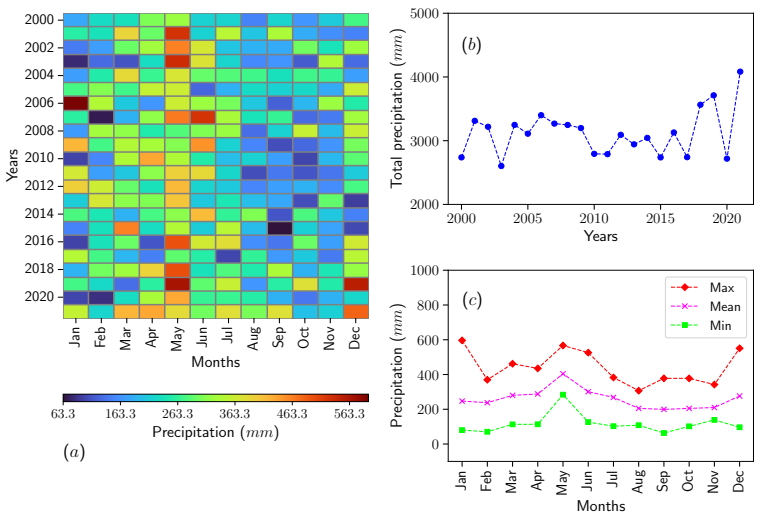


Figure 7. (a) Heatmap showing the precipitation by month in Cucuí station between the years 2000 and 2021, depicted using a colour palette. (b) Total precipitation in Cucuí station over the years. (c) Maximum, mean, and minimum precipitation values by month.

Figure 7 presents the observed precipitation at the Cucuí station. Figure 7(a) displays the heatmap of monthly precipitation between the years 2000 and 2021. Unlike Moura, Cucuí exhibits significant variations in precipitation intensity throughout the months, as indicated by the different colors. The highest precipitation values occur between January and July, with the maximum recorded in January 2006 at 596.3 *mm*.

May stands out as the month with the highest precipitation levels. Between August and November, the lowest values are observed, with the minimum recorded in September 2015 at 63.3 mm. However, December also shows high rainfall levels.

Figure 7(b) illustrates the total annual precipitation, revealing fluctuations between 2,500 and 3,500 mm/year. The highest observed value in the period was in 2021, reaching 4,083.6 mm. As shown in Fig. 3(c), this year was selected for flood forecasting using ANNs. Finally, the Fig.7(c) presents the maximum, mean, and minimum precipitation values. Unlike the Moura station, the seasonality of rainfall in Cucuí is less evident. However, the peak rainy period is observed between April and June, while the rest of the months shows less pronounced variations. This behavior clearly contributes to the observed variation in hydrological instability at Cucuí.

Other factors, including anthropogenic influences such as land use changes, deforestation, urbanization, and agriculture, can significantly impact the hydrological behavior of the Negro River basin. These human activities, alongside natural elements like climate variability, further modify flow patterns, water storage, and flood dynamics. However, these factors will be integrated into future research to enable a more precise and holistic understanding of the basin's hydrological processes.

3. Methods

This section offers a comprehensive description of the neural network architectures utilized in this study for predictive data analysis.

3.1. *Artificial Neural Networks*

Artificial Neural Networks (ANNs) are computational systems inspired by the human brain, designed to enable machines to detect patterns and address problems that are often complex for humans to analyze. While these networks are biological inspired, they do not conclusively replicate actual brain processes. Instead, they function as mathematical frameworks capable of learning from data (Chollet 2019; Géron 2022). The network begins with an input layer that processes historical data and ends with an output layer that generates the final predictions. The intermediate layers, known as hidden layers, play a critical role in identifying and extracting features from the input data, facilitating the model's ability to understand intricate patterns.

This study implements three types of ANNs for time series forecasting: Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and one-dimensional Convolutional Neural Networks(1D-CNN). Developed using TensorFlow and Keras in Python (Chollet 2019; Géron 2022), these libraries provided efficient tools for designing, training, and evaluating complex models. To validate the ANN architectures, *k-fold cross-validation* (Kohavi 1995; Silva, Spatti, and Flauzino 2020) was employed, dividing the dataset into k subsets, with $(k-1)$ used for training and the remaining subset for testing. This approach ensured robust evaluation and selection of the best-performing topology. The ANNs were trained using supervised learning to predict extreme flood and drought events, with carefully selected historical data as labels. Future research may incorporate unsupervised learning to uncover hidden patterns and improve model accuracy further.

Distinct ANN models were developed for the Moura and Cucuí stations, each tailored to their unique hydrological dynamics and designed to learn station-specific characteristics (see Fig. 2). Moura's water level (H) of the Negro River exhibits a consistent

pattern with a clear trend, which simplifies the network's learning process. As a result, simpler ANN models, trained on four years of data preceding extreme events, were sufficient to accurately forecast floods and droughts. In contrast, Cucuí's dynamics are far more complex, with historical data lacking consistent patterns over time. This complexity required advanced hyperparameter tuning and the inclusion of eight years of data for training, allowing the model to effectively capture the station's intricate behavior. To address the distinct hydrological behaviors of each station, the networks were customized with specific topologies and fine-tuned hyperparameters to suit their respective characteristics. For the ANNs validation, the same amount of data was utilized during training.

To achieve optimal performance with the employed networks, the amount of historical data can vary depending on the complexity of the problem, rather than being strictly limited to four or eight years as previously mentioned. The specific adjustments required for hyperparameters depend on the problem's intricacy, with larger volumes of training data often improving the learning capabilities of ANNs. This study focuses on the architectural features and unique characteristics of the networks while leaving out detailed mathematical explanations, as they are beyond the scope of this article. For those interested in exploring the mathematical foundations and operations of ANNs, it is suggested to consult the reference (Géron 2022).

3.1.1. Recurrent neural network

Recurrent Neural Networks (RNNs) are a class of ANNs specifically designed for sequential data processing, with the ability to retain information from previous inputs through directed cycles in their architecture. This capability to capture temporal dependencies makes RNNs particularly suitable for tasks such as time series forecasting, including hydrological data prediction. Figure 8 presents the schematic diagrams considered for the two stations, while Table 4 provides a complete description of the hyperparameters.

- **Moura**: The network was trained with four years of data. The first RNN layer comprises 50 units with the Tanh activation function (see Fig. 8(a)). The second and third RNN layers have 100 units each, also using Tanh activation, with a dropout rate of 0.1 applied in the second layer to reduce overfitting. The final output layer provides predictions of the water level (H) of the Negro River.
- **Cucuí**: Given the higher complexity of the data, eight years of training data were used. The RNN architecture includes three layers with 60, 120, and 120 units, respectively, and a dropout rate of 0.1 applied in the third layer (see Fig. 8(b)). ReLU activation is utilized across all layers, as it showed better performance than Tanh in k-fold cross-validation, effectively capturing nonlinear relationships. After the RNN layers, the data passes through a fully connected layer with 120 units using ReLU, followed by another fully connected layer with 60 units, concluding with the output layer that generates the predictions.

3.1.2. Long short-term memory

The Long Short-Term Memory (LSTM) network is an advanced version of the RNN, designed to overcome the challenges of capturing long-range dependencies. Its specialized memory cells and gating mechanisms allow it to effectively process extended sequences, making it ideal for tasks like predicting extreme hydrological events (Chollet

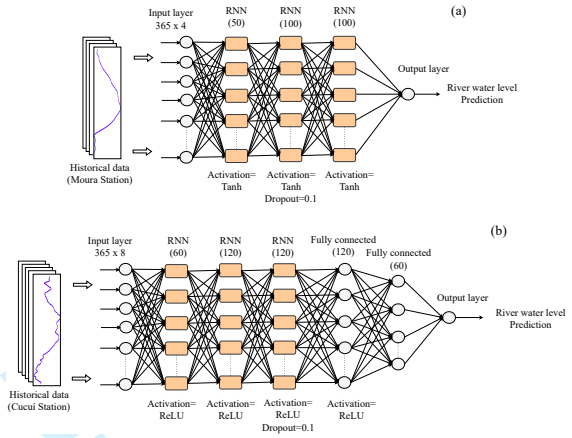


Figure 8. (a) Schematic diagram of the RNN architecture employed for the Moura station. The input data consists of four years of data before the extreme floods and droughts for training. (b) Diagram of the RNN for Cucuí station. The input data consists of eight years of data.

Table 4. Hyperparameters of the proposed RNN models.

Layer	Output	Dropout	Moura	Cucuí
Input	365 × 4	-	✔	-
	365 × 8	-	-	✔
RNN 1 + Tanh	50	-	✔	-
RNN 1 + ReLU	60	-	-	✔
RNN 2 + Tanh	100	-	✔	-
Dropout	100	0.1	✔	-
RNN 2 + ReLU	120	-	-	✔
RNN 3 + Tanh	100	-	✔	-
RNN 3 + ReLU	120	-	-	✔
Dropout	120	0.1	-	✔
Fully connected + ReLU	120	-	-	✔
Fully connected	60	-	-	✔
Output	1	-	✔	✔

2019; Géron 2022). LSTM's architecture helps preserve relevant information, which is key for tasks with significant temporal dependencies. This sub-subsection describes the LSTM models used in the study, detailing the two topologies for predictions. In Fig. 9, the diagrams of the two models used are shown, and in Table 5, the hyperparameters are listed.

- **Moura station**: The first LSTM layer consists of 50 units with Tanh and Sigmoid activation functions, while the second layer has 100 units and includes a 0.1 dropout rate. The third layer, with 100 units, uses only the Sigmoid activation function before the output layer.
- **Cucuí**: The network includes three LSTM layers with 60, 120, and 120 units, and a 0.1 dropout rate in the third layer. Following these layers, there are two fully connected layers with 120 and 60 units, respectively, before the output.

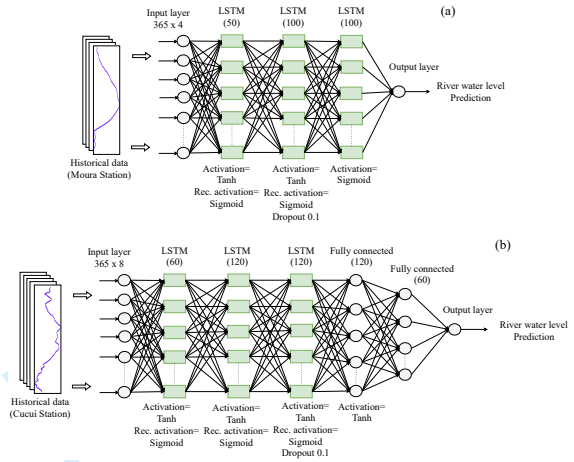


Figure 9. (a) Schematic diagram of the LSTM architecture employed for the Moura station. The input data consists of four years of data before the extreme floods and droughts. (b) Diagram of the LSTM used in Cucuí station. The input data consists of eight years of data.

Table 5. Hyperparameters of the proposed LSTM models.

Layer	Output	Dropout	Moura	Cucuí
Input	365 × 4	-	✔	-
	365 × 8	-	-	✔
LSTM 1 + Tanh + Sigmoid	50	-	✔	-
	60	-	-	✔
LSTM 2 + Tanh + Sigmoid	100	-	✔	-
	120	-	-	✔
Dropout	100	0.1	✔	-
LSTM 3 + Sigmoid	100	-	✔	-
LSTM 3 + Tanh + Sigmoid	120	-	-	✔
Dropout	120	0.1	-	✔
Fully connected + Tanh	120	-	-	✔
Fully connected	60	-	-	✔
Output	1	-	✔	✔

3.1.3. Convolutional neural network

The 1D Convolutional Neural Network (1D-CNN) is pivotal in advancing time series forecasting, particularly in understanding the complexities of hydrological phenomena. By capturing temporal patterns and correlations within the data, 1D-CNNs offer a robust solution, especially for addressing the variability of hydrological events. In Fig. 10, the diagrams are presented, and in Table 6, the hyperparameters are listed.

- **Moura**: The network consists of three convolutional layers, each using two kernels, with a dropout of 0.1 applied in the second layer. A kernel size of two indicates that the convolution windows are set to two time units. The first convolutional layer has 50 filters, while the subsequent two layers have 100 filters each. After these layers, a flattening layer with 36,500 elements is applied, followed by an output layer with one processing unit. ReLU activation functions are employed throughout the architecture.

- **Cucuí**: The architecture includes three convolutional layers, each using two kernels, and two pooling layers with a two stride configuration. The first convolutional layer uses 60 filters, while the following two layers use 120 filters each. After the convolutional layers, a flattening layer with 10,920 elements, followed by a dropout rate of 0.5. The network also includes a fully connected layer with 120 units, another fully connected layer with 60 units, and an output layer with one unit. ReLU activation functions are applied.

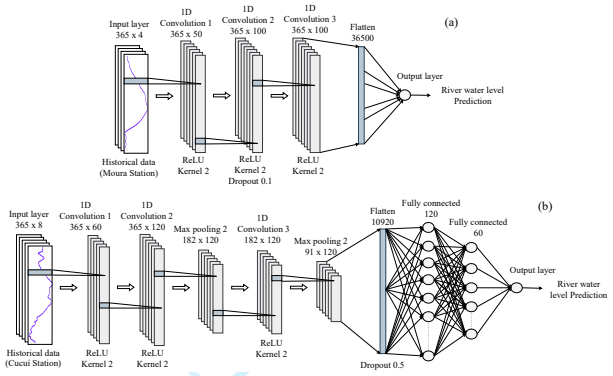


Figure 10. (a) Schematic diagram of the 1D-CNN architecture employed for the Moura station. The input data consists of four years of data before the extreme floods and droughts. (b) Diagram of the 1D-CNN used for the Cucuí station. The input data consists of eight years of data.

Table 6. Hyperparameters of the proposed 1D-CNN models.

Layer	Output	Kernel	Stride	Dropout	Moura	Cucuí
Input	365 × 4	-	-	-	✔	-
	365 × 8	-	-	-	-	✔
1D-Convolution 1 + ReLU	365 × 50	2	-	-	✔	-
	365 × 60	2	-	-	-	✔
1D-Convolution 2 + ReLU	365 × 100	2	-	-	✔	-
	365 × 120	2	-	-	-	✔
Dropout	365 × 100	-	-	0.1	✔	-
Max pooling	182 × 120	-	2	-	-	✔
1D-Convolution 3 + ReLU	365 × 100	2	-	-	✔	-
	182 × 120	2	-	-	-	✔
Max pooling	91 × 120	-	2	-	-	✔
Flatten	36500	-	-	-	✔	-
	10920	-	-	-	-	✔
Dropout	10920	-	-	0.5	-	✔
Fully connected	120	-	-	-	-	✔
Fully connected	60	-	-	-	-	✔
Output	1	-	-	-	✔	✔

3.2. *Evaluation of ANN Performance*

The techniques presented in this subsection are the statistical models used to evaluate the performance of the ANNs in making predictions.

3.2.1. Distance correlation

Distance Correlation (DC) is a statistical measure used to assess nonlinear relationships between variables by evaluating the association between the distances of paired points in each variable. Unlike traditional correlation metrics, DC captures nonlinear dependencies and is robust to outliers, making it particularly useful in time series analysis. DC can be applied in multidimensional spaces, $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, allowing for the assessment of dependencies in high-dimensional datasets. In this context, DC is used to assess the relationship between the observed (historical) data and the predictions made by the ANNs.

The procedure follows the methods described in previous studies (Székely, Rizzo, and Bakirov 2007; Mendes and Beims 2018; Mendes, da Silva, and Beims 2019; Mendes et al. 2022). For a data set $(X, Y) = \{(x_k, y_k) : k = 1, ..., n\}$, where $n > 2$, x_k denotes the observed data, and y_k represents the predicted values. The matrix R_{ij} (constructed from real data) for sample X is defined as:

$$R_{ij} = r_{ij} - \bar{r}_{i.} - \bar{r}_{.j} + \bar{r}_{..},$$

where $r_{ij} = |x_i - x_j|_p$ is the Euclidean distance between the elements, $\bar{r}_{i.} = \frac{1}{n}\sum_{j=1}^{n} r_{ij}$ is the mean of the rows, $\bar{r}_{.j} = \frac{1}{n}\sum_{i=1}^{n} r_{ij}$ is the mean of the columns, and $\bar{r}_{..} = \frac{1}{n^2}\sum_{i,j=1}^{n} r_{ij}$ is the general mean. Similarly, the matrix P_{ij} (constructed from predicted data) is computed for sample Y.

The empirical distance covariance for the joint sample (X, Y) is given by:

$$\sigma(X,Y) = \frac{1}{n}\left(\sum_{i,j=1}^{n} R_{ij}P_{ij}\right)^{1/2},$$

and the distance variances for the samples X and Y are:

$$\sigma(X) = \frac{1}{n}\left(\sum_{i,j=1}^{n} R_{ij}^2\right)^{1/2} \quad \text{and} \quad \sigma(Y) = \frac{1}{n}\left(\sum_{i,j=1}^{n} P_{ij}^2\right)^{1/2}.$$

Finally, the DC coefficient is calculated as:

$$\mathrm{DC}(X,Y) = \frac{\sigma(X,Y)}{\sqrt{\sigma(X)}\sqrt{\sigma(Y)}}.$$

3.2.2. Coefficient of determination

The coefficient of determination, or R-squared (R^2), quantifies the proportion of variance in the predicted values that can be explained by the variance in the observed values (Chicco, Warrens, and Jurman 2021). This measure can be calculated using the

following equation:

$$R^2 = 1 - \frac{\sum_{k=1}^{n} (x_k - y_k)^2}{\sum_{k=1}^{n} (x_k - \bar{x})^2}, \quad (1)$$

where n is the number of data points, x_k is the observed data, $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ is the arithmetic mean of the observed values, and y_k represents the predicted values. The value of R^2 ranges from $-\infty$ to 1. A value close to 1 indicates a good fit, while a more negative value signifies a poor fit, suggesting that the variances in the predicted values are less linearly related to the observed values.

4. Results

4.1. *Prediction of extreme floods at Moura station*

Figures 11(a)-(c) and (g)-(i) display the river Negro's water level (H) as a function of the months during the flood periods of 1989, 2009, 2012, 2014, 2019 and 2021, respectively. We call to attention again that all Moura flood periods match those of Manaus, making Moura's predictions of extreme floods of most relevance. The blue curves are the observed data at the Moura station, and all present a maximum between June and August of each year. Furthermore, smaller values of H occur over the years between January to February and from October to December. The orange curve shows predictions during the flood generated by the RNN, the lime curve represents predictions from the LSTM, and the magenta curve represents predictions from the 1D-CNN. Here, data from the 4 years preceding the actual year were used for the training. Overall, besides the cases between January to February and October to December of each year, there is a good visual agreement between the predictions and the real data. In Figs. 11(a) and (h), the maxima were predicted to occur earlier than the real data, while in Fig. 11(c) it was predicted to occur after the real maxima.

The performance of the predictions can be measured by the DC between observed and predicted values over the months of the corresponding years, as presented in Figs. 11 (d)-(f) and (j)-(l). Red curves for the DC between data and the predictions generated by the RNN, black curves for the DC between data and the LSTM predictions, and dashed cyan curves for the DC between data and the 1D-CNN predictions. For higher correlations (or anti-correlations), DC approaches 1, and it tends to 0 for vanishing correlations. While Figs. 11(f) and (l) present DC close to one for the whole year, in Figs. 11(d), (e), (j) and (k), the DC decays for the February. In Fig. 11(e), the DC is smaller between March and June and from September to December. In fact, besides Figs. 11(k) and (l), all other cases present a decrease of DC (sometimes small) around August or September.

To quantify the predictions we analyze next the margin of errors for the years 2012 and 2014, for which the predictions of the flood maxima and the corresponding month are not precise, as seen in Figs. 11(c) and 11(g). The margin errors for the years 2012 and 2014 are displayed in Fig. 12 for the distinct prediction methods. The blue curves repeat the data of the flood from Figs. 11(c) and 11(g). In all cases, it is possible to observe that close to the maxima, the margin errors increase, but lie below 5%, which are acceptable forecasting values.

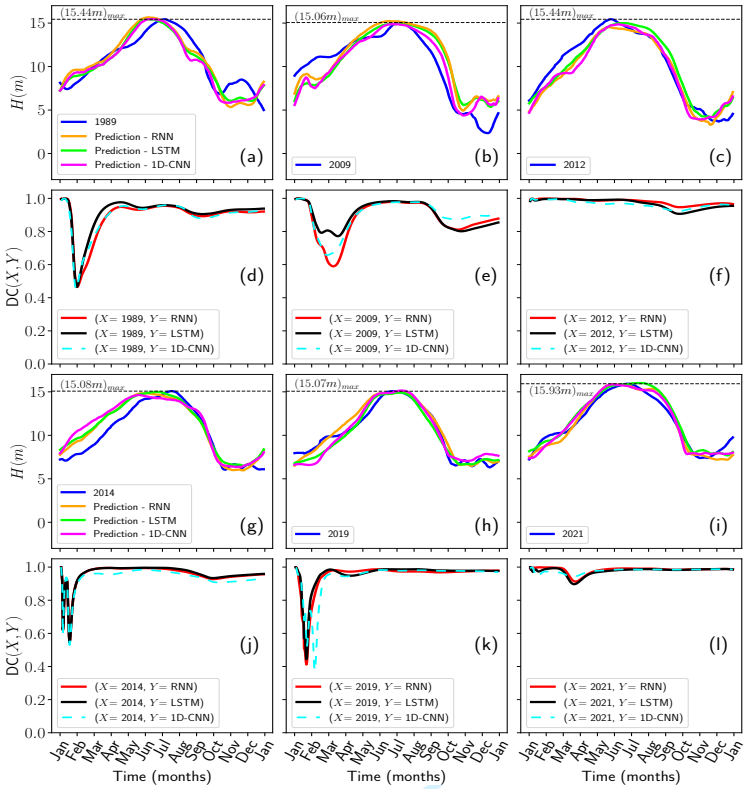


Figure 11. (a)-(c) and (g)-(i): Water level (H) at Moura station throughout the floods of 1989, 2009, 2012, 2014, 2019, and 2021, respectively. The continuos blue curve represents the observed data. The continuous orange, lime and magenta curves are predictions by the RNN, LSTM, and 1D-CNN. (d)-(f) and (j)-(l): DC between observed and predicted values.

Finally, Fig. 13 show the scatter plots between observed and predicted values for the (a) 2012 and (b) 2021 floods for all prediction methods. Furthermore, the corresponding coefficient of determination (R^2) is provided. The years 2012 and 2021 present the largest extrema in Moura and Manaus stations (see Fig. 3). While the dispersion in Fig. 13(b) is small and R^2 very close to one, in Fig. 13(a) the dispersion in larger and R^2 smaller, but still larger than 0.9. Such small dispersion and larger values of R^2 are a consequence of DC values close to one for all months in the years 2012 and 2021, as seen in Figs. 11(f) and (l).

4.2. *Prediction of extreme drought at Moura station*

Another relevant analysis concerns the prediction of extreme droughts. Figure 14(a) displays (blue curve) the water level (H) of the River Negro throughout the months during the 2009 drought. We observe minima in December. As depicted in Fig. 14(a), the other curves represent predictions generated by RNN, LSTM, and 1D-CNN. All three methods can predict the approximate minimum between December 2009 and

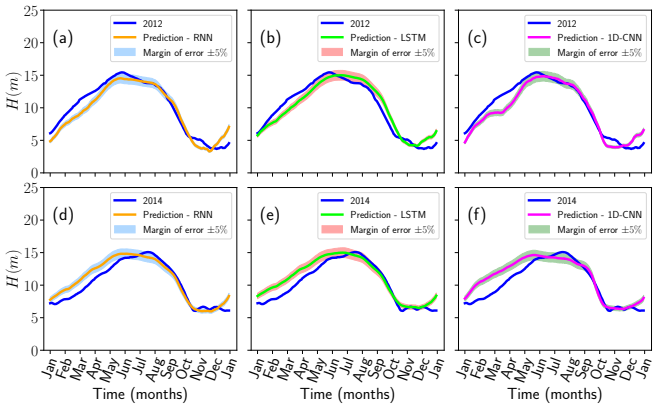


Figure 12. (a)-(c): Water level (H) at Moura station throughout the 2012 flood. The solid blue curve represents the observed data, while the continuous orange, lime, and magenta curves represent predictions by the ANN. The light blue, light red, and light green shaded regions indicate a margin of error of ±5% around the predicted values. (d)-(f): The same analysis is presented, but for the 2014 flood event.

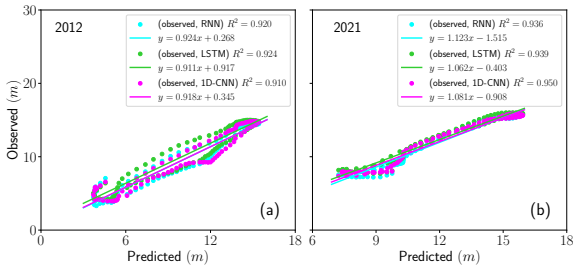


Figure 13. Scatter plots between observed and predicted values for the (a) 2012 and (b) 2021 floods. The linear fits between observed and predicted data are highlighted. The respective coefficients of determination (R^2) are presented.

January 2010, albeit with a slight delay in the days. Figure 14(b) illustrates the corresponding DC values. Apart from some oscillations around July and August 2009, the DC remains close to one for all other months, exhibiting a slight decrease between March and July. This demonstrates a strong correlation between predictions and real data. In Fig. 14(c), the scatter plot is presented. The R^2 values for the LSTM and 1D-CNN with the observed data are around 0.8. However, for the RNN with the observed data, the R^2 value exceeds 0.9.

4.3. *Prediction of extreme flood at Cucuí station*

We begin by outlining our analysis of the predictions generated by the neural networks trained on data from the Cucuí station. Similarly to Moura station, as depicted in Fig. 15(a), the solid blue curve illustrates the observed data during the extreme flood event of 2021, while the orange, lime, and magenta curves represent the predictions generated by the RNN, LSTM, and 1D-CNN models, respectively. It is noteworthy that the dynamic behavior of the water levels of the Negro River throughout the year

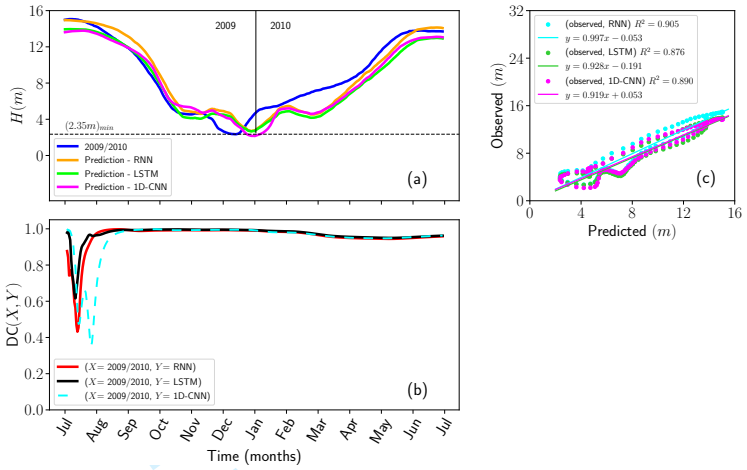


Figure 14. (a) Water level (H) at Moura station during the drought of 2009. The solid blue curve represents the observed data and the continuous orange, lime and magenta curves are the predictions by the RNN, LSTM, and 1D-CNN, respectively. (b) DC between observed and predicted values. (c) Scatter plot between observed and predicted values. The linear fits and the R^2 are presented.

exhibits significant variability, indicative of a complexity surpassing that observed in Moura. Consequently, we used eight years of data preceding 2021 to train the networks. Subsequently, our analysis reveals that the prediction curves exhibit some deviation from the observed data between January and late April. However, from May to mid-August, the predictions closely align with the actual measurements, demonstrating a notable correspondence with the peak of the extreme flood event during that year, as delineated by the dashed black line. Between August and November, the predictions display consistency among themselves but tend to exceed the observed measurements slightly. Finally, the prediction curves converge to the actual values from November to late December. Additionally, it is worth noting that the variations between the predictions and observed data during specific periods may be attributed to the inherent complexities of the hydrological dynamics in the Cucuí station. Despite these challenges, our models demonstrate promising performance in capturing the overall trends and significant fluctuations in water levels. Further refinements and fine-tuning of the models could potentially enhance their accuracy and reliability, paving the way for more effective flood prediction and mitigation strategies in the region.

In Fig. 15(b), the DC curves between the predictions and real values are displayed. The continuous red and black curves, along with the dashed cyan curve, represent the DC coefficient between the actual and predicted data by the ANN. It is evident that from January to the end of May, there is an oscillation of the DC around values of approximately 0.2 and 0.9. However, starting from June, the DC coefficient approaches 1.0 and remains constant until the end of December. Therefore, it is apparent that the predictions of the ANN exhibit a strong correlation with the observed data during the period of extreme flood that occurred in June. Additionally, they also demonstrate a good correlation during the onset of the ebbing period.

Figure 15(c) presents a scatter plot depicting the observed data at the Cucuí station and the data predicted. While the cyan points represent the real data with predictions by the RNN, the green points represent the real data with predictions by the

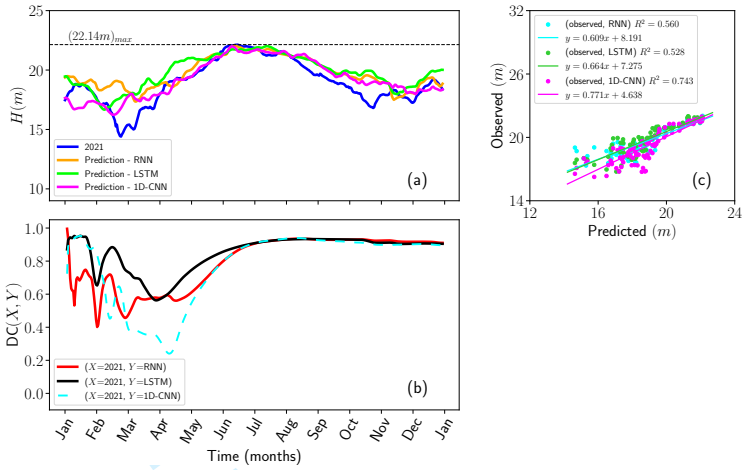


Figure 15. (a) Water level (H) at Cucuí station during the flood of 2021. The solid blue curve represents the observed data and the continuous orange, lime, and magenta curves are the predictions generated by the RNN, LSTM, and 1D-CNN, respectively. (b) DC between observed and predicted values. (c) Scatter plot between observed and predicted values. The linear fits and the R^2 are presented.

LSTM, and the magenta points the observed data with predictions by the 1D-CNN. Similarly, the same colors are used for the corresponding continuous lines representing the linear fits. It is noteworthy that between 14 and 18m, the points are slightly dispersed, whereas, from approximately 19m onwards, the points are well aligned along the continuous lines of the fits. For the RNN and LSTM with the observed data, the R^2 values are around 0.5. However, for the 1D-CNN with the observed data, the R^2 value exceeds 0.7. These results suggest that the 1D-CNN model exhibits a higher level of accuracy in predicting water levels at the Cucuí station compared to the RNN and LSTM models.

4.4. *Prediction of extreme drought at Cucuí station*

Utilizing eight years of historical data preceding the 2009 drought, ANNs were employed to forecast this event. It's noteworthy that for the drought periods, the data was arranged such that the ANNs were trained solely on droughts, meaning the data was organized to span from July of one year to June of the following year, completing a 12-month cycle. The same configuration was used for drought prediction at Moura. In Fig. 16(a), the years 2009/2010 are illustrated. Between July and December 2009, the ANNs predictions fluctuate around the observed data. From January to March 2010, the curves closely match the real data, indicating the extreme drought of that year. From April to June, the predictions slightly dip below the observed values. In Fig. 16(b), oscillations in the DC are noticeable between July and October, reaching values between 0.3 and 1.0 (maximum correlation). The fluctuations in the DC highlight the subtle shifts in correlation between predicted and observed data over various temporal intervals. In Fig. 16(c), the scatter plot between observed and predicted data is presented. In all cases, the R^2 exceeds 0.8.

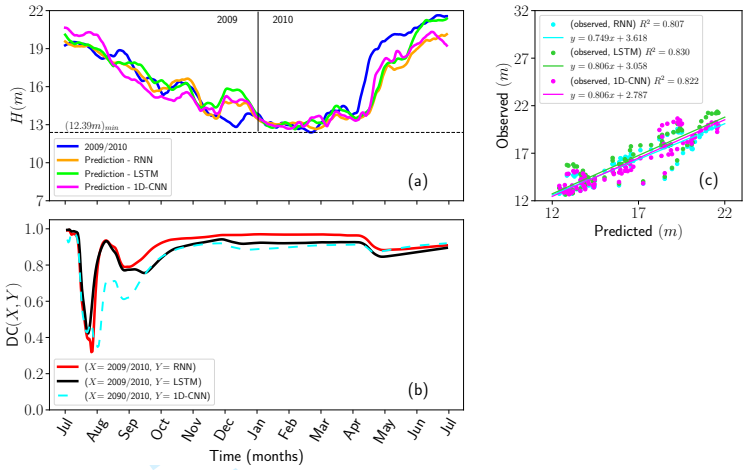


Figure 16. (a) Water level (H) at Cucuí station during the drought of 2010. The solid blue curve represents the observed data and the continuous orange, lime and magenta curves are the predictions generated by the RNN, LSTM, and 1D-CNN, respectively. (b) DC between observed and predicted values. (c) Scatter plot between observed and predicted values. The linear fits and the R^2 are presented.

5. Conclusions

The ability to predict extreme floods and droughts in the Negro River basin holds paramount importance. Not only does this forecasting capability enable proactive measures to mitigate the adverse impacts of these events, but it also contributes to a deeper understanding of the hydrological dynamics in the region. The significance of the Negro River basin extends beyond the Amazon, playing a crucial role in the global ecosystem and socio-economic landscape. Geological formations and precipitation play a fundamental role in the hydrological behavior of the basin. By employing Artificial Neural Networks (ANN) for prediction purposes, particularly leveraging historical data from stations like Moura and Cucuí, we demonstrate the efficacy of these models in capturing the complexities inherent in hydrological measurements. Despite the challenges posed by the intricate nature of these dynamics, the ANN models showcased commendable performance in forecasting extreme floods and droughts. Deep learning models can serve as valuable tools for competent authorities to make informed decisions well in advance, thus mitigating the impacts of these events on the Amazon region. Obviously, artificial intelligence is still directed towards solving specific problems. This means that, the network models need to be adapted to meet the particularities of the specific situation.

Specifically, we observe that the adjustment of each ANN model is influenced by the inherent complexity of each location. For the predictions curves at the Moura station using RNN, LSTM, and 1D-CNN, considering the six extreme floods of 1989 (15.44m), 2009 (15.06m), 2012 (15.44m), 2014 (15.08m), 2019 (15.07m), and 2021 (15.93m), we noticed that all of them approached or even surpassed the values of the maximum peaks observed each year. This is corroborated by the distance correlation (DC) values close to 1.0 throughout most of the period. The same analysis extends to the extreme drought at Moura station in 2009 (2.35m). At the Cucuí station, despite exhibiting a more complex behavior, the ANN predictions closely approximate the peak of the

extreme flood observed in 2021 (22.14m). The same applies to the extreme drought analyzed in 2010 (12.39m).

In addition, the utilization of ANN for predicting extreme floods and droughts implies a significant advance in the field of water resources management. The ability to accurately forecast floods and droughts not only enhances our preparedness but also empowers decision-makers with valuable insights to effective risk mitigation strategies. Moreover, the success of the ANN models, trained on historical data, underscores the importance of leveraging observational data for improving predictive capabilities in complex hydrological systems. As we continue to refine and optimize these models, incorporating advances of deep learning techniques, we move closer to achieving more robust and reliable predictions, ultimately contributing to sustainable water management practices. In future works, the integration of additional hydrological data, such as discharge, sediment transport, water quality parameters, ADCP (Acoustic Doppler Current Profiler) measurements, and remote sensing with satellite-derived information, will be explored. These datasets will provide a more comprehensive understanding of the hydrological system and enhance the ANN models' predictive performance. Incorporating these diverse data sources will allow for improved modeling of complex interactions within water systems, paving the way for even more accurate and reliable forecasting solutions.

Acknowledgements

C.L.E. thanks FAPEAM for financial support, N.P.F. thanks CAPES for supporting the Rio Negro TAR project, C.F.O.M., M.G., and M.W.B. thank the National Council for Scientific and Technological Development – CNPq (Brazilian agency) for financial support (Grant Numbers 311428/2022-3 and 310294/2022-3, respectively).

Disclosure statement

No potential conflict of interest was reported by the authors.

Availability of data

Data will be made available on reasonable request.

References

Abhishek, K., M. P. Singh, S. Ghosh, and A. Anand. 2012. "Weather Forecasting Model using Artificial Neural Network." *Procedia Technology* 4: 311–318. https://doi.org/https://doi.org/10.1016/j.protcy.2012.05.047.

Ahmad, Mumtaz, Md Abdullah Al Mehedi, Munshi Md Shafwat Yazdan, and Raaghul Kumar. 2022. "Development of Machine Learning Flood Model Using Artificial Neural Network (ANN) at Var River." *Liquids* 2: 147–160. https://doi.org/10.3390/liquids2030010.

ANA. 2024. "Agência Nacional de Águas e Saneamento Básico." `https://www.gov.br/ana/pt-br`.

Armitage, D. W., and H. K. Ober. 2010. "A comparison of supervised learning techniques in the classification of bat echolocation calls." *Ecological Informatics* 5: 465–473. https://doi.org/https://doi.org/10.1016/j.ecoinf.2010.08.001.

Brugnago, Eduardo L., Carolini C. Felicio, and Marcus W. Beims. 2022. "Forecasting the duration of three connected wings in a generalized Lorenz model." *International Journal of Bifurcation and Chaos* 32: 2230031. https://doi.org/10.1142/S0218127422300312.

Brugnago, Eduardo L., Jason A. C. Gallas, and Marcus W. Beims. 2020a. "Machine learning, alignment of covariant Lyapunov vectors, and predictability in Rikitake's geomagnetic dynamo model." *Chaos* 30: 083106. https://doi.org/10.1063/5.0009765.

Brugnago, Eduardo L., Jason A. C. Gallas, and Marcus W. Beims. 2020b. "Predicting regime changes and durations in Lorenz's atmospheric convection model." *Chaos* 30: 103109. https://doi.org/10.1063/5.0013253.

Brugnago, Eduardo L., Tony A. Hild, Daniel Weingärtner, and Marcus W. Beims. 2020. "Classification strategies in machine learning techniques predicting regime changes and durations in the Lorenz system." *Chaos* 30: 053101. https://doi.org/10.1063/5.0003892.

Chebii, Sally Jeruto, Micah Mukungu Mukolwe, and Basil Iro Ong'or. 2022. "River flow modelling for flood prediction using artificial neural network in ungauged Perkerra catchment, Baringo County, Kenya." *Water Practice and Technology* 17: 914–929. https://doi.org/10.2166/wpt.2022.034.

Chevuturi, Amulya, Nicholas P. Klingaman, Conrado M. Rudorff, Caio A. S. Coelho, and Jochen Schöngart. 2022. "Forecasting annual maximum water level for the Negro River at Manaus." *Climate Resilience and Sustainability* 1 (1): e18. https://doi.org/10.1002/cli2.18.

Chevuturi, Amulya, Nicholas P. Klingaman, Steven J. Woolnough, Conrado M. Rudorff, Caio A.S. Coelho, and Jochen Schöngart. 2023. "Forecasting annual maximum water level for the Negro River at Manaus using dynamical seasonal predictions." *Climate Services* 30: 100342. https://doi.org/10.1016/j.cliser.2023.100342.

Chicco, D., M. J. Warrens, and G. Jurman. 2021. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation." *PeerJ Comput Sci* 7: e623. https://doi.org/10.7717/peerj-cs.623.

Chollet, François. 2019. *Deep Learning with python.* 1st ed. Manning Publications Co.

Chowdhury, M., A. Alouani, and F. Hossain. 2010. "Comparison of ordinary kriging and artificial neural network for spatial mapping of arsenic contamination of groundwater." *Stochastic Environmental Research and Risk Assessment* 24: 1–7. https://doi.org/10.1007/s00477-008-0296-5.

Cireşan, D. C., A. Giusti, L. M. Gambardella, and J. Schmidhuber. 2013. "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, 411–418. Springer Berlin Heidelberg.

Eleutério, Cíntia L., Naziano P. Filizola, Alderlene P. de Brito, Mircea Galiceanu, and Carlos F. O. Mendes. 2024. "Identifying wildfires with convolutional neural networks and remote sensing: application to Amazon rainforest." *International Journal of Remote Sensing* 0 (0): 1–22. https://doi.org/10.1080/01431161.2024.2425119.

Frappart, Frédéric, Frédérique Seyler, Jean-Michel Martinez, Juan G. León, and Anny Cazenave. 2005. "Floodplain water storage in the Negro River basin estimated from microwave remote sensing of inundation area and water levels." *Remote Sensing of Environment* 99 (4): 387–399. https://doi.org/10.1016/j.rse.2005.08.016.

Garcia, Beatriz Nunes, Renata Libonati, and Ana M. B. Nunes. 2018. "Extreme Drought Events over the Amazon Basin: The Perspective from the Reconstruction of South American Hydroclimate." *Water* 10 (11). https://doi.org/10.3390/w10111594.

GeoSGB. 2024. "GeoSGB - Sistema de Informação Geográfica do Serviço Geológico do Brasil." Accessed: December 11, 2024, `https://geosgb.sgb.gov.br/`.

Géron, Aurélien. 2022. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* 3rd ed. O'Reilly Media.

Getirana, A. C. V. 2009. "Contribuições da altimetria espacial à modelagem hidrológica de grandes bacias na amazônia." PhD diss., Universidade Federal do Rio Janeiro,

Rio de Janeiro, RJ. http://www.coc.ufrj.br/pt/teses-de-doutorado/153-2009/1192-augusto-cesar-vieira-getirana.

Granato-Souza, Daniela, and David W. Stahle. 2023. "Drought and Flood Extremes on the Amazon River and in Northeast Brazil, 1790–1900." *Journal of Climate* 36 (20): 7213 – 7229. https://doi.org/10.1175/JCLI-D-23-0146.1.

Guyot, J.-L., M. Molinier, V. Guimarães, K. J. Cudo, and E. Oliveira. 1993. "Balanço hídrico da bacia do rio Negro." *Anais do X Simpósio Brasileiro de Recursos Hídricos, Gramado, 11p* .

Ha, S., D. Liu, and L. Mu. 2021. "Prediction of Yangtze River streamflow based on deep learning neural network with El Niño–Southern Oscillation." *Scientific Reports* 2: 11738. https://doi.org/10.1038/s41598-021-90964-3.

Han, H.-G., K.-L. Chen, and J.-F. Qiao. 2011. "An efficient self-organizing RBF neural network for water quality prediction." *Neural Networks* 24: 717–725. https://doi.org/https://doi.org/10.1016/j.neunet.2011.04.006.

Holanda, Janólfta Lêda Rocha, José Luiz Marmos, and Maria Adelaide Mansini Maia. 2014. "Geodiversidade do estado de Roraima." *CPRM* 252. https://rigeo.sgb.gov.br/handle/doc/16775.

Karagiorgi, Kasieczka G. Kravitz S. Nachman B., G., and D. Shih. 2022. "Machine learning in the search for new fundamental physics." *Nat. Rev. Phys.* 4: 399–412. https://doi.org/10.1038/s42254-022-00455-1.

Kessler, Calcavecchia F., J., and T. D. Kühne. 2021. "Artificial Neural Networks as Trial Wave Functions for Quantum Monte Carlo." *Adv. Theory Simul.* 4: 2000269. https://doi.org/10.1002/adts.202000269.

Kohavi, Ron. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, San Francisco, CA, USA, 1137–1143. Morgan Kaufmann Publishers Inc.

Kulichenko, M. et al. 2021. "The Rise of Neural Networks for Materials and Chemical Dynamics." *J. Phys. Chem. Lett.* 12: 6227–6243. https://doi.org/10.1021/acs.jpclett.1c01357.

Laraque, Alain, Jean Loup Guyot, and Naziano Filizola. 2009. "Mixing processes in the Amazon River at the confluences of the Negro and Solimões Rivers, Encontro das Águas, Manaus, Brazil." *Hydrological Processes* 23 (22): 3131–3140. https://doi.org/10.1002/hyp.7388.

Latrubesse, Edgardo M., and Elena Franzinelli. 2005. "The late Quaternary evolution of the Negro River, Amazon, Brazil: Implications for island and floodplain formation in large anabranching tropical systems." *Geomorphology* 70: 372–397. https://doi.org/10.1016/j.geomorph.2005.02.014.

Lek, S., and J.-F. Guégan. 2012. *Artificial neural network - Application and Evolution*. Springer Berlin, Heidelberg.

Li, Q., W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen. 2014. "Medical image classification with convolutional neural network." In *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, 844–848.

Lineros, Miriam Lóez, Antonio Madueño Luna, Pedro M. Ferreira, and Antonio E. Ruano. 2021. "Optimized Design of Neural Networks for a River Water Level Prediction System." *Sensors* 21: 6504. https://doi.org/10.3390/s21196504.

Maciel, Jussara Socorro Cury, Luna Gripp Simões Alves, Daniel de Oliveira, Bruno Gabriel Santos Corrêa, and Iraúna Maiconã Rodrigues de Carvalho. 2022. "Analysing the Flood Warning of Negro River in Manaus." *International Journal of Environmental Impacts: Management, Mitigation and Recovery* 5: 51–64. https://doi.org/10.2495/EI-V5-N1-51-64.

Maia, Maria Adelaide Mansini. 2010. "Geodiversidade do estado do Amazonas." *CPRM* 275. https://rigeo.sgb.gov.br/handle/doc/16624.

Marinho, Rogério Ribeiro, Naziano Pantoja Filizola Junior, and Édipo Henrique Cremon. 2020. "Analysis of Suspended Sediment in the Anavilhanas Archipelago, Rio Negro, Amazon Basin." *Water* 12 (4). https://doi.org/10.3390/w12041073.

Marinho, Rogério Ribeiro, Albert Reis Furtado, Vanessa Cristina Dos Santos, André Zumak Avezedo Nascimento, and Naziano Pantoja Filizola Junior. 2022. "Riverbed morphology and hydrodynamics in the confluence of complex mega rivers - A study in the Branco and Negro rivers, Amazon basin." *Journal of South American Earth Sciences* 118: 103969. https://doi.org/10.1016/j.jsames.2022.103969.

Marinho, Rogério Ribeiro, Paulo Rodrigo Zanin, and Naziano Pantoja Filizola Junior. 2022. "The Negro River in the Anavilhanas Archipelago: Streamflow and geomorphology of a complex anabranching system in the Amazon." *Earth Surface Processes and Landforms* 47 (4): 1108–1123. https://doi.org/10.1002/esp.5306.

Martinez, Jean-Michel, and Thuy Le Toan. 2007. "Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data." *Remote Sensing of Environment* 108 (3): 209–223. https://doi.org/10.1016/j.rse.2006.11.012.

Meade, Robert H., José M. Rayol, Sylvio C. Da Conceição, and José R. G. Natividade. 1991. "Backwater effects in the Amazon River basin of Brazil." *Environmental Geology and Water Sciences* 18 (2): 105–114. https://doi.org/10.1007/BF01704664.

Mendes, C. F. O., R. M. da Silva, and M. W. Beims. 2019. "Decay of the distance autocorrelation and Lyapunov exponents." *Phys. Rev. E* 99: 062206. https://doi.org/10.1103/PhysRevE.99.062206.

Mendes, C. F. O., L. Menon, R. M. da Silva, and M. W. Beims. 2022. "Detecting Correlations in Desynchronized Chaotic Chimera States." *International Journal of Bifurcation and Chaos* 32 (05): 2230012. https://doi.org/10.1142/S0218127422300129.

Mendes, Carlos F. Oliveira, and Marcus W. Beims. 2018. "Distance correlation detecting Lyapunov instabilities, noise-induced escape times and mixing." *Physica A: Statistical Mechanics and its Applications* 512: 721–730. https://doi.org/10.1016/j.physa.2018.08.028.

Mendes, Túlio A. A., Rielva S. C. Nascimento, Renata S. Veras, Marcelo E. Almeida, and Luiz G. Knauer. 2021. "Geological knowledge advances on the Alto Rio Negro region, northwestern Amazonian Craton, Brazil: a review." *Journal of the Geological Survey of Brazil* 4 (3): 209–222. https://doi.org/10.29396/jgsb.2021.v4.n3.2.

Reis, M. M. T., D. S. Ladislau, M. W. S. Ribeiro, C. C. Guimarães, A. J. V. Paiva, D. C. Mattos, P. H. R. Aride, and A. T. Oliveira. 2022. "Socioeconomic aspects and profile of fishing according to fishers of commercial edible fish in the municipality of Barcelos, middle Negro River, Amazonas, Brazil." *Brazilian journal of biology* 82: 1–14. https://doi.org/10.1590/1519-6984.264210.

SGB. 2024. "Serviço Geológico do Brasil (SGB-CPRM)." Accessed: December 11, 2024, `https://www.sgb.gov.br/`.

Sharma, Priyanka, Surjeet Singh, and Survey D. Sharma. 2022. "Artificial Neural Network Approach for Hydrologic River Flow Time Series Forecasting." *Agricultural Research* 11: 465. https://doi.org/10.1007/s40003-021-00585-5.

Silva, I. N., D. H. Spatti, and R. A. Flauzino. 2020. *Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas*. 1st ed. São Paulo: Artliber editora Ltda.

Siqueira, Luan Ferreira, and Naziano Filizola. 2021. "Estudo hidrológico do efeito de barramento hidráulico no rio Tarumã-Açú, Manaus-AM." *Revista Brasileira de Geomorfologia* 22 (2). https://doi.org/10.20502/rbg.v22i2.1752.

Sit, M., B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir. 2020. "A comprehensive review of deep learning applications in hydrology and water resources." *Water Science and Technology* 82: 2635–2670. https://doi.org/10.2166/wst.2020.369.

Steiner, N. Y., D. Hissel, Ph. Moçotéguy, and D. Candusso. 2011. "Diagnosis of polymer electrolyte fuel cells failure modes (flooding & drying out) by neural networks modeling." *International Journal of Hydrogen Energy* 36: 3067–3075. https://doi.org/https://doi.org/10.1016/j.ijhydene.2010.10.077.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov. 2007. "Measuring and testing dependence by correlation of distances." *Ann. Statist.* 35: 2769–2794. https://doi.org/10.1214/009053607000000505.

Tian, H., and Z. Shang. 2006. "Artificial neural network as a classi-

fication method of mice by their calls." *Ultrasonics* 44: e275–e278. https://doi.org/https://doi.org/10.1016/j.ultras.2006.06.012.

Wang, L., Z. Q. Lin, and A. Wong. 2020. "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images." *Scientific Reports* 10: 19549. https://doi.org/https://doi.org/10.1038/s41598-020-76550-z.

Wu, Zhu Y. Wang Y., Y. D., and G. Chiribella. 2024. "Learning quantum properties from short-range correlations using multi-task networks." *Nat. Commun.* 15: 8796. https://doi.org/10.1038/s41467-024-53101-y.

Yan, H., Y. Jiang, J. Zheng, C. Peng, and Q. Li. 2006. "A multilayer perceptron-based medical decision support system for heart disease diagnosis." *Expert Systems with Applications* 30: 272–281. https://doi.org/https://doi.org/10.1016/j.eswa.2005.07.022.

Artigo 3 - Identifying wildfires with convolutional neural networks and remote sensing: application to Amazon rainforest

Artigo publicado e escolhido como destaque na revista internacional indexada **International Journal of Remote Sensing**. Como já mencionado na introdução, os resultados deste terceiro artigo referem-se à disciplina de Geologia Ambiental, cursada no mestrado durante o segundo semestre de 2023.

Eleutério, C. L., Filizola, N. P., de Brito, A. P., Galiceanu, M., & Mendes, C. F. O. (2024). Identifying wildfires with convolutional neural networks and remote sensing: application to Amazon Rainforest. International Journal of Remote Sensing, 46(7), 2665–2688.
https://doi.org/10.1080/01431161.2024.2425119
Acesse o artigo aqui

INTERNATIONAL JOURNAL OF REMOTE SENSING
https://doi.org/10.1080/01431161.2024.2425119

Taylor & Francis
Taylor & Francis Group

Identifying wildfires with convolutional neural networks and remote sensing: application to Amazon rainforest

Cíntia L. Eleutério [a], Naziano P. Filizola [a], Alderlene P. de Brito [a], Mircea Galiceanu [b] and Carlos F. O. Mendes [b]

[a]Departamento de Geociências, Universidade Federal do Amazonas, Manaus, Brazil; [b]Departamento de Física, Universidade Federal do Amazonas, Manaus, Brazil

ABSTRACT

We employ a Convolutional Neural Network (CNN) to classify areas affected by wildfires in the Amazon Rainforest. We focus on areas of the state of Amazonas (Brazil), which accounts for a significant portion of Brazil's wildfires and which in recent years have experienced a notable increase of these incidents. The satellite imagery used in this study was sourced from Landsat 8 and Landsat 9 with a specific combination of spectral bands 7 (shortwave infrared – SWIR 2), 5 (near-infrared), and 2 (visible blue), which were chosen for optimal colour representation, thus conducive to wildfire detection. These images were extensively utilized to train the CNN. Following the training phase, the CNN demonstrated significant efficiency in classifying images not included in the training dataset. This underscores the robustness and generalization capabilities of the CNN model, showcasing its potential for effective wildfire detection beyond the training dataset. Our approach demonstrates a great potential for wildfire detection, serving as a complement to well-established large-scale monitoring systems, such as MODIS and VIIRS, which are widely used for broad, continuous fire detection. Our methodology is not restricted only to the Amazonian ecosystem and has the potential to significantly assist competent authorities in combating and managing such incidents, providing an advanced and more localized approach to wildfire detection.

ARTICLE HISTORY
Received 14 July 2024
Accepted 28 October 2024

KEYWORDS
Amazon Rainforest; Wildfires; Artificial Intelligence; Convolutional Neural Network; Remote Sensing

1. Introduction

The Amazon Rainforest represents a globally significant asset, while exerting a strong influence both socio-environmentally and economically to the northern region of Brazil. Spanning nine South American countries, its ecological importance is underscored by its role in biodiversity preservation and global climate regulation. The region is home for numerous species of animals and vegetation, as well as various indigenous communities, whose cultures are deeply intertwined with the forest preservation (González and Kröger 2020).

The Amazon region, with its extensive resources and rich biodiversity, provides a crucial influence to the local economy and it also substantially contributes to the global

CONTACT Carlos F. O. Mendes cfabio.mendes@gmail.com Departamento de Física - Universidade Federal Do Amazonas, Manaus 69077-000, Brazil

ARTICLE TEMPLATE

Identifying wildfires with convolutional neural networks and remote sensing: application to Amazon Rainforest

Cíntia L. Eleutério[a], Naziano P. Filizola[a], Alderlene P. de Brito[a], Mircea Galiceanu[b], Carlos F. O. Mendes[b]

[a]Departamento de Geociências, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brazil; [b]Departamento de Física, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brazil

ARTICLE HISTORY
Compiled January 15, 2025

Abstract
We employ a Convolutional Neural Network (CNN) to classify areas affected by wildfires in the Amazon Rainforest. We focus on areas of the state of Amazonas (Brazil), which accounts for a significant portion of Brazil's wildfires and which in recent years have experienced a notable increase of these incidents. The satellite imagery used in this study was sourced from Landsat 8 and Landsat 9 with a specific combination of spectral bands 7 (shortwave infrared - SWIR 2), 5 (near-infrared), and 2 (visible blue), which were chosen for optimal color representation, thus conducive to wildfire detection. These images were extensively utilized to train the CNN. Following the training phase, the CNN demonstrated significant efficiency in classifying images not included in the training dataset. This underscores the robustness and generalization capabilities of the CNN model, showcasing its potential for effective wildfire detection beyond the training dataset. Our approach demonstrates a great potential for wildfire detection, serving as a complement to well-established large-scale monitoring systems, such as MODIS and VIIRS, which are widely used for broad, continuous fire detection. Our methodology is not restricted only to the Amazonian ecosystem and has the potential to significantly assist competent authorities in combating and managing such incidents, providing an advanced and more localized approach to wildfire detection.

KEYWORDS
Amazon Rainforest; Wildfires; Artificial Intelligence; Convolutional Neural Network; Remote Sensing.

1. Introduction

The Amazon Rainforest represents a globally significant asset, while exerting a strong influence both socio-environmentally and economically to the northern region of Brazil. Spanning nine South American countries, its ecological importance is underscored by its role in biodiversity preservation and global climate regulation. The region is home for numerous species of animals and vegetation, as well as various indigenous communities, whose cultures are deeply intertwined with the forest preservation (González and Kröger 2020). The Amazon region, with its extensive resources and rich biodiversity, provides a crucial influence to the local economy and it also substantially contributes to

CONTACT Carlos F. O. Mendes. Email: cfabio.mendes@gmail.com

the global economic landscape. In the last decades, the Amazon is facing an escalating crisis characterized by rampant deforestation, unsustainable land-use practices, and the imminent threat of wildfires (Russo Lopes and Bastos Lima 2022; Cardil et al. 2020), compelling the international community to acknowledge the delicate balance between environmental conservation and socio-economic development (Pokorny, P. Pacheco, and Entenmann 2021).

From 1970 onwards, the temperature has tended to increase, intensifying the frequency and the duration of droughts in the Amazon region (Gloor et al. 2015). Most Earth system models predict for the 21st century an increase of the intensity of the dry season in the Amazon Rainforest (Aragão et al. 2018). This is directly related to radiation forces and to decreasing production of aerosol from the Northern Hemisphere (Booth et al. 2012), which tend to cause anomalous variations in sea surface temperatures (SST) and, consequently, lead to fluctuations of the precipitations amount over the Amazon region. If this new climate configuration is confirmed, the Amazon is expected to become an amplified system prone to wildfires (Aragão et al. 2014). Therefore, frequently referred to as the lungs of the Earth, currently the Amazon Rainforest is facing an escalating threat due to the increasing number of wildfires. The environmental repercussions of these fires extend far beyond regional boundaries, impacting global climate patterns and biodiversity. Remote sensing through satellite imagery has become a vital tool in monitoring and addressing this environmental crisis (Young and Onoda 2017). According to reference (Brando et al. 2020), the ability to detect and respond to wildfires in the Amazon is crucial for preserving the delicate ecological balance of this vital ecosystem and the future of this region depends on decisive rapid action.

This paper applies the Convolutional Neural Network (CNN) in the classification of wildfires in the Amazon region by using remote sensing data. Exploring the great potential of the deep learning techniques, CNN presents a powerful solution for the automated detection of the wildfires, offering a timely response to mitigate their devastating effects. Deep learning is a subset of Artificial Intelligence (AI) and Machine Learning and employs algorithms that enhance their performance over time through an exposure to increasing data volumes. This technological synergy leverages statistical procedures that help machines to acquire skills, such as identifying wildfires, through experiential learning. The integration of deep learning techniques to wildfire detection significantly improves accuracy and response time (Park et al. 2022). Deep learning utilizes Artificial Neural Networks (ANN) and by comprising multiple layers stacked sequentially facilitates the knowledge assimilation when have to deal with extensive datasets during its training process. This intricate interplay of technologies underscores the dynamic landscape of AI and its multifaceted applications in addressing critical environmental challenges, such as wildfires in the Amazon region. Highlighting the importance of this investigation, we emphasize the importance of using AI in preserving the ecological integrity of the Amazon Rainforest.

The CNN have become a power tool in computer vision, enabling feature extraction and pattern recognition within images across various domains in the modern science. In this context, we mention only several applications of CNN in image classification. The CNN model has been employed to the classification and diagnosing of COVID-19 (Ismael and Şengür 2021; Wang, Lin, and Wong 2020), the detection of pneumonia using chest x-ray images (Sharma et al. 2020), the analysis of collisions in High Energy Physics (Madrazo et al. 2019), the prediction of molecular properties for aerosol chemistry and health effects (Krüger et al. 2022), the cell type identification in stem cell images (Kusumoto and Yuasa 2019), the intelligent waste identification and recycling

(Wu et al. 2023), the change detection of deforestation in the Brazilian Amazon using Landsat data (de Bem et al. 2020), the wildfire detection (Khan, Patil, and Ali Haider 2020), and also to forest fire smoke detection for unmanned aerial vehicle images (Kim and Muminov 2020).

In this paper, we have implemented a deep learning model for image classification, specifically utilizing CNN through the python libraries TensorFlow and Keras (Chollet 2019; Géron 2022), a model type widely employed in computer vision applications for extracting features from data, such as images. Our primary goal was to employ CNN to detect wildfires in the Amazon region using Landsat 8 and Landsat 9 satellites imagery. We have trained the model with recent images obtained from EarthExplorer (USGS 2023a), which depicted wildfires in the Amazon Rainforest. After the training process, we achieved a satisfactory level of performance by using the CNN, enabling us to apply the network to wildfire classification through satellite images.

The paper is organized as follows. Section 2 provides an overview of the wildfires in Brazil, Brazilian biomes, and a description of the region under study. Section 3 explores the history of wildfires in the state of Amazonas. Section 4 provides the main technical aspects related to the Landsat 8 and Landsat 9 satellites. In Section 5, the artificial neural networks are briefly described, with a particular emphasis on CNNs. Section 6 presents the results. Finally, Section 7 summarizes our findings.

2. Wildfires in Brazil

In this section, we consider the comprehensive assessment of wildfire events in Brazil during 2023, by employing the dataset available from the National Institute for Space Research (INPE) on the TerraBrasilis web platform (INPE 2024). The website provides access to an extensive archive of fire hotspots detected through the reference satellite (AQUA). This comprehensive database offers valuable information about the precise frequency of fire occurrences throughout the country, mainly in the Amazon region, which has been significantly affected during the year 2023 (Observatory 2023). Furthermore, this repository serves as a crucial tool for monitoring fire-prone regions, thus contributing to the development of more effective wildfire management and prevention strategies. Wildfires have become a significant environmental and ecological concern in Brazil, with their impact extending beyond national borders, affecting not only the rich biodiversity and ecosystems of the country but also global climate patterns. The INPE's data provides insights into the spatial and temporal distribution of these wildfires.

In this study we examine data related to the wildfire occurrences in all Brazilian states and collected between January 1st and December 31st, 2023. In Fig. 1(a) we depict the number of wildfire incidents per state, with Pará leading with over 40000 wildfires, corresponding to 21.97% of the total cases. Mato Grosso and Maranhão follow, each of them with more than 20000 wildfires (11.44% and 11.12%, respectively). Having a number of wildfire incidents between 10000 and 20000, the states of Amazonas, Piauí and Bahia account together for 23.42% of the occurrences. The states ranging from Tocantins to Distrito Federal report fewer than 10000 wildfires. Notably, the decreasing trend in wildfire incidents across states follows an exponential fit, illustrated by dark dashed line in the figure, which is equal to $y = 35564.35\exp(-0.22x)$, with 35564.35 representing the scale coefficient and -0.22 giving the decay rate. These results show small deviation from the Pareto principle (or the 80/20 law), which states that roughly 80% of a quantity belongs to only 20% of the population (Barabási and

Pósfai 2016). In our case, the six states with the highest amount of wildfires correspond to 22% of the states and they share 67.95% from the total number of fires.

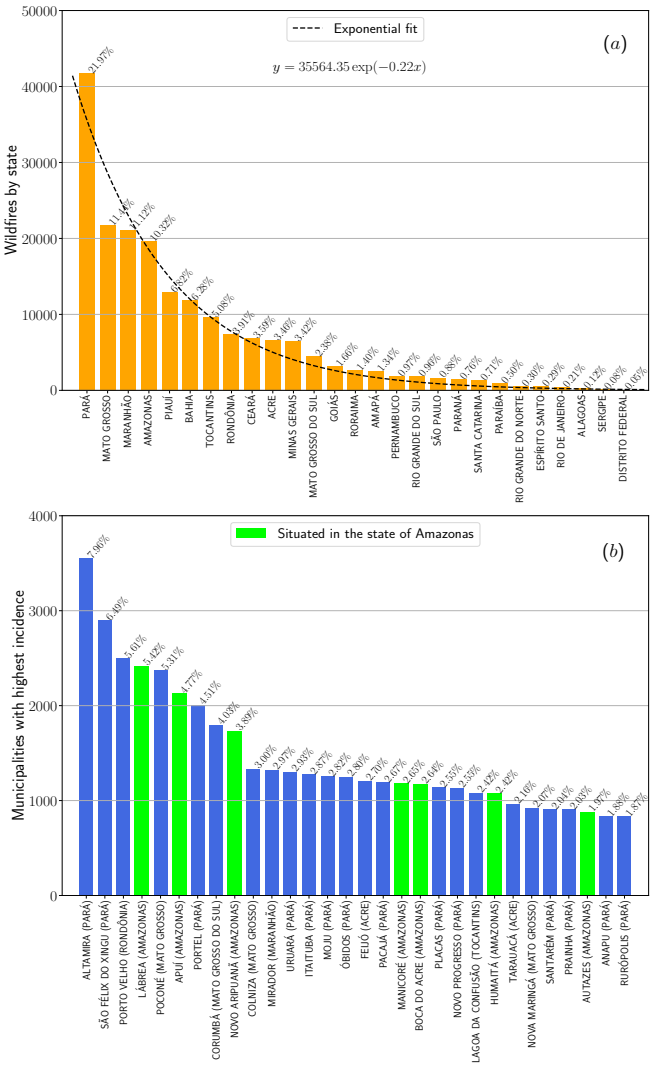


Figure 1. (a) Number of detected wildfires by states in Brazil. The dark dashed curve represents an exponential fit applied to the data. (b) Brazilian municipalities with the highest incidence of wildfire. The lime-colored bars highlight the municipalities situated in the state of Amazonas.

In Fig. 1(b) we show a comprehensive insight to the 30 Brazilian municipalities that exhibited the highest incidences of wildfires in the year 2023. Remarkably, Altamira emerges as the municipality with the most extensive record of detected fires, surpassing 3000 cases, which correspond to a substantial 7.96% of the total cases. It follows São Félix do Xingu and Porto Velho, the capital of Rondônia state, each reporting over 2500 wildfires incidents, together representing 12.1% of the total. Subsequently, we note three municipalities situated in the state of Amazonas (the first three lime-colored bars) - Lábrea, Apuí, and Novo Aripuanã - contributing to 14.17% of the detected cases. In the range of 0 to 1000 incidents, 7 municipalities, spanning from

Tarauacá to Rurópolis, correspond to 14.02% from the total. Notably, 7 of the top 30 municipalities with the highest wildfires incident rates are located in the state of Amazonas. This analysis shows the spatial distribution of wildfire occurrences throughout Brazil, with particular emphasis on the state of Amazonas. Understanding these geographical variations is of paramount importance for the development of targeted fire management and conservation strategies in each region. In the next subsection, we address the wildfires problem within the various Brazilian biomes.

2.1. *Brazilian Biomes*

The richness of biodiversity within the various Brazilian biomes plays a crucial role in maintaining global ecological balance. Brazil hosts an impressive array of ecosystems, including the Amazon, Cerrado, Atlantic Forest, Pantanal, and Caatinga, each with its unique species diversity and characteristics. According to a recent study (Oliveira et al. 2022), these biomes play an important role in Brazil and the wildfires are causing significant damage to all these ecosystems. Due to climate change it is predicted that in the near future the wildfires impact will increase. Thus, protecting these biomes is crucial for biodiversity, ecosystem services, and environmental stability. Effective policies are needed to mitigate this growing threat. These biomes are essential in mitigating climate change, acting as carbon sinks and safeguarding endangered species. The conservation and sustainable management of these areas benefit not only Brazil but also play a fundamental role in our planet's health, making the preservation of these ecosystems a global priority. However, it is important to note that recently, these biomes have been facing a significant increase in wildfires, as can be seen in Table 1, according to INPE data published in 2023 (INPE 2024).

Table 1. The quantity of wildfires detected in the Brazilian biomes from January 1st until December 31st, 2023.

Biome	**Wildfire incidents**	**Percentage** (%)
Amazon Rainforest	98639	51.94
Cerrado	50713	26.70
Caatinga	21550	11.35
Atlantic Forest	11702	6.16
Pantanal	6580	3.46
Pampa	717	0.38

According to Table 1, the biome with the highest incidence of wildfires is the Amazon Rainforest. During the year 2023, a total of 98639 wildfires were recorded in this region, accounting for 51.94% of all cases, representing more than half of all biomes from the country. In second place, Cerrado has 50713 (26.70%) cases detected, while Caatinga and Atlantic Forest together account for 33252 wildfires, which correspond to 17.51%. On the other hand, Pantanal and Pampas are the two biomes least affected by wildfires in 2023, with both combined totaling 7297 (3.84%) detected fires. These findings underscore the critical need for effective wildfire prevention and management strategies. A particular attention should be given to the Amazon Rainforest, as it plays a crucial role in global biodiversity and climate regulation. Furthermore, for this purpose the utilization of AI through CNN, shows a significant potential in enhancing our efforts to the prevention and management of the wildfires (James et al. 2023).

Forest fires have devastating consequences on multiple fronts. The destruction of

vegetation can lead to an irreversible extinction of native species, some of which possess unique medicinal properties, which are found only in the Amazon region (Shanley and Luz 2003). Furthermore, wildfires compromise soil fertility, rendering it barren and hindering the natural recovery of affected areas (Bomfim et al. 2020). The loss of fauna and flora is equally alarming, with fires destroying vital habitats and triggering significant imbalances in the local food chain (Driscoll et al. 2021). Additionally, the intense carbon dioxide (CO_2) emissions into the atmosphere contribute significantly to aggravate the greenhouse effect and global warming, intensifying the global climate challenges and making large fires more frequent in the Amazon Rainforest (Aragão et al. 2014, 2018). Deforestation caused by the wildfires also increases water runoff, raising river levels, as the reduction of vegetation cover decreases water infiltration into the soil and reduces evapotranspiration rates (Guzmán-Rojo et al. 2024). In the next subsection, we present our considered region.

2.2. *Rainforest of the Amazonas state*

In this subsection, we highlight the focal region of interest in this paper, which is displayed in Fig. 2: the state of Amazonas. The figure indicates that the highest concentration of wildfires in the state of Amazonas according to INPE data of 2023, is situated in the south municipalities. The color varies between blue and brown, with blue denoting lower concentrations of detected incidents and brown signifying higher occurrences of wildfires. Intermediate colors represent values falling between these two extremes. Particularly noteworthy are the municipalities of Lábrea and Apuí, identified as the top two with the highest number of detected fire hotspots between January 1st and December 31st, 2023, followed by Novo Aripuanã as the third municipality. These municipalities are distinctively marked with a brown color according to the figure's color palette. This geographical analysis provides valuable insights for understanding the distribution and intensity of wildfires in the southern region of the Amazonas state.

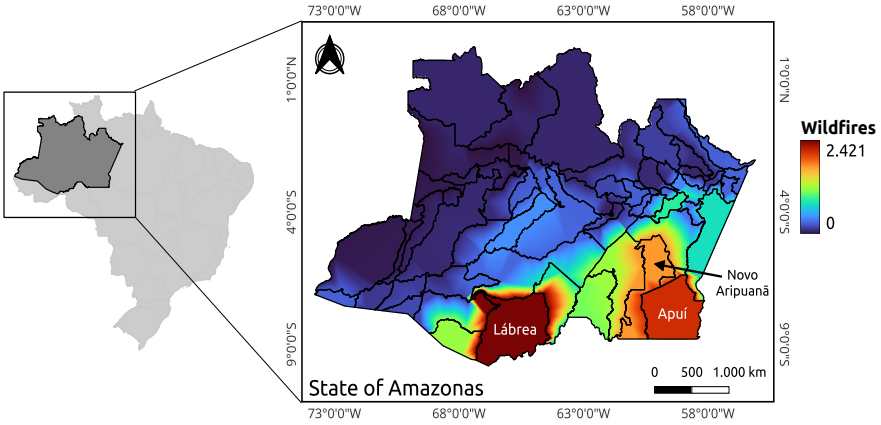


Figure 2. The distribution of wildfires in the state of Amazonas based on INPE data in 2023. Colors indicate the distribution of wildfire occurrences among the municipalities within the state. The highlighted municipalities of Lábrea, Apuí and Novo Aripuanã, situated in the south of the state correspond to the highest incidences.

The satellite images utilized for CNN analysis are specifically concentrated in the Amazonas state regions. These images are sourced from the Landsat 8 and Landsat 9 satellites and cover the period from September to October, 2023 and 2024. This

period was chosen for experiencing the highest frequency of wildfires in the state of Amazonas, with smoke dispersing across a significant part of the state. According to IPAM Amazônia (IPAM 2020), within the Amazon region there are three distinct types of wildfires. First, there are the fires for agricultural management, which are intentionally used by various farmers, including traditional communities, to clear land for pest control and soil renewal in pasture areas. Second, recent deforestation fires are intentionally set to burn cleared vegetation, a cost-effective method aiding soil enrichment for activities like pasture planting. Finally, forest fires, whether accidental or intentional, which are rapidly spreading through live forest, posing a threat to the ecosystem's integrity. In the next section, we present a 25-years (1999-2023) history of wildfires in the Amazonas state.

3. History of wildfires in the state of Amazonas

In this section, we present a brief history of wildfires that occurred between the years 1999 and 2023 within the state of Amazonas. Here, we analyze the incidence data, identifying seasonal patterns and trends over the period. In Fig. 3(a), a heatmap is presented showing the incidences of wildfires in the state of Amazonas. It is important to stress that here we consider data provided by INPE (INPE 2024). It can be observed that between January and June, the number of incidences are low, as indicated by the blue color in the color palette. However, there is a considerable increase in the number of wildfires between the months of July and November. In December, the number of wildfires decreases again. These findings are strongly related to the rain season in Amazonas. Between 1999 and 2001, the incidences of wildfires were low throughout the year. The intensity of wildfires began to increase significantly starting with year 2002, reaching its local maximum in 2005. After a decrease until 2011, the number of wildfires started a steady increase, with this ascending trend being still maintained. The last five years correspond to the highest incidences of wildfires, as evidenced by the brown color.

All the above mentioned observations were confirmed also in Fig. 3(b). In this figure, the dashed blue curve with circles shows that the last two years have recorded the highest incidences, namely 21217 wildfires in 2022 (the maximum value) and 19604 wildfires in 2023. Between 2011 and 2023, the number of total wildfires has been progressively rising, following an exponential curve, $6851.25\exp(0.09x)$. In Fig. 3(c), the maximum (diamond symbols), mean (x symbols), and minimum (square symbols) numbers of wildfires between 1999 and 2023 are shown. The figure clearly indicates that from January to June, there were almost no wildfires, with a significant increase occurring between July and November. The highest number of wildfire counts were 8588 and 8659 for August and September, respectively. For these two months, the mean value was 3737 in August and 3272 wildfires in September. This seasonal pattern reflects the climatic conditions and human activities contributing to the occurrence of wildfires. Understanding these trends is crucial for implementing effective wildfire prevention and control strategies.

Table 2 presents the monthly mean values of maximum $\langle T_{\max}(^{\circ}\mathrm{C})\rangle$ and minimum $\langle T_{\min}(^{\circ}\mathrm{C})\rangle$ temperatures, as well as the total or accumulated monthly precipitation, for the three municipalities with the highest incidences in the Amazonas state during 2023, as shown in Fig. 1(b): Lábrea, Apuí, and Novo Aripuanã. The data shown in the table were obtained from the National Institute of Meteorology (INMET) (INMET 2024). The months selected were August, September, and October 2023, which are the

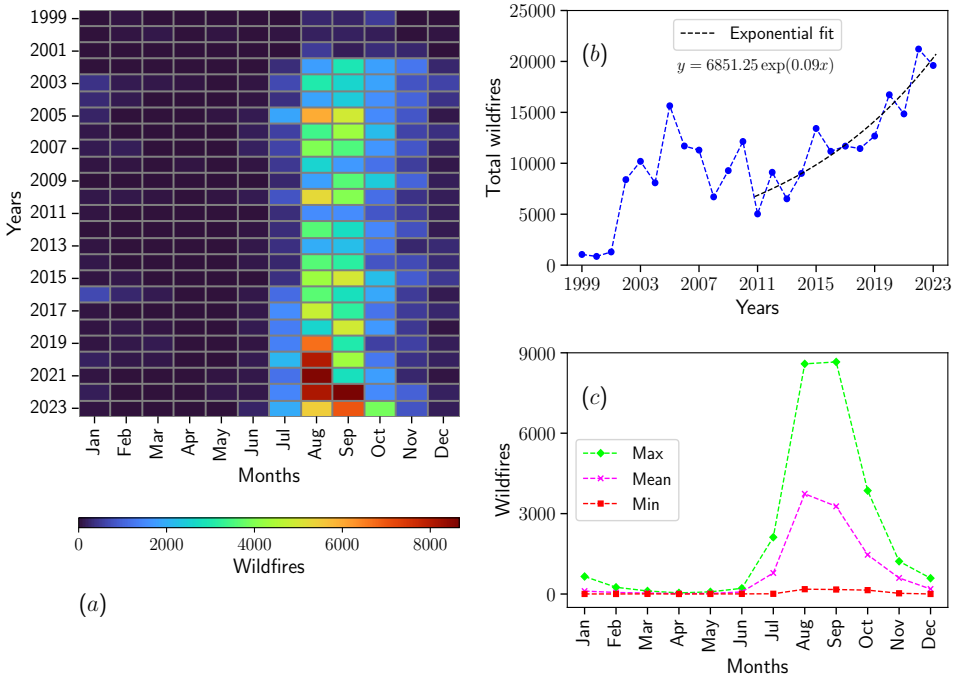


Figure 3. (a) Heatmap showing the number of wildfire incidents by month in Amazonas state between 1999 and 2023, depicted using a color palette. (b) Historical time series of total wildfire incidents in Amazonas over the years. The dark dashed curve represents an exponential fit applied to the data. (c) Comparison of maximum, mean, and minimum incident values by month.

months with the highest numbers of wildfires, as seen in Fig. 3(c). Additionally, Table 2 shows that the mean, maximum and minimum temperatures in the municipalities are quite high, significantly contributing to the large number of wildfires in these regions. For maximum temperatures, the lowest mean value was 26.89°C in August for Apuí, and for minimum temperatures, it was 25.59°C in August, also for Apuí. Regarding accumulated precipitation, the municipality of Apuí recorded the lowest amounts of rainfall observed during the three months considered. The low amount of rainfall also significantly contributes to the occurrence of wildfires in these regions. The lack of precipitation increases the dryness of the soil and vegetation, facilitating the spread of fires.

4. Landsat 8 and 9

The Landsat 8 and Landsat 9 satellites are equipped with advanced instruments that capture images across various spectral bands, essential for wildfire detection and classification (USGS 2023b,c). Their multispectral and thermal capabilities enable the identification of temperature anomalies and the differentiation between vegetation types and burned areas. Both satellites feature the Operational Land Imager (OLI), which captures images in visible, near-infrared (NIR), and shortwave infrared (SWIR) bands. These bands are critical for detecting vegetation changes: healthy vegetation are strongly reflected in the NIR, while burned areas are showing low reflectance. The

Table 2. The mean values of maximum and minimum temperatures, as well as the total precipitation, for the three municipalities with the highest number of detected wildfires from August to October, 2023.

Municipalities	**Months**	$\langle T_{\max}(^{\circ}\mathrm{C})\rangle$	$\langle T_{\min}(^{\circ}\mathrm{C})\rangle$	**Total precipitation** (mm)
Lábrea	Aug	28.75	27.47	33.0
	Sep	29.22	27.86	131.4
	Oct	28.38	28.11	48.8
Apuí	Aug	26.89	25.59	10.6
	Sep	28.41	27.03	9.6
	Oct	28.15	26.68	3.6
Novo Aripuanã	Aug	28.93	27.63	64.2
	Sep	29.44	28.03	34.6
	Oct	30.29	28.78	166.8

SWIR bands are also crucial for detecting moisture in soil and vegetation, helping to identify fire-affected regions.

Additionally, the Thermal Infrared Sensor (TIRS), which is present on both satellites, operates in two thermal infrared bands, providing valuable data on heat sources and surface temperature. While the TIRS offers a spatial resolution of 100 meters, the OLI bands most relevant for wildfire monitoring - Band 7 (SWIR 2), Band 5 (NIR), and Band 2 (Blue) - have a spatial resolution of 30 meters, allowing for more detailed mapping of burned areas and vegetation health. Table 3 provides detailed information about the sensors and spectral bands of Landsat 8 and 9. These spectral bands are particularly useful for identifying temperature anomalies and assessing vegetation health. Landsat 8 and 9 data can be downloaded from EarthExplorer (USGS 2023a).

Table 3. The technical information about the sensors and spectral bands of the Landsat 8 and Landsat 9 satellites. Highlighted bands are used for wildfire detection.

Sensors	**Bands**	**Wavelengths** (μm)	**Spatial resolutions** (m)
OLI/OLI-2	1 Coastal Aerosol	0.43 - 0.45	30
	2 Blue	**0.45 - 0.51**	**30**
	3 Green	0.53 - 0.59	30
	4 Red	0.64 - 0.67	30
	5 Near-Infrared	**0.85 - 0.88**	**30**
	6 SWIR 1	1.57 - 1.65	30
	7 SWIR 2	**2.11 - 2.29**	**30**
	8 Panchromatic (PAN)	0.50 - 0.68	15
	9 Cirrus	1.36 - 1.38	30
TIRS/TIRS-2	10 TIRS 1	10.6 - 11.19	100
	11 TIRS 2	11.5 - 12.51	100

5. Artificial Neural Network

The ANN simulate the operations of the human brain in the field of AI, enabling computational models to discern intricate patterns and address challenges that may surpass human comprehension. These models acquire knowledge from available information, capturing complex patterns to perform tasks such as classification and pattern recognition. While inspired by biology, there is no conclusive evidence supporting the idea that neural networks faithfully replicate brain functionality. Nevertheless, they

can be conceptualized as a mathematical framework that learns from data (Chollet 2019; Géron 2022). Expanding upon the neural network framework, we now shift our focus to CNN, a specialized subset optimized for image analysis and visual pattern recognition.

5.1. *Convolutional neural network*

The CNN have profoundly changed the computer vision and image processing, enabling noteworthy progress in tasks such as object detection, and medical imaging (Yang and Yu 2021). Their hierarchical feature learning and adaptability make them a cornerstone in the development of AI systems for visual data analysis. In this subsection, we provide a comprehensive exposition of the CNN model employed in this article. Our aim is to offer a detailed exploration of the specific CNN architecture utilized within the context of this research, thus providing an in-depth understanding of its application.

In Fig. 4 is shown a schematic representation of the proposed CNN model. In this study, we employ three sets of convolutional and pooling layers in the feature learning phase, along with three fully-connected layers in the classification stage. The input layer establishes a fixed dimension for input images derived from the Landsat 8 and Landsat 9 satellites. The image undergoes convolution operations using shared weights within the convolutional layer. Subsequently, the pooling layer reduces the image's dimensions while striving to preserve essential information. In the classification phase, these extracted features are integrated within the fully-connected layers. Finally, the output layer reveals the image classification, which, in our context, corresponds to two conditions: *wildfire* and *no wildfire*. The CNN employs binary classification, where the label 0 corresponds to wildfire, and 1 to no wildfire. For a more detailed description of the CNN structure, please consult the references (Chollet 2019; Géron 2022; Phung and Rhee 2019).

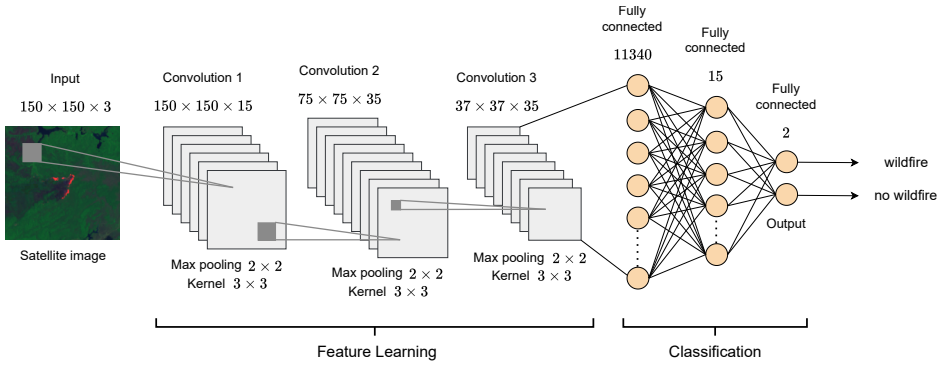


Figure 4. Schematic diagram of the proposed CNN architecture.

In Table 4, the hyperparameters considered for the construction of the proposed CNN are presented. The convolutional layers employ a 3×3 kernel, while the pooling layers utilize a 2×2 stride. The first convolutional layer is configured with 15 filters, and the subsequent two convolutional layers each employ 35 filters. This is followed by a fully connected layer with 11340 artificial neurons, a second fully connected layer with 15 neurons with a dropout rate of 0.5, and an output layer with 2 neurons. Dropout is a regularization technique used in CNN to prevent overfitting. Throughout

the CNN architecture, Rectified Linear Unit (ReLU) activation functions are utilized, culminating in the Sigmoid activation function at the output layer. ReLU activation function outputs the input directly if positive, otherwise, it outputs zero, aiding in faster training and mitigating the vanishing gradient problem. Sigmoid activation function maps inputs to a range between 0 and 1, useful for binary classification by squashing output values.

In the output layer, we employ 2 neurons to accommodate the binary nature of the output - either indicating the presence of a wildfire or the absence thereof. This design is well-suited for our scenario, where we have only two possible conditions: wildfire and no wildfire. It's noteworthy that, for the detection of wildfires of small datasets, a modest computational resource suffices. This CNN is adequate for this purpose, as detecting wildfires and fire-ravaged areas does not require extensive computational complexity. Clearly, the size of the CNN must adapt according to the number of images to be trained by the network and the complexity of the problem. For the construction of the CNN code, we use the python libraries TensorFlow and Keras (Chollet 2019; Géron 2022), ensuring robust and efficient implementation.

Table 4. The hyperparameters of the proposed CNN model.

Layer	**Output**	**Kernel**	**Stride**	**Dropout**
Input	150 × 150 × 3	-	-	-
Convolution 1	150 × 150 × 15	3 × 3	-	-
Relu	150 × 150 × 15	-	-	-
Max pooling	75 × 75 × 15	-	2 × 2	-
Convolution 2	75 × 75 × 35	3 × 3	-	-
Relu	75 × 75 × 35	-	-	-
Max pooling	37 × 37 × 35	-	2 × 2	-
Convolution 3	37 × 37 × 35	3 × 3	-	-
Relu	37 × 37 × 35	-	-	-
Max pooling	18 × 18 × 35	-	2 × 2	-
Flatten	11340	-	-	-
Fully connected	11340	-	-	-
Dropout	11340	-	-	0.5
Fully connected	15	-	-	-
Relu	15	-	-	-
Fully connected	2	-	-	-
Sigmoid	2	-	-	-

In addition to the structural hyperparameters of the CNN, such as convolutional layers, max pooling, and activation functions, other essential hyperparameters were configured to optimize the model's performance. The chosen optimizer was Adaptive Moment Estimation (Adam), which combines the benefits of first- and second-order gradient methods is dynamically adjusting the learning rate for each model parameter. We used a learning rate of 0.001, which is a common value that provides a good balance between convergence speed and training stability and prevents large oscillations in weight adjustments. The batch size was set to 32, allowing efficient processing and accelerating training through more frequent gradient updates. The training was conducted over 100 epochs, with 12 steps per epoch, reflecting the number of gradient updates per epoch based on the dataset size and batch size. These complementary hyperparameters were carefully tuned to ensure the stability and the efficiency of the CNN training process, used for the task of wildfire image classification.

5.2. *CNN architecture and hyperparameter tuning*

To ensure robust validation of the CNN architecture, we utilized k-fold cross-validation, a widely recognized statistical method (Kohavi 1995; Silva, Spatti, and Flauzino 2020). This technique is specifically designed to evaluate the generalization capability of the neural network by assessing its performance on different subsets of data. The primary goal of k-fold cross-validation is to minimize the risk of overfitting and provide a more reliable measure to estimate the model performance on new data.

In this method, the dataset is divided into k equally sized partitions, or folds. The first fold is considered to be the testing set, while the remaining $k-1$ folds are used for training. This process is repeated k times, and in each iteration, a different fold is held out as the testing set. The final performance metric is obtained by averaging the results over all k iterations, providing a more comprehensive behavior across different subsets of data. By systematically rotating the training and testing sets, k-fold cross-validation ensures that every observation in the dataset is used for both training and validation. This approach reduces the bias associated with a single train-test split. Additionally, k-fold cross-validation aids in identifying the best-performing CNN architecture among various candidates. The architecture that consistently shows superior performance across multiple folds is selected for deployment, as it is more likely to generalize well to new, unseen data. In Fig. 5, we display a schematic representation of k-fold cross-validation method, illustrating how the folds and trials are divided. Each fold serves as a validation set while the remaining folds are used for training, ensuring comprehensive evaluation of the model's performance.

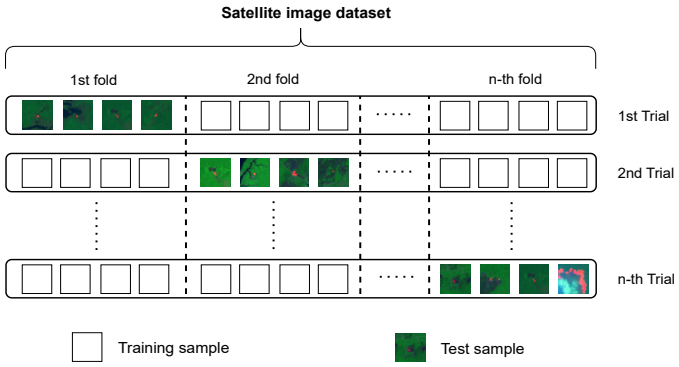


Figure 5. Schematic representation of k-fold cross-validation.

The method of k-fold cross-validation plays a crucial role also in the selection and tuning of the model's hyperparameters. During each iteration of the cross-validation process, different combinations of hyperparameters such as learning rate, number of layers, filter sizes, and the number of neurons per layer, are tested to determine the configuration that yields the best results. Each partition of the dataset is used both for training and validation, thus the impact of each set of hyperparameters is assessed across multiple scenarios, which reduces the risk of overfitting to any specific partition. The main advantage of using k-fold cross-validation on hyperparameter tuning is that the method provides a more robust evaluation of the tested configurations. By averaging the performance metrics over all folds, this method offers a more comprehensive understanding of how the model performs overall, avoiding reliance on a single

train-test split. Thus, k-fold cross-validation not only aids in identifying the best CNN architecture but also ensures that the selected hyperparameters generalize well across different subsets of the data, making the model more reliable and effective for real-world applications.

6. CNN in the wildfire classification

In this section, we present the models of images captured by the Landsat 8 and Landsat 9 satellites that reveal environmental changes, specifically concentrating on spectral bands: 7 (shortwave infrared - SWIR 2), 5 (near-infrared), and 2 (visible blue) (see Table 3). The fusion of these spectral bands generated a composite image, adept at highlighting regions affected by wildfires in the Amazonas state. These satellites images not only serve as a means to identify fire-prone areas but also play a crucial role in the training and validation of the CNN. The utilization of Landsat 8 and Landsat 9 imagery, with its distinctive band combination, emerges as a robust tool for monitoring and comprehending the impact of wildfires on the landscape. The synergy of satellites imagery and AI enables more efficient monitoring to mitigate the impacts of wildfires. Utilizing satellites data, and employing AI, notably the CNN, facilitates timely detection and assessment of wildfires areas. This integrated approach offers a proactive solution, reducing the potential impacts of wildfires by enhancing our ability to respond swiftly and effectively to environmental changes. Moreover, the model can be used in supporting competent authorities, providing valuable insights and aiding in the overall management and response to wildfire incidents.

In Fig. 6, we illustrate an example of a Landsat 9 satellite image used for training and validation of the CNN. In Fig. 6(a), the image is displayed in colors, generated from the combinations of spectral bands 7, 5, and 2. This image corresponds to the date of October 4th, 2023, and is located in the state of Amazonas, close to the city of Manaus. In Fig. 6(b), we depict the location of the image in relation to the Amazonas state and also Brazil. Figure 6(c) zooms in on the bottom part of the image shown in (a), revealing four wildfires in this region. In this image, it is possible to identify four wildfires in a small region. The wildfire in the lower part of the image exhibits a significantly larger size compared to the other three, providing a visual indication of its heightened intensity. Additionally, the image captures the discernible spread of smoke from this larger wildfire, offering valuable insights into the environmental impacts. The smoke generated by these wildfires can pose significant environmental and human health risks. It not only compromises air quality but also contains fine particles that can impact respiratory health (Aguilera et al. 2021). A thorough analysis of the environmental and health impacts associated with these events is crucial.

The total dataset for the CNN's training and validation consisted of 484 standardized images, each with an input size of 150 pixels in width and 150 pixels in height. All these images were derived from a set of 26 Landsat 8 and Landsat 9 images. In Appendix A, the identifiers of the 26 Landsat 8 and 9 images used in this research, obtained from (USGS 2023a), can be observed. Of the 484 images, a total of 242 are wildfire images and 242 are no wildfire images. Out of these, 200 wildfire images and 200 no wildfire images were used for CNN training, while the remaining 42 wildfire images and 42 no wildfire images were allocated for validation. The scale used for the figures was 1 : 50000, as this scale is important to clearly show the pixels that display the fires. The CNN was trained with two classes: one representing wildfire occurrences and the other representing no wildfire instances. Consequently, the number of images (with

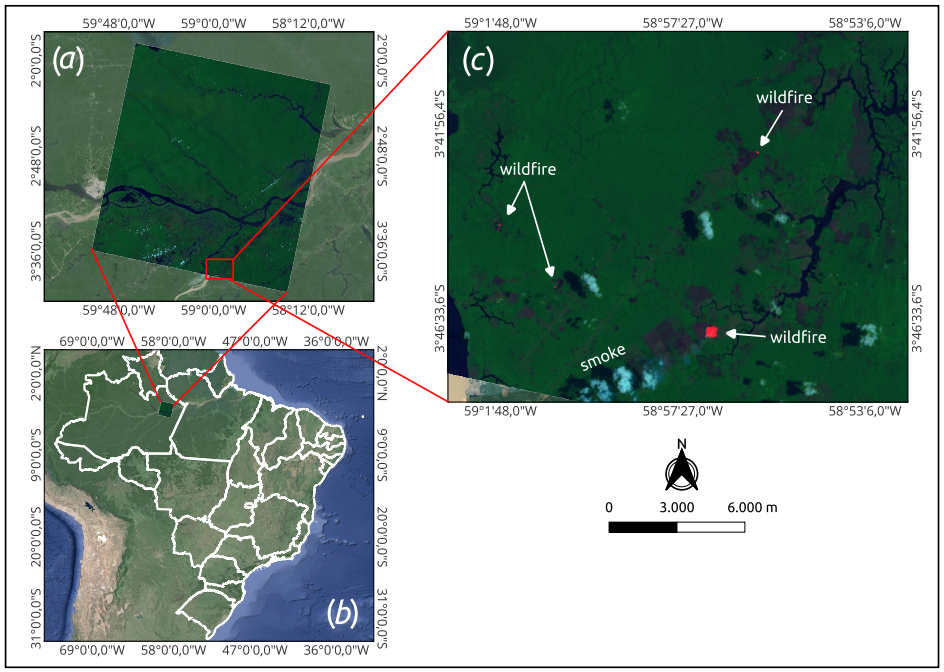


Figure 6. (*a*) Landsat 9 satellite image highlighted in the Amazon region, combining spectral bands 7, 5, and 2. (*b*) Location of the satellite image in relation to the state of Amazonas and Brazil. (*c*) Zooming in on a small region of the satellite image, with a focus on four wildfires in that area. This type of image was used in the training and validation of the CNN. This image corresponds to the date of October 4th, 2023.

and no wildfires) was balanced to ensure a balanced learning approach. Additionally, it is mentioned that in the classification of images, the network demonstrates effective efficiency.

The method for identifying wildfires in the Landsat 8 and 9 satellite images used specific spectral bands (7, 5, and 2) to detect fire activities, focusing on elevated thermal emissions and vegetation stress. This approach allowed for the selection of regions exhibiting significant wildfire signals, ensuring that areas with high thermal emissions and associated vegetation stress due to fire were captured. Regarding the image dimensions, at 150 × 150 pixels, the presence of wildfire was determined based on a distinct and observable portion of the image revealing fire activity. Although wildfires vary in size, the images used for training were designed to include events with sufficient spatial representation, allowing clear differentiation of wildfire activity from other phenomena. An image was labeled as "wildfire" if it showed clear evidence of fire, regardless of whether the burn occupied a small or large fraction of the image. This approach aligns with the goal of training the CNN to identify patterns and indicators of fires, even when located in just a part of the image. It is recognized that the variability in wildfire size may influence the performance of the network. However, the model is able to generalize well in classifying wildfires of different sizes.

The Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS) systems are widely employed by INPE for large-scale, continuous monitoring of wildfires across Brazil. These systems are designed to provide nearly real-time data with moderate spatial resolutions (INPE 2023, 2015).

This capability enables the detection of heat sources and wildfires over extensive areas, playing a crucial role in global fire and deforestation monitoring, particularly in the Amazonian region. However, their moderate resolution may limit the ability to capture finer details in remote areas or smaller fire outbreaks. In this context, the CNN model, which utilizes high-resolution Landsat images combined with specific spectral bands, represents a good alternative for a more precise wildfire detection in localized smaller regions, but lacks the frequent temporal coverage of MODIS and VIIRS due to longer revisit times. Our approach complements the above-mentioned large-scale monitoring systems by providing a highly detailed and efficient monitoring in critical environmental preservation zones. Therefore, combining the frequent temporal data from MODIS and VIIRS with the spatial precision of our model, it is possible to achieve a more comprehensive and efficient approach to wildfire monitoring.

6.1. *Data preprocessing and cloud interference*

Cloud interference is a common challenge in satellite imagery, particularly with data from Landsat 8 and 9. In this study, no specific preprocessing steps, such as cloud masking or removal, were applied to the images. This decision was made intentionally to evaluate the robustness of the CNN under realistic conditions where cloud interference is inevitable. By utilizing the data as provided by public Landsat repositories, the study aimed to simulate real-world challenges in wildfire detection. While this approach introduces potential noise, it emphasizes the ability of the CNN to autonomously identify relevant patterns despite such variations. The results suggest that the model can effectively generalize and perform well even in the presence of cloud interference. Future studies may explore the inclusion of cloud detection and removal techniques, which could enhance the reliability and performance of the model further. However, the results of this study highlight the capability of the CNN to manage data with potential cloud-related noise without extensive preprocessing.

6.2. *CNN performance in wildfire classification*

In the final testing phase of the model, after completing the training and validation process, a confusion matrix was generated to evaluate the robustness of the CNN in distinguishing between wildfire and no wildfire. For this test, we used 40 images that were not part of the 484 images used during the training and validation processes. These new 40 images were divided between the two classes, with 24 wildfire images and 16 no wildfire images. The resulting confusion matrix is presented in Table 5. From the table, it can be observed that the model correctly identified 23 wildfire cases, with only 1 false negative among the 24 images labeled as wildfire. Furthermore, among the 16 no wildfire images, the model correctly classified all of them, *i.e.*, no false positives were recorded.

Table 5. Confusion matrix obtained from the final test phase.

	Predicted wildfire	**Predicted no wildfire**
Actual wildfire	23	1
Actual no wildfire	0	16

The following measures can be derived from the confusion matrix: Precision, Re-

call, F1-Score, and Accuracy (Sokolova and Lapalme 2009). These metrics provide an overview of the model's performance in the classification task and they are computed by knowing four key components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). TP (TN) corresponds to the number of examples correctly predicted as positive (negative), while FP (FN) gives the number of examples incorrectly predicted as positive (negative).

- **Precision** represents the proportion of true positives from all examples predicted as positive (TP + FP). In other words, it reflects the model's ability to avoid false positives. Precision is calculated as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{1}$$

  Our model made no errors in identifying the wildfire class: the Precision value equals 1.00, which means that all images predicted as wildfires truly belonged to that class.
- **Recall** is also known as sensitivity and it measures the proportion of true positives identified correctly from all actual positives (TP + FN). Thus, the Recall is given by:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{2}$$

  Our model achieved a Recall of 0.958, indicating that it correctly identified 95.8% of all wildfire images. We had only 1 false negative, meaning that a wildfire image was incorrectly classified as no wildfire.
- **F1-Score** is the harmonic mean of Precision and Recall. It provides a balanced metric that accounts for both false positives and false negatives. The F1-Score is calculated as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

  The F1-Score value of 0.979 gives our model's overall classification performance, obtained by considering both Precision and Recall.
- **Accuracy** measures the ratio between correct predictions (true positives and true negatives) and the total predictions made by the model. Accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

  The Accuracy of our model is equal to 0.975, which means that 97.5% of all images in the test set were correctly classified.

The performance metrics are summarized in Table 6, showing that the model achieved a very good efficiency in distinguishing between wildfire and no wildfire classes.

In summary, the confusion matrix and derived performance metrics demonstrate the effectiveness of the CNN in the task of classifying wildfire and no wildfire images. The

Table 6. Performance metrics for the final test phase.

Metric	Value
Precision	1.000
Recall	0.958
F1-Score	0.979
Accuracy	0.975

high precision and accuracy indicate that the model is highly reliable in predicting wildfires.

6.3. *Loss and accuracy trends*

In Fig. 7, the loss and accuracy of CNN are depicted. Figure 7(a) illustrates the loss over epochs. The continuous red curve represents the training loss calculated by binary crossentropy, starting at approximately 0.5 and gradually decreasing to just below 0.2 after 100 epochs. The curve show a tendency to converge towards values close to zero. Simultaneously, the continuous blue curve represents the validation loss, which exhibits the same behavior, starting at approximately 0.5 and gradually decreasing to just below 0.2 after 100 epochs. The model does not show signs of overfitting, as the training and validation loss curves converge closely. Both curves maintain a similar downward trend, indicating consistent generalization to unseen data. In Fig. 7(b), the accuracy over epochs is presented, with the continuous lime color curve representing the training accuracy. The training accuracy starts at approximately 0.7 and shows convergence towards 1.0 around epoch 100. Conversely, the continuous magenta curve illustrates the validation accuracy, showing similar behavior and reaching around 1.0 by epoch 100. Both curves demonstrate notable accuracy at epoch 100. In Fig. 7, a moving average was applied to smoothen the curves, with a window size of 10. Therefore, only 90 epochs are displayed in Fig. 7 to better highlight the trend without excessive fluctuations.

Binary crossentropy is a commonly employed loss function in binary classification tasks. It quantifies the divergence between predicted probabilities and actual labels, with a focus on penalizing incorrect predictions more heavily. The term epochs refers to the number of times the entire dataset is used to train a neural network. During each epoch, the data is fed through the network, and the weights are adjusted based on the discrepancies between predictions and actual labels. This iterative process is repeated across multiple epochs to incrementally improve the model's performance.

Table 7 provides an overview of the CNN capability to identify or classify images containing and not containing wildfires. As an illustration, we selected 9 images (from the 40) that were not seen by the CNN during both the training and validation processes. Their performance is detailed in the table. After a more detailed analysis of Table 7 we observe that there are only one image with a probability below 95%. Thus, all 9 images not previously seen by the CNN were successfully predicted their classifications with a high probability. The neural network's proficiency in classifying wildfires based on satellite imagery underscores its critical importance in the early and effective detection of forest fire events. The utilization of these technologies provides an advanced and responsive approach to monitor extensive areas, significantly contributing to the prevention and swift response to critical wildfire situations.

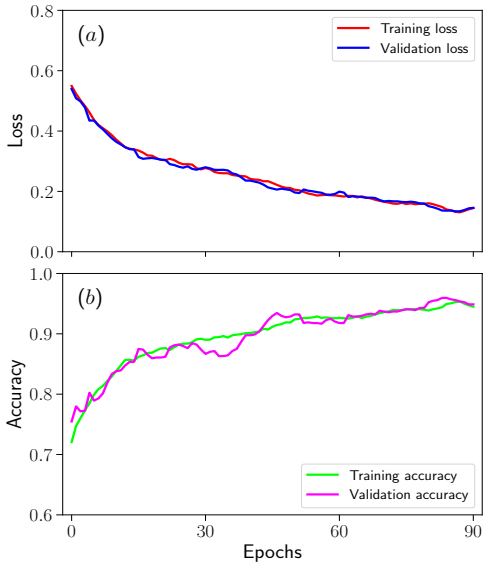


Figure 7. (a) The training and validation loss, and (b) the training and validation accuracy in relation to the number of epochs.

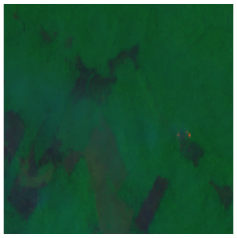
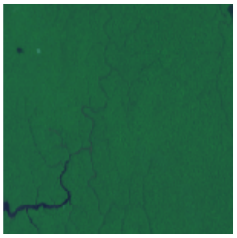
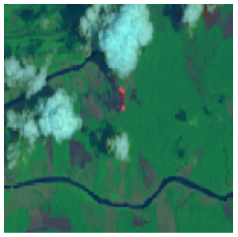
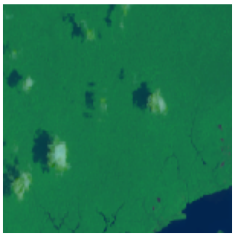
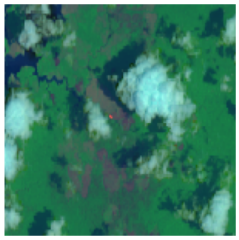
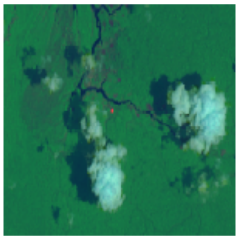
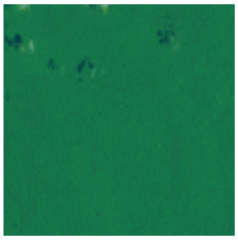
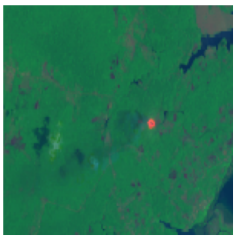
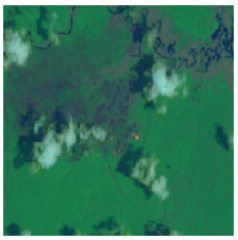
7. Conclusions

In this paper, a Convolutional Neural Network (CNN) was employed for the classification of wildfires in the Amazon region, specifically utilizing Landsat 8 and Landsat 9 satellites images captured in the state of Amazonas. The trained network, utilizing a combination of spectral bands 7, 5, and 2, demonstrated remarkable efficiency in image classification, achieving a very good accuracy, *i.e.* 93% for the training accuracy and 94% for the validation accuracy. Notably, the implementation of a compact neural network proved to be efficient. Additionally, it is emphasized that the methodology is not limited to Landsat 8 and Landsat 9. By combining the spectral bands in a suitable manner to unveil fire occurrences, the images from other satellites can be integrated, enhancing the versatility of the approach.

Increasing the number of training images will undoubtedly lead to a more robust model. As the dataset expands, it may be necessary to adjust the network architecture to effectively handle the additional data. However, these modifications should be executed with careful management of the hyperparameters, utilizing methods like cross-validation, as employed in this research, to ensure optimal performance and mitigate the risk of overfitting. Moreover, the quality of the images is crucial, as high-quality images directly enhance the model's effectiveness. Therefore, choosing the appropriate height and width in pixels for the images is essential. The chosen scale for the images is of fundamental importance, as it makes the wildfire evident in the visual data. Factors such as image resolution, batch size, learning rate, and the choice of activation functions also significantly influence the model's learning capacity. Thus, a holistic approach that considers these aspects will further strengthen the model's performance in classifying images accurately.

The findings of this research hold significant implications for environmental monitoring and disaster response. The ability to accurately classify and detect forest fires using satellite imagery, especially in a region as ecologically crucial as the Amazon,

Table 7. Components of Landsat 8 and Landsat 9 satellites images (150×150 pixels) and the CNN classifications between *wildfire* and *no wildfire*.

 Class: wildfire; Prob.: 82%.	 Class: no wildfire; Prob.: 96%.	 Class: wildfire; Prob.: 100%.
 Class: no wildfire; Prob.: 95%.	 Class: wildfire; Prob.: 100%.	 Class: wildfire; Prob.: 100%.
 Class: no wildfire; Prob.: 97%.	 Class: wildfire; Prob.: 100%.	 Class: wildfire; Prob.: 100%.

provides a valuable tool for early intervention and mitigation. Notably, the Amazon Rainforest biome accounts for a significant percentage, 51.94%, of wildfires in Brazil. In the state of Amazonas, specifically, the period most affected by wildfires is between July and November, and has been experiencing an increase in wildfires in recent years. Moreover, the simplicity and efficiency of the implemented neural network show the feasibility of incorporating such technologies into real-world applications.

In Brazil the MODIS and VIIRS systems play a crucial role in large-scale wildfire monitoring, providing nearly real-time data with moderate resolutions. However, their ability to detect details in remote areas or smaller fire outbreaks is limited. In this regard, the CNN model, utilizing high-resolution Landsat images, could serve as a valuable addition, enabling more detailed analyses in specific regions. By combining the wide temporal coverage of these sensors with the spatial precision of our model, we can significantly enhance wildfire monitoring in critical environmental preservation zones.

In conclusion, this study not only contributes to the field of remote sensing and deep learning but also offers practical insights for relevant authorities. The integration

of satellite images and AI presents a proactive approach to addressing the escalating challenges posed by forest fires, promoting a well-informed and timely response to protect our ecosystems. Although the method followed standard training and testing practices, its application for environmental monitoring in a challenging region represents a significant contribution. The approach dealt with adverse conditions such as dense vegetation cover, atmospheric interference (clouds and smoke), and the absence of extensive preprocessing steps. These decisions reinforce the practical applicability of the model, aiming to provide a realistic and scalable solution for authorities and institutions in combating forest fires in the Amazon. The study highlights the relevance of the Amazonian context and suggests future extensions that could explore more sophisticated network architectures and incorporate additional preprocessing. Other applications can be explored by the CNN and satellite images, such as monitoring and control of Amazonian deforestation.

Acknowledgements

C.L.E. thanks the Amazonas State Research Support Foundation (FAPEAM) and C.F.O.M. and M.G. thank the National Council for Scientific and Technological Development (CNPq) for financial support.

Disclosure statement

The authors declare no competing interests.

Availability of data

The data (images) of the Landsat 8 and 9 satellites used in this study are openly available in reference (USGS 2023a). The identifiers of the images can be seen in Appendix A. Additionally, the 484 images used in the training and validation processes of our CNN can also be provided upon request.

Appendix A. Identifiers of the images

The images of the Landsat 8 and 9 satellites are openly available in (USGS 2023a). Table A1 shows the image identifiers.

References

Aguilera, R., et al. 2021. "Wildfire smoke impacts respiratory health more than fine particles from other sources: observational evidence from Southern California." *Nature Commununications* 12: 1493. https://doi.org/10.1038/s41467-021-21708-0.

Aragão, L. E. O. C., et al. 2018. "21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions." *Nature Communications* 9 (1): 536. https://doi.org/10.1038/s41467-017-02771-y.

Aragão, Luiz E. O. C., Benjamin Poulter, Jos B. Barlow, Liana O. Anderson, Yadvinder Malhi, Sassan Saatchi, Oliver L. Phillips, and Emanuel Gloor. 2014. "Environmental

Table A1. Identifiers of the Landsat 8 and 9 images used in this study.

ID images
LC08_L2SP_230062_20231028_20231101_02_T1
LC08_L2SP_230063_20231028_20231101_02_T1
LC09_L2SP_230062_20231004_20231005_02_T1
LC09_L2SP_230063_20231004_20231005_02_T1
LC09_L2SP_230064_20231004_20231005_02_T1
LC09_L2SP_231064_20231011_20231012_02_T1
LC08_L2SP_230062_20230926_20231003_02_T1
LC08_L2SP_230063_20230926_20231003_02_T1
LC09_L2SP_230063_20231020_20231024_02_T1
LC09_L2SP_229065_20240913_20240914_02_T1
LC08_L2SP_001066_20240908_20240914_02_T1
LC08_L2SP_230062_20240827_20240831_02_T1
LC08_L2SP_230063_20240827_20240831_02_T1
LC08_L2SP_231065_20240903_20240906_02_T1
LC09_L2SP_229063_20240913_20240914_02_T1
LC09_L2SP_229065_20240828_20240829_02_T1
LC09_L2SP_230062_20240904_20240906_02_T1
LC09_L2SP_230063_20240904_20240906_02_T1
LC09_L2SP_230064_20240904_20240906_02_T1
LC09_L2SP_230065_20240904_20240906_02_T1
LC09_L2SP_230066_20240904_20240906_02_T1
LC09_L2SP_231062_20240911_20240912_02_T1
LC09_L2SP_231063_20240911_20240912_02_T1
LC09_L2SP_231064_20240911_20240912_02_T1
LC09_L2SP_232065_20240902_20240903_02_T1
LC09_L2SP_233064_20240909_20240910_02_T1

change and the carbon balance of Amazonian forests." *Biological Reviews* 89 (4): 913–931. https://doi.org/10.1111/brv.12088.

Barabási, Albert-László, and Márton Pósfai. 2016. *Network science*. Cambridge University Press.

Bomfim, Barbara, Lucas C. R. Silva, Ben H. Marimon-Júnior, Beatriz S. Marimon, Timothy A. Doane, and William R. Horwath. 2020. "Fire Affects Asymbiotic Nitrogen Fixation in Southern Amazon Forests." *Journal of Geophysical Research: Biogeosciences* 125 (2): e2019JG005383. https://doi.org/https://doi.org/10.1029/2019JG005383.

Booth, B., et al. 2012. "Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability." *Nature* 484: 228–232. https://doi.org/10.1038/nature10946.

Brando, Paulo, et al. 2020. "Amazon wildfires: Scenes from a foreseeable disaster." *Flora* 268: 151609. https://doi.org/10.1016/j.flora.2020.151609.

Cardil, Adrián, et al. 2020. "Recent deforestation drove the spike in Amazonian fires." *Environmental Research Letters* 15 (12): 121003. https://doi.org/10.1088/1748-9326/abcac7.

Chollet, François. 2019. *Deep Learning with python*. 1st ed. Manning Publications Co.

de Bem, Pablo Pozzobon, et al. 2020. "Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks." *Remote Sensing* 12 (6). https://doi.org/10.3390/rs12060901.

Driscoll, Don A., et al. 2021. "How fire interacts with habitat loss and fragmentation." *Biological Reviews* 96 (3): 976–998. https://doi.org/https://doi.org/10.1111/brv.12687.

Géron, Aurélien. 2022. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 3rd ed. O'Reilly Media.

Gloor, M., et al. 2015. "Recent Amazon climate as background for possible ongoing and future changes of Amazon humid forests." *Global Biogeochemical Cycles* 29 (9): 1384–1399. https://doi.org/10.1002/2014GB005080.

González, Nidia Catherine, and Markus Kröger. 2020. "The potential of Amazon indigenous agroforestry practices and ontologies for rethinking global forest governance." *Forest Policy and Economics* 118: 102257. https://doi.org/10.1016/j.forpol.2020.102257.

Guzmán-Rojo, Mónica, Jeanne Fernandez, Paul d'Abzac, and Marijke Huysmans. 2024. "Impacts of Wildfires on Groundwater Recharge: A Comprehensive Analysis of Processes, Methodological Challenges, and Research Opportunities." *Water* 16 (18). https://doi.org/10.3390/w16182562.

INMET. 2024. "Instituto Nacional de Meteorologia." Accessed 2023-06-30. `https://portal.inmet.gov.br/`.

INPE. 2015. "Intercomparação de produtos AQUA&TERRA/MODIS e Suomi-NPP/VIIRS sobre o Brasil." Accessed: 2024-09-20, `http://www.dsr.inpe.br/sbsr2015/files/p1232.pdf`.

INPE. 2023. "Sensores Ópticos de Resolução Espacial Moderada." Accessed: 2024-09-20, `http://www.dsr.inpe.br/DSR/areas-de-atuacao/sensores-plataformas/sensores-opticos-resolucao-espacial-moderada`.

INPE. 2024. "Plataforma web TerraBrasilis do INPE." Accessed 2024-06-26. `https://terrabrasilis.dpi.inpe.br/`.

IPAM. 2020. "Tudo o que voce queria saber sobre fogo na amazônia mas não sabia para quem perguntar." Accessed 2023-11-18. `https://ipam.org.br/cartilhas-ipam/foconofogo/`.

Ismael, Aras M., and Abdulkadir Şengür. 2021. "Deep learning approaches for COVID-19 detection based on chest X-ray images." *Expert Systems with Applications* 164: 114054. https://doi.org/10.1016/j.eswa.2020.114054.

James, George L., et al. 2023. "An Efficient Wildfire Detection System for AI-Embedded Applications Using Satellite Imagery." *Fire* 6 (4). https://doi.org/10.3390/fire6040169.

Khan, Muhammad Safeer, Rajvardhan Patil, and Syed Ali Haider. 2020. "Application of Convolutional Neural Networks For Wild Fire Detection." 2: 1–5. https://doi.org/10.1109/SoutheastCon44009.2020.9368283.

Kim, S. Y., and A. Muminov. 2020. "Forest Fire Smoke Detection Based on Deep Learning Approaches and Unmanned Aerial Vehicle Images." *Sensors (Basel)* 23 (12): 5702. https://doi.org/10.3390/s23125702.

Kohavi, Ron. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, San Francisco, CA, USA, 1137–1143. Morgan Kaufmann Publishers Inc.

Krüger, Matteo, et al. 2022. "Convolutional neural network prediction of molecular properties for aerosol chemistry and health effects." *Natural Sciences* 2 (4): e20220016. https://doi.org/10.1002/ntls.20220016.

Kusumoto, D., and S. Yuasa. 2019. "The application of convolutional neural network to stem cell biology." *Inflammation and Regeneration* 39 (14). https://doi.org/10.1186/s41232-019-0103-3.

Madrazo, Celia Fernández, Ignacio Heredia, Lara Lloret, and Jesús Marco de Lucas. 2019. "Application of a Convolutional Neural Network for image classification for the analysis of collisions in High Energy Physics." *EPJ Web Conf.* 214: 06017. https://doi.org/10.1051/epjconf/201921406017.

Observatory, Earth. 2023. "Drought Fuels Wildfires in the Amazon." Accessed 2023-10-22. `https://earthobservatory.nasa.gov/images/151965/drought-fuels-wildfires-in-the-amazon`.

Oliveira, Ubirajara, et al. 2022. "Determinants of Fire Impact in the Brazilian Biomes." *Frontiers in Forests and Global Change* 5. https://doi.org/10.3389/ffgc.2022.735017.

Park, Minsoo, Dai Quoc Tran, Jinyeong Bak, and Seunghee Park. 2022. "Advanced wildfire detection using generative adversarial network-based augmented datasets and weakly

supervised object localization." *International Journal of Applied Earth Observation and Geoinformation* 114: 103052. https://doi.org/https://doi.org/10.1016/j.jag.2022.103052.

Phung, Van Hiep, and Eun Joo Rhee. 2019. "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets." *Applied Sciences* 9 (21). https://doi.org/10.3390/app9214500.

Pokorny, B., W. de Jong P. Pacheco, and S. K. Entenmann. 2021. "Forest frontiers out of control: The long-term effects of discourses, policies, and markets on conservation and development of the Brazilian Amazon." *Ambio* 50 (12): 2199–2223. https://doi.org/10.1007/s13280-021-01637-4.

Russo Lopes, Gabriela, and Mairon G. Bastos Lima. 2022. "Understanding deforestation lock-in: Insights from Land Reform settlements in the Brazilian Amazon." *Frontiers in Forests and Global Change* 5. https://doi.org/10.3389/ffgc.2022.951290.

Shanley, Patricia, and Leda Luz. 2003. "The Impacts of Forest Degradation on Medicinal Plant Use and Implications for Health Care in Eastern Amazonia." *BioScience* 53 (6): 573–584. https://doi.org/10.1641/0006-3568(2003)053[0573:TIOFDO]2.0.CO;2.

Sharma, Harsh, Jai Sethia Jain, Priti Bansal, and Sumit Gupta. 2020. "Feature Extraction and Classification of Chest X-Ray Images Using CNN to Detect Pneumonia." 227–231. https://doi.org/10.1109/Confluence47617.2020.9057809.

Silva, I. N., D. H. Spatti, and R. A. Flauzino. 2020. *Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas*. 1st ed. São Paulo: Artliber editora Ltda.

Sokolova, Marina, and Guy Lapalme. 2009. "A systematic analysis of performance measures for classification tasks." *Information Processing & Management* 45 (4): 427–437. https://doi.org/https://doi.org/10.1016/j.ipm.2009.03.002.

USGS. 2023a. "EarthExplorer." Accessed 2023-10-10. `https://earthexplorer.usgs.gov/`.

USGS. 2023b. "Landsat 8 Satellite." Accessed 2023-10-07. `https://www.usgs.gov/landsat-missions/landsat-8`.

USGS. 2023c. "Landsat 9 Satellite." Accessed 2023-12-07. `https://www.usgs.gov/landsat-missions/landsat-9`.

Wang, L., Z. Q. Lin, and A. Wong. 2020. "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images." *Scientific Reports* 10: 19549. https://doi.org/10.1038/s41598-020-76550-z.

Wu, Ting-Wei, Hua Zhang, Wei Peng, Fan Lü, and Pin-Jing He. 2023. "Applications of convolutional neural networks for intelligent waste identification and recycling: A review." *Resources, Conservation and Recycling* 190: 106813. https://doi.org/10.1016/j.resconrec.2022.106813.

Yang, Ruixin, and Yingyan Yu. 2021. "Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis." *Frontiers in Oncology* 11. https://doi.org/10.3389/fonc.2021.638182.

Young, O. R., and M. Onoda. 2017. *Satellite Earth Observations in Environmental Problem-Solving*. Springer Singapore.

Artigo 4 - Hydrological Modeling using Multiple Gaussians: application to Negro River Basin

Artigo em fase de preparação para submissão em revista internacional idexada:

Hydrological Modeling using Multiple Gaussians: application to Negro River Basin

Carlos F. O. Mendes[a], Cíntia L. Eleutério[b], Naziano P. Filizola[b], Mircea Galiceanu[a]

[a]*Departamento de Física, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brasil*
[b]*Departamento de Geociências, Universidade Federal do Amazonas, Manaus, 69077-000, AM, Brasil*

Abstract

We use multiple Gaussian distributions to model the data from some hydrometric stations across the Negro River basin. To characterize hydrological seasonality we use statistical and unsupervised learning techniques to analyze the water level data. The use of Gaussian fitting yields a good approach to identify key hydrological patterns, providing a robust framework for distinguishing seasonal variations and long-term trends. Furthermore, our findings highlight the variability and similarities of the water level dynamics for the chosen stations of Serrinha, Santa Maria do Boiaçú, and Moura. The proposed methodology effectively captures the observed hydrological behavior, accurately reproducing the water level fluctuations recorded at these stations. This approach provides valuable information on the hydrological dynamics of the basin, contributing to improved water resource management and the evaluation of hydrological patterns in the Amazon region.

Keywords: Hydrological Modeling, Gaussian Fitting, Negro River Basin.

1. Introduction

The Negro River basin is an extensive and complex hydrological system within the larger Amazon basin. Spanning multiple South American countries, its largest portion is situated in Brazil. The basin is known for its unique geological structure and intricate network of interconnected channels, which make its hydrodynamics highly intricate and a subject of continuous study [1, 2, 3]. The Negro River serves as the main watercourse, playing a fundamental role at its confluence with the Solimões River near Manaus, where they merge to form the Amazon River [4]. Beyond its ecological importance, the basin is fundamental to regional economies and local communities [5]. Advances in hydrological modeling have significantly improved the understanding of its dynamics, which is highly relevant nowadays as extreme floods and droughts became more frequent .

The Negro River basin is part of the Amazon basin, recognized as one of the largest tropical basins in the world. The Amazon basin covers approximately 6.1×10^6 km^2, representing about 5% of the planet's continental landmass. This vast hydrological system discharges annually around 6.6×10^{12} m^3 of freshwater into the ocean, accounting for 16% to 20% of the global continental water reserves [6]. Its climatic and topographic diversity is remarkable. The western portion is dominated by the Andes, which covers only 12% of the total area, but contribute with the most of sediments transported by the major rivers to the lowlands [7]. Rainfall also varies significantly, with annual rates ranging from 100 mm to 5.000 mm. Approximately 70% of the basin is covered by tropical rainforest, solidifying the Amazon Basin as one of the most biodiverse and globally significant ecosystem [8].

Gaussian functions have a wide range of applications across different fields due to their mathematical properties and ability to model real-world phenomena. In statistics, they are frequently used for fitting data distributions, allowing researchers to make inferences based on functional data [9]. In signal processing, Gaussian functions serve as essential tools for smoothing and filtering signals, for the multivariate curve resolution of non-bilinear data [10]. In machine learning, Gaussians are employed in algorithms such as

Gaussian Mixture Models (GMM), which allow estimation of the density of complex data sets [11]. Furthermore, Gaussian filters are commonly used in image processing to denoize digital images, demonstrating their importance in visual data analysis [12]. Finally, in environmental sciences the Gaussian functions are often used to model atmospheric pollutant dispersion [13]. The use of Gaussian components in hydrology is crucial for accurately modeling complex hydrological phenomena. This approach captures the spatial and temporal variability of hydrological data, which are influenced by various interacting factors. Combining Gaussians, researchers have modeled the surface water fraction from MODIS and Landsat images [14], precipitation [15], spatial distribution of watershed-scale hydrologic models [16], and other processes. This technique improves the predictions, but also helps to identify anomalies and assess the risks of extreme events, such as floods and droughts, contributing to efficient water resource management. Thus, Gaussians serve as a valuable tool for understanding and managing of hydrological systems.

In this study we apply multiple Gaussians to model the water level behavior of rivers in the Negro River basin. The initial approach involves combining data from the Serrinha station, located on the Negro River, and the Santa Maria do Boiaçú station, situated on the Branco River. This merge of information seeks to analyze and understand the behavior of the monitoring point at Moura, situated downstream of the confluence of these two rivers. The combination of gaussians can effectively serve to analyze the hydrological behavior of the region, providing a better understanding of the dynamics of both the rivers and how their interaction influences the water levels at Moura.

The paper is organized as follows, Sec. 2 presents the localization of the considered stations within the Negro River basin and the analyzed data. In Sec. 3 we provide an overview of the Gaussian fitting models, while in Sec. 4 we describe the statistical quantities considered in this paper. Sec. 5 is devoted to the results and their discussions. Finally in Sec. 6 we present our conclusions and some future perspectives.

2. Study Area

2.1. Negro River Basin and Fluviometric Stations

This study utilizes hydrological data from three fluviometric stations located within the Negro River basin: Santa Maria do Boiaçú (SMB), Serrinha, and Moura. The Negro River basin and the locations of these stations are depicted in Fig. 1. The elevation of the terrain, depicted by colors, was obtained from OpenTopography [17]. The hydrological dataset was sourced from the Brazilian hidrometric network: National Water and Basic Sanitation Agency (ANA) [18]. Detailed characteristics of these stations, such as data periods, drainage areas, and the rivers on which they are established, are summarized in Table 1.

Each station is under the administration of ANA and is managed by the Geological Survey of Brazil (SGB). Strategically located across various points in the Negro River basin, these stations provide important hydrological information. Their datasets play a fundamental role in the monitoring, evaluation, and management of water resources in the Negro River basin, contributing significantly to understanding regional hydrodynamics and supporting water resource planning and management strategies. ANA provides a wide range of hydrological data, including Water Level (WL), discharge, sediment transport, and water quality parameters, among others. However, in this study, the focus is exclusively on WL data.

Table 1: Data period of the measurements, drainage area, geographic coordinates, and the rivers corresponding to the three stations.

Station	**Data period**	**Drainage area (km^2)**	**River**	**Latitude**	**Longitude**
Serrinha	2000/01 - 2023/24	293,000	Negro	00°28'54"S	64°49'38"W
SMB	2000/01 - 2023/24	184,000	Branco	00°30'19"S	61°47'11"W
Moura	2000/01 - 2023/24	593,000	Negro	01°27'24"S	61°38'05"W

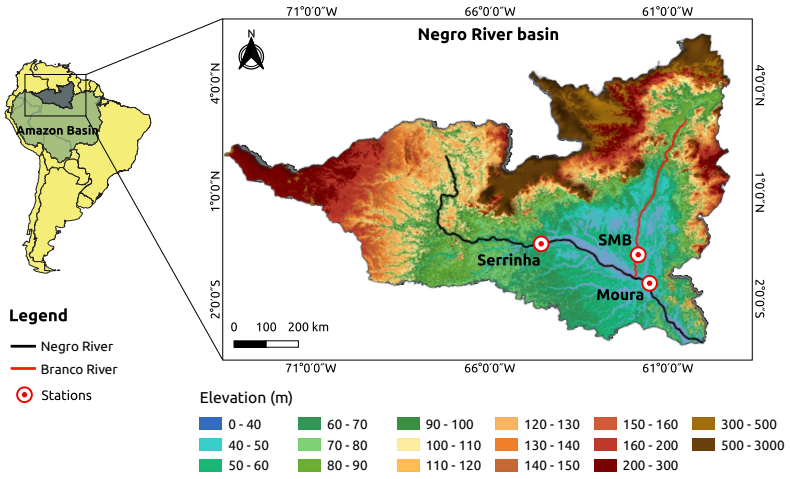


Figure 1: Negro River basin, highlighting the locations of the stations: Serrinha, Santa Maria do Boiaçú, and Moura. The map in the left displays the position of the Negro River basin as a sub-basin of the Amazon River basin within South America.

2.2. Hydrological Patterns and Data Integration

The relationship between Moura and the combined data from SMB and Serrinha is examined from a hydrological perspective. This combination is essential for understanding the dynamics of WL in Moura, justified by the proximity of the stations and the hydrological interactions that occur between the Negro and Branco Rivers post confluence [19]. The interplay of these rivers creates a complex hydrological system that influences the flow patterns and WL, making it fundamental to analyze the data from these stations. Typically, hydrological data are reported based on the calendar year (January to December). However, in this study, a hydrological year spanning from March of one year to February of the following year is adopted, denoted as $\{Y_n/Y_{n+1} : n = 2000, \ldots, 2023\}$, where Y_n represents the starting year and Y_{n+1} is the subsequent year. Thus, each hydrological year covers the whole year interval $\{(m, Y_n) \mid m = [3, 12]\} \cup \{(m, Y_{n+1}) \mid m = [1, 2]\}$, ensuring continuity with seasonal hydrological patterns. This approach allows for a more comprehensive representation of the seasonal cycle, ensuring that both the onset of the flooding process and the complete dry season recession are captured. By structuring the data in this manner, the analysis encompasses the full flood wave propagation and drought.

Figure 2(a) shows the combined WL from SMB and Serrinha (SMB + Serrinha), represented by a continuous blue curve. The period analyzed spans from 2000/01 to 2023/24. A detailed examination reveals notable similarities between the WL at Moura and the combined curve at various points (see Fig. 2(b)). However, an exact equivalence cannot be expected due to multiple influencing factors, including climatic variations such as rainfall distribution and intensity [20, 21, 22], geological characteristics of the basin [22, 23], and the presence of tributaries that contribute additional flow downstream from Serrinha and SMB. The hydraulic effect, in particular, plays a fundamental role, occurring when a river with higher discharge impounds the flow of a smaller tributary, leading to water accumulation in the tributary's basin and modifying flow velocity and WL in both rivers [24, 25]. Additionally, anthropogenic influences, such as land use changes and deforestation [26], can alter runoff patterns and sediment transport [27], further impacting the hydrological balance.

Figure 3 shows the variation of WL for the stations as a function of the months. The blue shaded region represents the WL variation over the last 24 years, from 2000/01 to 2023/24. The highlighted dashed curves correspond to the years 2022/23 (black curve), 2021/22 (orange curve), and 2020/21 (red curve). These

years are key for the analysis as they serve as examples for the application of Gaussian fitting, which is discussed later. Figure 3(a) presents the WL of SMB, Fig. 3(b) shows Serrinha, Fig. 3(c) represents SMB + Serrinha, and Fig. 3(d) depicts Moura. It is observed that, over the past 23 years, the three highlighted years tend to be positioned towards the upper part of the blue shaded region, indicating higher WL levels during these years. Figure 3 justifies the choice of starting in March, as it clearly delineates the typical hydrological cycle. From March to July, the WL exhibit a rising trend, reaching their peak between the months of June and July. This is followed by a recession phase from August to October, and finally, a drought period extending from October to February of the following year. This seasonal pattern highlights the importance of considering March as the starting point for data analysis, aligning with the natural dynamics of the WL variation.

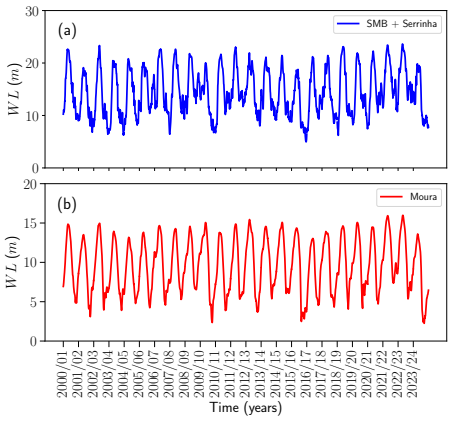


Figure 2: WL for fluviometric stations: (a) SMB + Serrinha and (b) Moura, considering the years from 2000/01 to 2023/24.

2.3. Comparative Analysis of Combined WL Data

A comparative analysis of the WL from SMB, Serrinha, and Moura stations is presented, with the aim of understanding the hydrological dynamics and exploring the relationships between these stations. Figure 4 illustrates the WL for the years 2022/23 (Fig. 4(a)), 2021/22 (Fig. 4(b)), and 2020/21 (Fig. 4(c)). The continuous cyan and orange curves represent the WL of SMB and Serrinha, respectively, while the green dashed curve corresponds to the combined sum of these two stations (SMB + Serrinha). The continuous red curve shows the data from the Moura station. This comparison highlights the seasonal variations in WL across the analyzed years and aims to assess how the combined sum of the water levels from SMB and Serrinha compares to the behavior observed at Moura.

The objective is to investigate whether this data combination can provide insights into the hydrodynamics of the Negro River basin, considering the complex interactions within the hydrological system, after the confluence of the Negro and Branco rivers. This combination is supported by the geographical proximity of the stations and the hydrological interactions between the rivers. Thus, the comparative analysis of these dynamics shows that, although the summed levels from SMB and Serrinha cannot faithfully reproduce the hydrodynamics at Moura, they exhibit many similarities and provide a valuable baseline for understanding certain aspects of the WL behavior in the basin.

Figure 5 employs boxplots to provide a more straightforward representation of the WL distribution at the stations. A boxplot is a statistical tool that visually summarizes data distribution through key percentiles. It displays the upper limit (maximum peak – flood) and the lower limit (minimum – drought), along with the first quartile (Q_1), third quartile (Q_3), and the median (orange line). The upper and lower limits of the

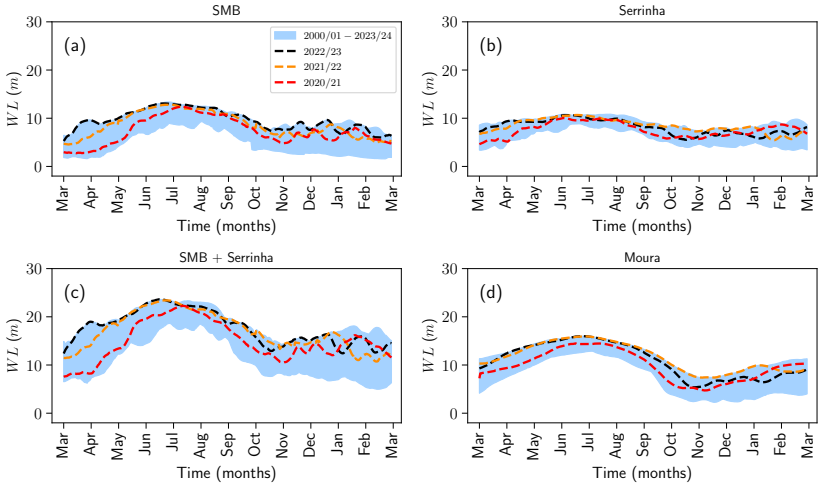


Figure 3: WL for the stations: (a) SMB, (b) Serrinha, (c) SMB + Serrinha, and (d) Moura. The blue-shaded region represents the variation from 2000/01 to 2023/24. Dashed curves highlight the years 2022/23 (black), 2021/22 (orange), and 2020/21 (red).

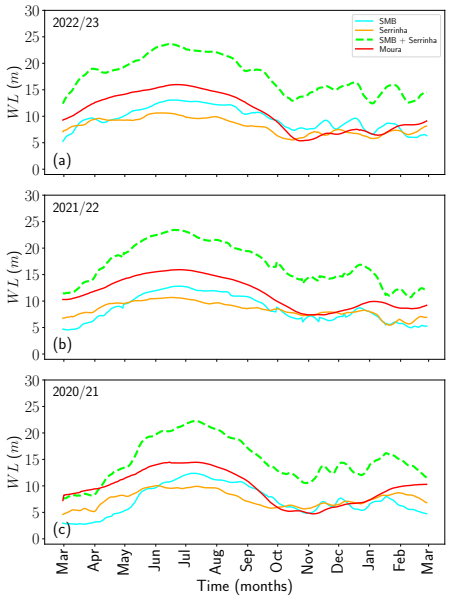


Figure 4: WL for the stations SMB, Serrinha, SMB + Serrinha, and Moura for the years (a) 2022/23, (b) 2021/22, and (c) 2020/21.

boxplots correspond to the maximum and minimum measured WL values, excluding potential outliers. The median represents the central tendency of the data, while the Q_1 and Q_3 indicate the interquartile range.

For SMB, a slight increase is observed in the last three analyzed years, as indicated by the median, upper, and lower limits. In contrast, Serrinha shows little variation over the years. However, for the summed levels of SMB + Serrinha, the dynamics of SMB prevail, as seen in the median and the upper and lower limits. The WL dynamics at Moura station is mostly influenced by Serrinha, given the similarities in median, Q_1, Q_3, and upper limits. The boxplot is a valuable statistical tool for illustrating WL dynamics without displaying all data points in a graph. Naturally, it can be applied to any parameter beyond WL.

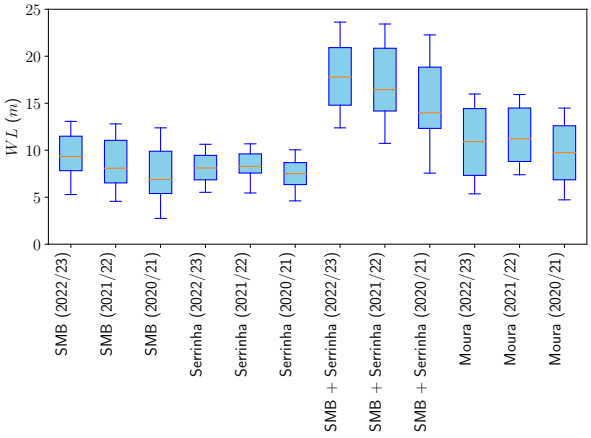


Figure 5: Boxplots representing the WL distribution for SMB, Serrinha, SMB + Serrinha, and Moura.

3. Gaussian Fitting Models

3.1. Multiple Gaussian Fitting with Optimized Components

Curve fitting is a common technique used to model complex patterns and has innumerous applications in data analysis [28]. Considering the range of available techniques, Gaussian functions are particularly useful for describing distributions with prominent peaks and symmetrical dispersions around a central value. In this study, a multiple Gaussian fitting model is employed to characterize the hydrological seasonal cycle in the Negro River basin. A Gaussian function is commonly defined as

$$f(x) = A \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \tag{1}$$

where A is the amplitude of the peak, μ is the mean (representing the central position of the curve), and σ is the standard deviation, which determines the width of the curve. The smooth and symmetric shape of the Gaussian function makes it particularly suitable for fitting the WL curve over the year, allowing for a precise representation of hydrological variations. This approach efficiently captures seasonal patterns and anomalies in the data, facilitating a deeper understanding of the hydrological dynamics.

For fitting multiple Gaussian components, a combination of $n = 3$ Gaussian functions is employed, represented as:

$$f(x) = \sum_{i=1}^{3} A_i \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right\}, \tag{2}$$

where A_i, μ_i, and σ_i are the parameters of the i-th Gaussian. This approach allows us to capture local variations and multiple peaks observed in the data, providing a detailed representation of phenomena with multiple seasonal patterns. In the fitting process, the positions of the Gaussian components were set as $\mu_1 = 100$ (days), $\mu_2 = 183$ (days), and $\mu_3 = 250$ (days), corresponding to key phases of the hydrological cycle. These values were chosen based on an initial visual inspection of the data and refined during the fitting process. The choice of $n = 3$ was sufficient for effectively reproducing the hydrological behavior in the considered stations, enabling robust modeling of the WL dynamics. However, there may be cases where additional Gaussian components are required to achieve a more accurate fitting, depending on the complexity of the hydrological patterns. In such cases, the model can be expanded by incorporating as many Gaussian components as needed to better capture the dynamics.

The initial parameters for each of the three Gaussian functions, namely amplitude, mean, and standard deviation, were set as an initial guess based on the data distribution. These initial mean values serve as starting points for the optimization process and do not represent their final locations. The optimization was then performed using the `curve_fit` function from the *SciPy* library [29], which iteratively adjusts the parameters to minimize the difference between the observed data and the fitted model. By fitting these three Gaussian components, the seasonal variations in WL can be modeled, with each Gaussian component representing a distinct phase. By utilizing non-linear curve fitting, it provides valuable insights into the temporal dynamics of the basin, helping to identify the hydrological cycle.

3.2. Gaussian Mixture Model

To model the distribution of hydrological data, the *Gaussian Mixture Model* (GMM) [29] is employed. This is a statistical technique effective in describing complex distributions, especially when the data can be represented as a combination of multiple Gaussian distributions. The GMM is a probabilistic approach that assumes data points originating from a combination of multiple Gaussian distributions with unknown parameters [29]. The GMM is an unsupervised learning technique used for clustering and density estimation. It assumes that the data is a mixture of multiple Gaussian distributions, each representing a different cluster. The GMM is described by the following probability density function

$$f(x) = \sum_{i=1}^{n} \pi_i \mathcal{N}(x|\mu_i, \sigma_i^2), \tag{3}$$

where $\mathcal{N}(x|\mu_i, \sigma_i^2)$ is the probability density of a normal distribution for the i-th component, with mean μ_i and variance σ_i^2, and π_i is the weight of each component in the mixture. The number n of components is determined by minimizing the *Bayesian Information Criterion* (BIC) [30].

The `GaussianMixture` class from the *scikit-learn* library was used for fitting the GMM [29]. This class utilizes the Expectation-Maximization (EM) algorithm to estimate the parameters of the mixture without the need for labeled data, including the means, variances, and weights. BIC scores were calculated for different values of n (the number of components), and the model with the lowest BIC score was selected. In hydrological analyses, GMM proves to be a valuable tool for identifying distinct phases or patterns in WL, helping to model seasonal variations, detect anomalies, and classify different hydrological regimes, all without prior knowledge of the system's behavior. Its ability to handle complex, multimodal data distributions makes it particularly suitable for the analysis of hydrological time series, where data often exhibit multiple underlying processes.

The number of components, n, is chosen based on BIC, given by

$$BIC = -2 \log L + k \log N, \tag{4}$$

where L is the likelihood function of the model, k is the number of model parameters (number of components multiplied by the parameters of each component), and N is the total number of observations. The value of n that minimizes BIC is selected as the best-fitting model for the data. By automatically selecting the number of components using BIC, the model ensures that the complexity is appropriate for the data.

The fitting process starts with an initial estimation of the parameters for each Gaussian component, which are refined through an optimization procedure. This iterative process uses the EM algorithm, which alternates between two steps:

- Expectation (E-step): Calculates the probability that each data point belongs to each Gaussian component, given the current parameters;
- Maximization (M-step): Updates the parameters of the mixture (weights, means, and variances) to maximize the likelihood based on the probabilities calculated in the E-step.

4. Statistical Metrics

4.1. Distance Correlation Coefficient

The *Distance Correlation* (DC) measures the dependence between random vectors using Euclidean distances, assuming a value of zero only when the vectors are completely independent [31, 32, 33, 34]. It provides a more general measure of dependence than traditional correlation coefficients, as it can capture both linear and nonlinear relationships between variables. Consider a joint sample $(\boldsymbol{X}, \boldsymbol{Y}) = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, the matrice S_{ij} is constructed as follows. For the sample data $\boldsymbol{X}$

$$S_{ij} = s_{ij} - \bar{s}_{i.} - \bar{s}_{.j} + \bar{s}_{..} \tag{5}$$

where $s_{ij} = |x_i - x_j|$ is a matrix formed by the Euclidean distance between the pairs of sample elements, $\bar{s}_{i.} = \frac{1}{n}\sum_{j=1}^{n} s_{ij}$ is the row arithmetic mean, $\bar{s}_{.j} = \frac{1}{n}\sum_{i=1}^{n} s_{ij}$ is the column arithmetic mean, and $\bar{s}_{..} = \frac{1}{n^2}\sum_{i,j=1}^{n} s_{ij}$ is the overall mean of the matrix formed by the Euclidean distances s_{ij}. The matrix G_{ij} for the sample $\boldsymbol{Y}$ is obtained similarly to S_{ij}.

The distance covariance is calculated as

$$\sigma_n^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n^2}\sum_{i,j=1}^{n} S_{ij}G_{ij}. \tag{6}$$

and the terms $\sigma_n(\boldsymbol{X})$ and $\sigma_n(\boldsymbol{Y})$ represent the distance variances of the samples $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$\sigma_n^2(\boldsymbol{X}) = \frac{1}{n^2}\sum_{i,j=1}^{n} S_{ij}^2, \text{ and } \sigma_n^2(\boldsymbol{Y}) = \frac{1}{n^2}\sum_{i,j=1}^{n} G_{ij}^2. \tag{7}$$

The DC coefficient is given by

$$DC(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\sigma_n(\boldsymbol{X}, \boldsymbol{Y})}{\sqrt{\sigma_n(\boldsymbol{X})}\sqrt{\sigma_n(\boldsymbol{Y})}}. \tag{8}$$

4.2. Coefficient of Determination, Mean Squared Error and Mean Absolute Error

The quality of regression models is often assessed using the *Coefficient of Determination* (R^2), the *Mean Squared Error* (MSE), and the *Mean Absolute Error* (MAE). These metrics provide insights into how well a model fits the observed data. The R^2 quantifies the proportion of the variance in the observed data that is explained by the model [35]. It is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{9}$$

where x_i are the observed values, y_i are the fitted values, $\bar{x}$ is the mean of the observed values, and n is the number of observations. The values of R^2 range from $-\infty$ to 1. A value of $R^2 = 1$ indicates a perfect fit. When R^2 is negative, it implies that the model performs worse than simply using the mean of the observed values as a predictor.

The MSE measures the average squared difference between the observed and fitted values. It is given by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2. \tag{10}$$

A smaller MSE indicates higher accuracy. The MAE is another widely used metric that calculates the average absolute differences between observed and fitted values

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|. \tag{11}$$

Unlike MSE, which penalizes larger errors more heavily due to squaring, MAE treats all errors equally, making it a useful metric when outliers are present.

5. Results

5.1. Analysis of DC Between Stations

Figure 6 presents the heatmaps of DC between the stations from 2000/01 to 2023/24. The correlations are computed considering all years with each other, and the main diagonal of the DC matrix corresponds to the maximum value ($DC = 1.0$), as it represents the correlation of each year with itself. In Fig. 6(a), the DC between SMB + Serrinha and itself is depicted, revealing a generally strong correlation. Most correlations exceed 0.8, as indicated by the color scale, with redder hues representing higher values. Several regions exhibit strong correlations, such as the period from 2000/01 to 2004/05, as well as the years 2006/07 and 2007/08 in relation to 2000/01–2004/05. A similar pattern is observed between 2021/22 and 2023/24. Other correlated periods can also be identified in the figure. Years with moderate correlations are highlighted in blue, with DC values ranging between 0.55 and 0.75. These include 2005/06, 2008/09, 2009/10, and 2012/13, which show lower similarity to other years. However, only a few years exhibit consistently moderate correlations. Overall, the heatmap indicates relatively high DC values across the dataset, as evidenced by the predominant red tones, which reflect stronger correlations between the WL of different years.

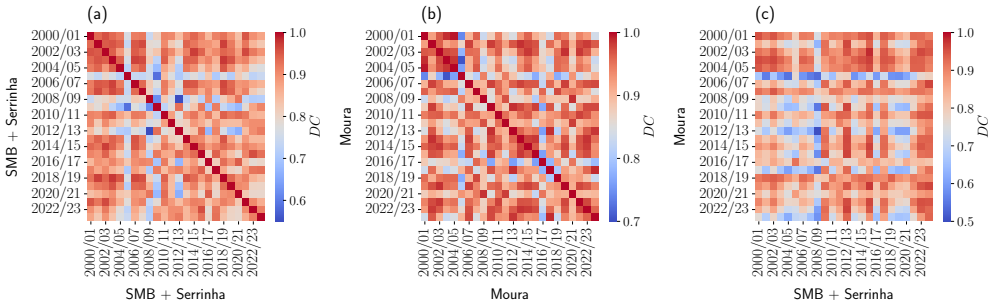


Figure 6: Heatmaps of DC between: (a) SMB + Serrinha with itself, (b) Moura with itself, and (c) Moura with SMB + Serrinha.

Figure 6(b) illustrates the DC of Moura with itself. The correlations in Moura are notably higher than those observed in Fig. 6(a), as the lowest DC values remain around 0.7. High-correlation regions similar to those of previous case also appear in this figure, with the most prominent ones occurring between 2000/01 and 2004/05, as well as from 2013/14 to 2015/16. Other regions with more intense red tones can be observed. As expected, the main diagonal in the matrix represents the maximum DC values. Some years, such as 2005/06, 2008/09, and 2016/17, exhibit moderate correlations, highlighted in blue colors, with DC values

ranging between 0.7 and 0.85. These results suggest that in Moura, the WL remains more stable over the years during the hydrological seasonal cycle, unlike in SMB + Serrinha, where greater fluctuations are observed.

Finally, Fig. 6(c) presents the DC between Moura and SMB + Serrinha. The DC values range from 0.5 to 1.0, with a predominance of red tones, indicating correlations mostly between 0.8 and 1.0. The high-correlation regions observed in the previous figures persist here. For instance, the years 2000/01 to 2004/05 in Moura exhibit strong correlations with nearly all years of SMB + Serrinha. Similarly, the years 2021/22 to 2023/24 in SMB + Serrinha show high correlation with all years of Moura. Other regions of strong correlation can also be identified in the figure. Conversely, some years in Moura exhibit lower correlation with SMB + Serrinha, including 2005/06, 2008/09, 2009/10, 2012/13, 2017/18, and 2023/24. Meanwhile, the years in SMB + Serrinha that show lower correlation with Moura are 2004/05, 2005/06, 2008/09, 2016/17, 2019/20, and 2020/21, with DC values ranging between 0.5 and 0.75. This still indicate a moderate correlation. Figure 6(c) highlights the overall strong hydrological correlation between Moura and SMB + Serrinha, reinforcing the interdependence of their WL variations over time.

5.2. Application of Multiple Gaussian Fitting

Figure 7 presents the fitting using three optimized Gaussian components, as described in subsection 3.1. This figure illustrates how only three individual Gaussian components, when summed, best represent the observed WL data. These components are highlighted with solid curves: blue (1st Gaussian), magenta (2nd Gaussian), and red (3rd Gaussian). The resulting fitting, corresponding to the sum of these components, is represented by the dashed black curve. In Fig. 7, only the summed case is considered for SMB + Serrinha (solid lime curve). However, this approach can be applied to any WL dataset from any hydrometric station. Subsequently, its application will be demonstrated to the Moura. To the right of the fitting plots, scatter plots compare the observed data with the fitted Gaussian values. These plots also display statistical metrics for error and correlation assessment, including R^2, MSE, MAE, and the function ($y \propto ax + b$) of the best linear fit between the datasets.

Figure 7(a) illustrates the results for the year 2022/23. The first Gaussian component accounts for the initial oscillation (repiquete) observed between March and April. The second Gaussian, provides the best fit for the flood process occurring from May to August. As the water recession begins, the third Gaussian takes over, fitting the period from August to February of the following year. The dashed black curve, representing the sum of the three components, provides the overall best fit to the observed data. Between November and February, some oscillations occur, and the model smooths over these variations. Figure 7(b) presents the scatter plot between the Gaussian fitted data and the observed data. It can be observed that the points in the figure follow a linear trend, with the fitting represented by the black dashed line. The R^2 metric shows a value of 0.98, with MSE equal to 0.74 and MAE equal to 0.65.

Figure 7(c), corresponding to the year 2021/22, shows that the first Gaussian fits the entire flood period, from March to August. The second Gaussian represents a small portion of the fit during the water recession period, from August to October. The third Gaussian fits the entire drought period, from October to February. As in the previous case, the drought period shows some oscillations, and the fit smooths over them. However, the dashed black curve that represents the fitting is very well aligned with the actual data. Figure 7(d) presents the statistical metrics in the scatter plot. The value of R^2 shows a result of 0.96, with MSE and MAE providing reasonable values. The best linear fit between the data is represented by the dashed curve.

Finally, Fig. 7(e) presents the WL of the year 2020/21. Similarly in Fig. 7(a), the first Gaussian component captures a slight oscillation occurring between March and April. The second Gaussian accurately fits the flood process, covering the water level rise and part of the recession, spanning from April to August. From August to February, which corresponds to the remainder of the recession and the low-water period, the third Gaussian component provides the best fit. Once again, the model adjusts smoothly to the oscillations that occur during the drought season. Statistical metrics comparing the fitted data to the WL for this year are shown in Fig. 7(f). For the three examples presented in Fig. 7, this case exhibits the best performance, as it shows the highest R^2 value and the lowest MSE and MAE. This indicates that the Gaussian model provides a highly accurate fit.

A notable finding from Fig. 7 is that only three Gaussian components can serve as an effective model for flood forecasting in Amazon rivers. The flood cycle, which comprises the rising phase, peak, and recession, can be well represented by Gaussian functions. This approach not only allows for flood peak predictions but also provides a reliable model for future scenario analysis. With data from the rising phase, the Gaussian model can be fitted to the trend and extrapolated into the future, offering valuable insights for flood management and preparedness in the region.

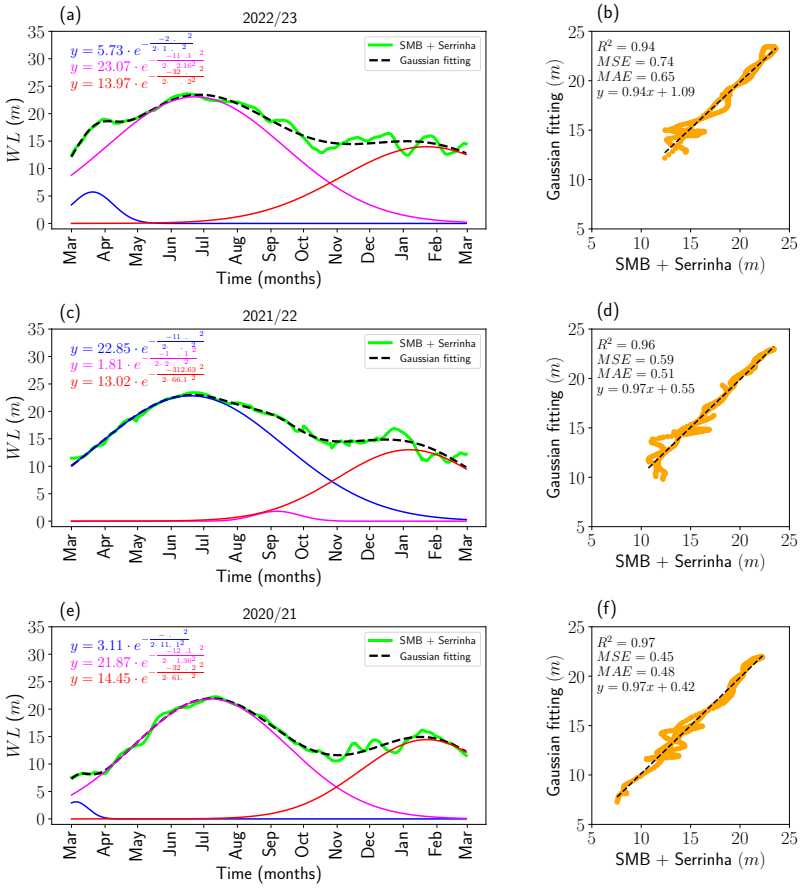


Figure 7: Gaussian fitting in WL for SMB + Serrinha during the years (a) 2022/23, (c) 2021/22, and (e) 2020/21. The corresponding scatter plots in (b), (d), and (f) illustrate the fitted versus observed values and the statistical metrics.

Table 2 presents the fitted parameters for the years 2022/23, 2021/22, and 2020/21, as shown in Fig. 7. These parameters represent the Gaussian components that were optimized to best fit the observed WL data for each year. The table includes the mean (μ_i), standard deviation (σ_i), and amplitude (A_i) for each of the Gaussian components, which are key to capturing the hydrological dynamics of the flood and drought processes in the Negro River basin. By comparing these fitted parameters across different years, it is possible to observe trends and variations in the hydrological regime.

Table 2: Fitted parameters of the three Gaussian components, given in Eq. (2), for the SMB + Serrinha.

Years	A_i	μ_i	σ_i
2022/23	5.73	20.50	18.97
	23.07	115.14	82.16
	13.97	328.47	78.82
2021/22	22.85	110.87	85.90
	1.81	190.51	20.90
	13.02	312.63	66.15
2020/21	3.11	5.09	11.51
	21.87	129.18	71.36
	14.45	328.72	61.70

5.3. Generalization of Multiple Gaussian Fitting

As a generalization of the model, three Gaussian components are now applied to fit WL for the SMB + Serrinha and Moura from 2000/01 to 2023/24, covering a total of 24 years of hydrological data. Figure 8 presents the heatmaps corresponding to the hydrological regimes of these stations. The WL values are normalized through the color scale to enhance comparability across different periods. Red shades indicate drought periods, while blue shades represent flood periods. Intermediate colors, such as yellow, green, and cyan, correspond to transition phases between these extremes. In Fig. 8(a), the heatmap for SMB + Serrinha is shown. It can be observed that drought periods persist between March and April, particularly pronounced between the years 2000/01–2005/06 and 2014/15–2021/23. The flood period occurs predominantly from May to September, while drought conditions prevail from October to February.

Applying the three Gaussian fitting to each year yields the best result, which is plotted in the heatmap shown in Fig. 8(b). A strong similarity can be observed between the real and fitted data, Fig. 8(a) and Fig. 8(b), respectively. If the short-term fluctuations in the real data, known as repiquetes, are interpreted as "noise", they are effectively smoothed in Fig. 8(b). As previously discussed in subsection 5.2, the Gaussian fitting provides a robust approximation, capturing the main hydrological patterns while attenuating the influence of small oscillations. Repiquetes are characterized by abrupt and unexpected shifts in river level variations [36]. Changes in WL are associated with health concerns like malaria, impacting the seasonal dynamics of the disease and playing a role in its incidence [37].

To further demonstrate the generalization of the model, WL from Moura is also analyzed over the same time period. In Fig. 8(c), minimal to negligible drought is observed between March and April. Instead, a transition period is evident, indicating rising water levels. The flood period extends from May to September, followed by the recession phase starting in September. From October to February, the dry season is predominant. As observed in Figs. 8(a) and 8(b), there is a strong similarity between the adjustments of the Gaussian component and the real data. Since the hydrological regime at Moura is more stable, with fewer fluctuations, the heatmap in Fig. 8(d) closely resembles the real data, further supporting the generalization of the model. Given the proximity of these stations, the same procedure can be applied. However, it is important to highlight that for stations in other regions of the Amazon basin, additional Gaussian components might be required to achieve a more accurate fitting, as hydrological regimes vary significantly across different stations.

Figure 9 presents the fitted parameters of the three Gaussian components for each year, specifically for the case shown in Fig. 8(b). In Fig. 9(a), the A_i are displayed, where the blue color represents the first Gaussian, orange corresponds to the second, and green denotes the third. The same color scheme is applied in Fig. 9(b) for the $|\mu_i|$ and in Fig. 9(c) for the σ_i values. All plots are shown on a semi-logarithmic scale. In Figs. 9(a) and 9(b), the parameters exhibit an almost uniform distribution across the years, except for 2016/17, where the first Gaussian shows significantly higher values. This variation may be attributed to an unusual hydrological event. In particular, the occurrence of significant repiquetes at the beginning or end of the series forces the Gaussian fitting to adjust in a way that increases the amplitude or other parameters; however, this does not pose an issue. In order to properly accommodate the variations a very high peak

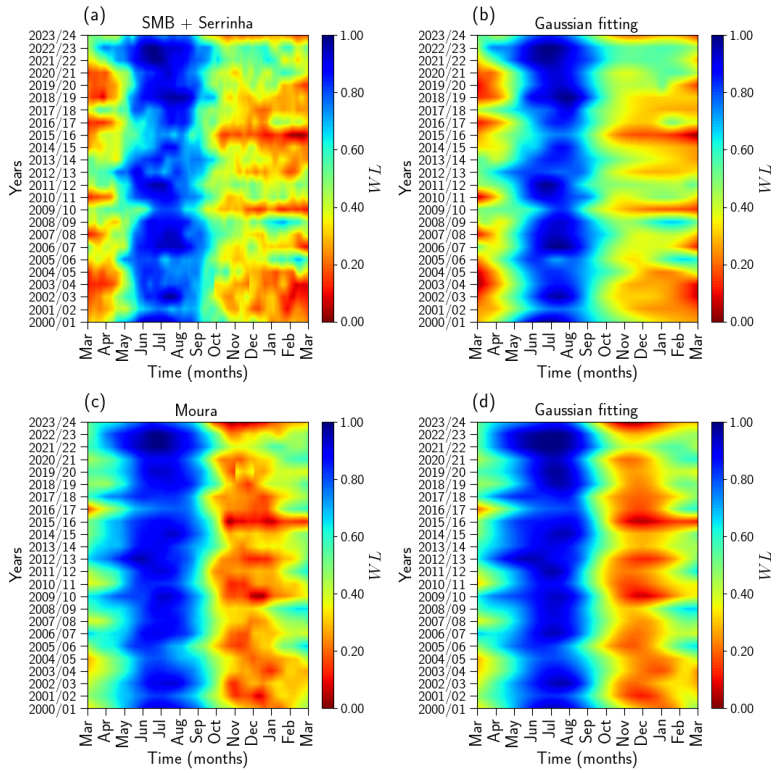


Figure 8: Heatmaps of normalized WL obtained from real data between 2000/01 and 2023/24 for (a) SMB + Serrinha and (c) Moura. The corresponding heatmaps obtained through Gaussian fitting are shown in (b) and (d).

appears for the amplitude. In Fig. 9(c), values of σ_i show some variations but generally follow a consistent pattern, likely reflecting the natural variability of seasonal transitions while maintaining the underlying structure of the hydrological cycles.

All statistical metrics defined in Section 4, such as, DC, R^2, MSE, and MAE, are computed by considering as two datasets the observed data and the Gaussian fitting of Fig. 8. Fig. 10 presents these results, confirming the correlation between the model outputs and the observed data. Figs. 10(a)–(d) correspond to the WL of SMB + Serrinha versus Gaussian fitting, while Figs. 10(e)–(h) represent the WL of Moura versus Gaussian fitting. In all heatmaps in Fig. 10, the main diagonal represents the metric for each station year versus its corresponding fitted year. In particular, for DC and R^2, a maximum correlation is observed, with values close to 1, indicated by a dark red coloring. In contrast, for MSE and MAE, the values reach zero, represented by dark blue. These diagonals clearly illustrate how closely the fitted data align with the observed values at the stations.

A strong correlation is observed when we compare the real data of WL to the Gaussian fitting, for every year. In Figs. 10(a) and 10(b), the heatmaps predominantly feature red tones. The DC values range from 0.6 to 1, indicating a moderate to strong correlation between the series. Years such as 2005/06, 2008–2009/10, and 2012/13 exhibit moderate correlation, while the remaining years show high correlation. For R^2, the

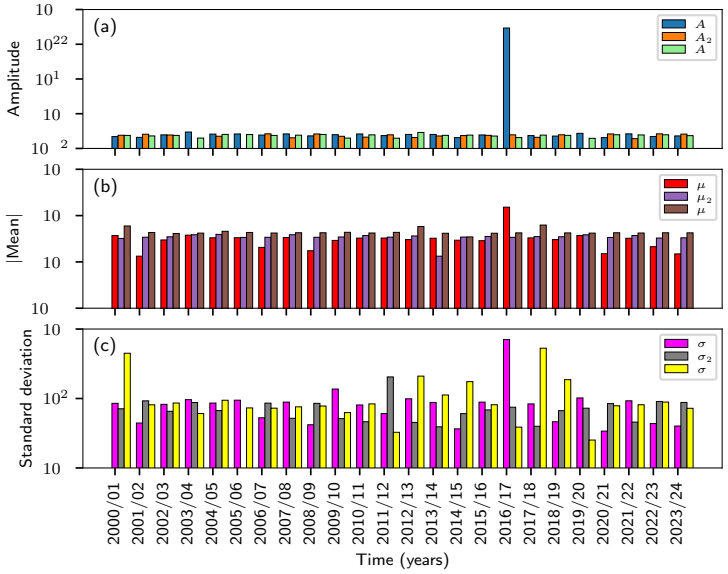


Figure 9: Semi-logarithmic plots of (a) A_i, (b) $|\mu_i|$, and (c) σ_i from 2000/01 to 2023/24 for SMB + Serrinha, Fig. (8)(b). The colors represent the three Gaussian components in sequence from left to right.

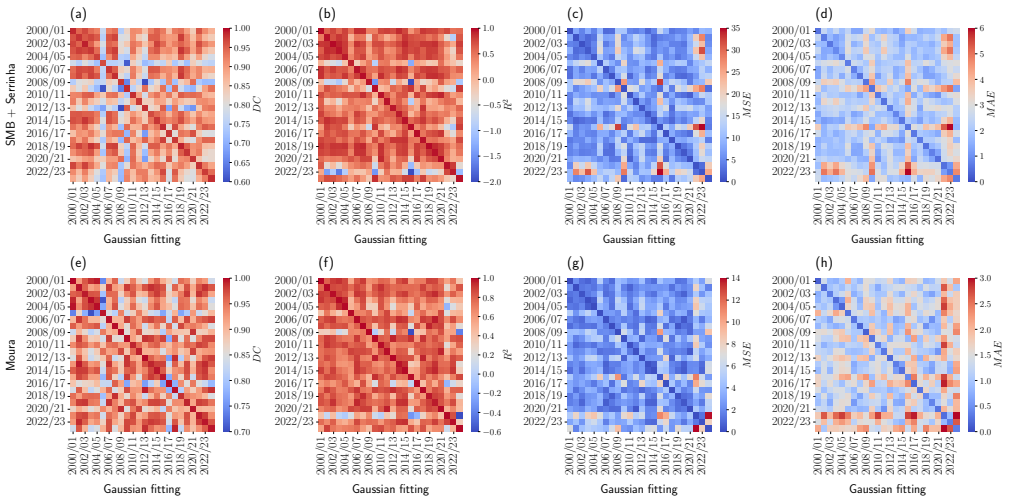


Figure 10: Statistical metrics comparing WL and fitted data for SMB + Serrinha and Moura: (a, e) DC, (b, f) R^2, (c, g) MSE, and (d, h) MAE.

years 2005/06, 2008/09, and 2022/23 display negative values, while all others are close to 1. In Figs. 10(c) and 10(d), the values remain close to zero, except for the years 2008–2009/10, 2015/16, and 2021–2022/23, which show higher MSE and MAE values. However, both figures exhibit similar patterns, with blue tones predominating.

For Figs.10(e) and 10(f), the observed behavior closely mirrors the previous case. The values of DC and R^2 are nearly 1, with red tones predominant in the figures. In Fig.10(e), only the years 2005/06, 2008/09, 2016/17, and 2020/21 show correlations close to 0.7, though still indicating a strong correlation. In Fig.10(f), the years 2005/06, 2008/09, and 2021/22 display values approaching zero or even negative, suggesting weaker relationships. For Fig.10(g), which represents MSE, the years 2021/22 and 2023/24 exhibit significantly higher values, while all other years are marked in blue, indicating minimal error. Finally, Fig.10(h) is similar to Fig.10(d), with the later years, 2021–2023/24, showing the highest values of MAE. We conclude that in addition to the strong correlation between observed and fitted data for each year, there is also a robust correlation across all others.

5.4. GMM for Fitting WL Distributions

In this subsection, the GMM is applied to fitting the distribution of the SMB + Serrinha and Moura data from 2000/01 to 2023/24. The number of components used in the GMM is automatically determined, following the procedure outlined in subsection 3.2. This approach ensures an objective and consistent modeling of the data distributions throughout the analyzed years, providing a clearer understanding of the data structure and a reliable basis for comparisons between observed and estimated distributions.

In Fig. 11(a), the probability density of the WL distributions for SMB + Serrinha is shown over 24 years, from 2000/01 to 2023/24. This figure reveals two distinct regions: the first, between 5 and 17.5 m, represents the larger group corresponding to the drought period, and the second, from 17.5 to just above 22.5 m, corresponds to the flood period. It is important to note that the distribution WL becomes a valuable tool, as it clearly delineates the different periods of the hydrological regime. In this case, the GMM automatically chose $n = 6$ Gaussians for the best fitting of the distribution. The best fitting is shown by the solid red curve, while each Gaussian components are represented by a dashed curve, with colors corresponding to the highlighted legend. Two distinct data clusters are clearly observed.

In Fig. 11(b), the probability density for Moura is shown. In this figure, two predominant regions are observed: one between 2 and 8 m, representing the drought period, and another between 12 and 16 m, representing the flood period. In addition, there is a notable region between 8 and 12 m, highlighting the transition period between flood and drought, and vice versa. For Moura, the GMM is adjusted by $n = 5$ Gaussian components. The continuous blue curve represents the best fitting of the distribution. In this case, the use of GMM also clearly highlights the different data clusters. This fitting clearly distinguishes the different hydrological periods and transitions within the data.

In Fig. 12, the GMM was fitted independently for each year, automatically selecting two Gaussian components to model the seasonal distribution of WL. The evolution of the mean values of these components over time can be seen in the figures, which highlight distinct hydrological periods. The mean values of the Gaussian components are obtained by fitting the GMM to WL for each year, where the model estimates the parameters of two Gaussian distributions, and the mean of each component corresponds to the central value around which the data clusters, in this case the flood and drought periods. The results illustrate the capacity of GMM to identify seasonal variations and long-term trends in river height dynamics, offering insights into hydrological regime shifts.

Figures 12(a) and 12(b) illustrate the mean values representing the flood (μ_1) and drought (μ_2) periods for SMB + Serrinha and Moura, respectively. The blue curve with points corresponds to flood periods, capturing peak water levels. It is important to note that these values do not necessarily match exactly the observed measurements but provide a close approximation of the WL. This makes the GMM-based approach a valuable tool for identifying hydrological regimes within the basin's rivers. Similarly, the orange curve with points represents the minimum values of drought periods. In both figures, the variability and similarity between the upper and lower curves are evident. For exemplo, the minimum points in 2015/16 and 2023/24 correspond to extreme droughts, while 2021/22 stands out as a year of extreme flooding. The

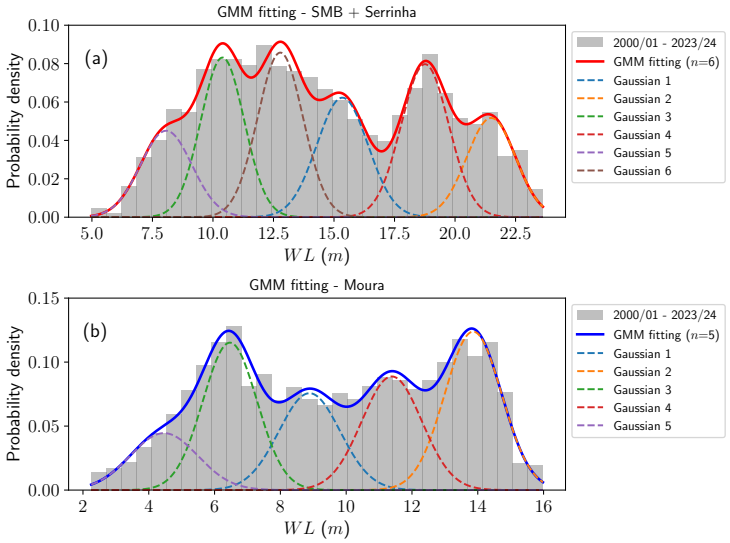


Figure 11: Probability density of the distribution fitted by GMM for: (a) SMB + Serrinha and (b) Moura. The legend highlights the sequence of Gaussian components by color.

black points in Fig. 12(b) represent the values observed at the Moura station, highlighting their proximity to the mean values of the GMM. These estimated means effectively highlight both typical and extreme hydrological events and make GMM a good tool also for predicting such occurrences.

6. Conclusions

This study applied multiple Gaussians to analyze the hydrological regime of the Negro River basin, focusing on the stations of Serrinha, SMB, and Moura. Through the combination of statistical and machine learning techniques (GMM), key hydrological patterns were identified, allowing for the characterization of seasonal variations and long-term trends. The results indicate that the fitting models adequately represent WL dynamics, effectively capturing the variability between flood and drought periods. This approach provides a structured framework for examining hydrological behaviors in complex river systems.

The applied models allowed the quantification of similarities and differences in hydrological trends across different locations, highlighting regional patterns in water level fluctuations. These findings contribute to a more comprehensive understanding of the hydrological processes of the basin, supporting strategies for the management of water resources. The methodology presented can be extended to other subbasins of the Amazon basin with similar characteristics, offering a flexible and data-driven approach for hydrological studies.

Future applications of this framework include its extension to other hydrological variables, such as precipitation, discharge, and sediment transport, to improve regional hydrological assessments. Furthermore, the potential of unsupervised learning techniques, such as GMM, to forecast future variations WL should be further explored by integrating predictive modeling approaches to anticipate extreme hydrological events and support proactive water resource management.

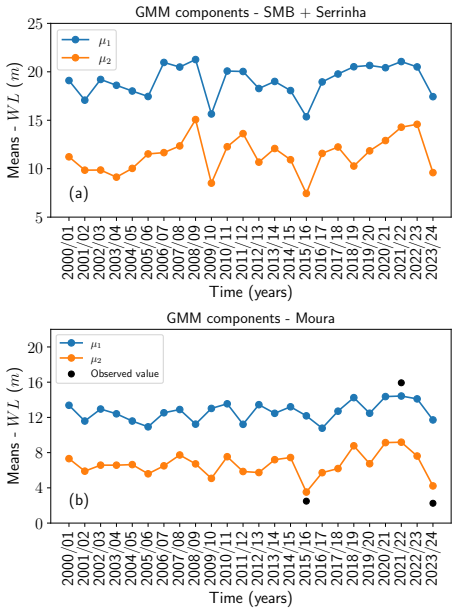


Figure 12: Means (μ_1, μ_2) in GMM components to characterize the flood and drought periods for (a) SMB + Serrinha and (b) Moura.

Acknowledgements

C.F.O.M. and M.G. thank the National Council for Scientific and Technological Development – CNPq (Brazilian agency) for financial support (Grant Numbers 151896/2024-0 and 311428/2022-3).

Disclosure statement

No potential conflict of interest was reported by the authors.

Availability of data

Data will be made available on reasonable request.

References

[1] Rogério Ribeiro Marinho, Paulo Rodrigo Zanin, and Naziano Pantoja Filizola Junior. The negro river in the anavilhanas archipelago: Streamflow and geomorphology of a complex anabranching system in the amazon. *Earth Surface Processes and Landforms*, 47(4):1108–1123, 2022. doi:10.1002/esp.5306.

[2] Rogério Ribeiro Marinho, Naziano Pantoja Filizola Junior, and Édipo Henrique Cremon. Analysis of suspended sediment in the anavilhanas archipelago, rio negro, amazon basin. *Water*, 12(4), 2020. doi:10.3390/w12041073.

[3] Rogério Ribeiro Marinho, Albert Reis Furtado, Vanessa Cristina Dos Santos, André Zumak Avezedo Nascimento, and Naziano Pantoja Filizola Junior. Riverbed morphology and hydrodynamics in the confluence of complex mega rivers - a study in the branco and negro rivers, amazon basin. *Journal of South American Earth Sciences*, 118:103969, 2022. doi:10.1016/j.jsames.2022.103969.

[4] Alain Laraque, Jean Loup Guyot, and Naziano Filizola. Mixing processes in the amazon river at the confluences of the negro and solimões rivers, encontro das Águas, manaus, brazil. *Hydrological Processes*, 23(22):3131–3140, 2009. doi:10.1002/hyp.7388.

[5] M. M. T. Reis, D. S. Ladislau, M. W. S. Ribeiro, C. C. Guimarães, A. J. V. Paiva, D. C. Mattos, P. H. R. Aride, and A. T. Oliveira. Socioeconomic aspects and profile of fishing according to fishers of commercial edible fish in the municipality of barcelos, middle negro river, amazonas, brazil. *Brazilian journal of biology*, 82:1–14, 2022. doi:10.1590/1519-6984.264210.

[6] Michel Molinier, Jean-Loup Guyot, E. de Oliveira, and V. Guimaraes. Les régimes hydrologiques de l'Amazone et de ses affluents. In *L'hydrologie tropicale : géosciences et outil pour le développement : mélanges à la mémoire de Jean Rodier*, Publication - AISH, pages 209–222. AISH, Wallingford, 1996. URL: https://www.documentation.ird.fr/hor/fdi:010009811.

[7] Naziano Filizola and Jean Loup Guyot. Suspended sediment yields in the amazon basin: an assessment using the brazilian national data set. *Hydrological Processes*, 23(22):3207–3215, 2009. doi:10.1002/hyp.7394.

[8] Naziano Filizola and Jean-Loup Guyot. O fluxo de sedimentos em suspensão nos rios da amazônia brasileira. *Revista Brasileira de Geociencias*, 41:566–576, 12 2011. URL: https://www.documentation.ird.fr/hor/fdi:010070669.

[9] Petr Čoupek, Viktor Dolník, Zdeněk Hlávka, and Daniel Hlubinka. Fourier approach to goodness-of-fit tests for gaussian random processes. *Statistical Papers*, 65(5):2937–2972, 2024. doi:10.1007/s00362-023-01510-4.

[10] Santiago Cavanillas, José Manuel Díaz-Cruz, Cristina Ariño, and Miquel Esteban. Parametric signal fitting by gaussian peak adjustment: A new multivariate curve resolution method for non-bilinear voltammetric measurements. *Analytica Chimica Acta*, 689(2):198–205, 2011. doi:10.1016/j.aca.2011.01.017.

[11] Weiguo Lu, Deng Ding, Fengyan Wu, and Gangnan Yuan. An efficient gaussian mixture model and its application to neural networks. *Knowledge-Based Systems*, 310:112942, 2025. doi:10.1016/j.knosys.2024.112942.

[12] Mehdi Mafi, Harold Martin, Mercedes Cabrerizo, Jean Andrian, Armando Barreto, and Malek Adjouadi. A comprehensive survey on impulse and gaussian denoising filters for digital images. *Signal Processing*, 157:236–260, 2019. doi:10.1016/j.sigpro.2018.12.006.

[13] Joel B. Johnson. An introduction to atmospheric pollutant dispersion modelling. *Environmental Sciences Proceedings*, 19(1), 2022. doi:10.3390/ecas2022-12826.

[14] Jiayong Liang and Desheng Liu and. Automated estimation of daily surface water fraction from modis and landsat images using gaussian process regression. *International Journal of Remote Sensing*, 42(11):4261–4283, 2021. doi:10.1080/01431161.2021.1892859.

[15] Vasiliki D. Agou, Andrew Pavlides, and Dionissios T. Hristopulos. Spatial modeling of precipitation based on data-driven warping of gaussian processes. *Entropy*, 24(3), 2022. doi:10.3390/e24030321.

[16] Tessa Maurer, Francesco Avanzi, Carlos A. Oroza, Steven D. Glaser, Martha Conklin, and Roger C. Bales. Optimizing spatial distribution of watershed-scale hydrologic models using gaussian mixture models. *Environmental Modelling Software*, 142:105076, 2021. doi:10.1016/j.envsoft.2021.105076.

[17] OpenTopography. OpenTopography - High-Resolution Topography Data and Tools, 2025. Accessed: March 22, 2025. URL: https://opentopography.org/.

[18] ANA. Agência nacional de águas e saneamento básico, 2024. URL: https://www.gov.br/ana/pt-br.

[19] Rogério Ribeiro Marinho, Albert Reis Furtado, Vanessa Cristina Dos Santos, André Zumak Avezedo Nascimento, and Naziano Pantoja Filizola Junior. Riverbed morphology and hydrodynamics in the confluence of complex mega rivers - a study in the branco and negro rivers, amazon basin. *Journal of South American Earth Sciences*, 118:103969, 2022. doi:10.1016/j.jsames.2022.103969.

[20] J.-L. Guyot, M. Molinier, V. Guimarães, K. J. Cudo, and E. Oliveira. Balanço hídrico da bacia do rio negro. *Anais do X Simpósio Brasileiro de Recursos Hídricos, Gramado, 11p*, 1993. URL: https://www.documentation.ird.fr/hor/fdi:39480.

[21] Assessment of different precipitation datasets and their impacts on the water balance of the negro river basin. *Journal of Hydrology*, 404(3):304–322, 2011. doi:10.1016/j.jhydrol.2011.04.037.

[22] Cíntia L. Eleutério, Carlos F. O. Mendes, Marcus W. Beims, Mircea Galiceanu, and Naziano P. Filizola. Multi-station data integration for flood and drought prediction in the negro river basin using convolutional neural networks. *Submited for publication*, 2025.

[23] Túlio A. A. Mendes, Rielva S. C. Nascimento, Renata S. Veras, Marcelo E. Almeida, and Luiz G. Knauer. Geological knowledge advances on the alto rio negro region, northwestern amazonian craton, brazil: a review. *Journal of the Geological Survey of Brazil*, 4(3):209–222, 2021. doi:10.29396/jgsb.2021.v4.n3.2.

[24] Robert H. Meade, José M. Rayol, Sylvio C. Da Conceição, and José R. G. Natividade. Backwater effects in the amazon river basin of brazil. *Environmental Geology and Water Sciences*, 18(2):105–114, 1991. doi:10.1007/BF01704664.

[25] Luan Ferreira Siqueira and Naziano Filizola. Estudo hidrológico do efeito de barramento hidráulico no rio tarumã-açu, manaus-am. *Revista Brasileira de Geomorfologia*, 22(2), 2021. doi:10.20502/rbg.v22i2.1752.

[26] Cassiano D'Almeida, Charles J. Vörösmarty, George C. Hurtt, José A. Marengo, S. Lawrence Dingman, and Barry D. Keim. The effects of deforestation on the hydrological cycle in amazonia: a review on scale and resolution. *International Journal of Climatology*, 27(5):633–647, 2007. doi:10.1002/joc.1475.

[27] Rogério Ribeiro Marinho, Tristan Harmel, Jean-Michel Martinez, and Naziano Pantoja Filizola Junior. Spatiotemporal dynamics of suspended sediments in the negro river, amazon basin, from in situ and sentinel-2 remote sensing data. *ISPRS International Journal of Geo-Information*, 10(2), 2021. doi:10.3390/ijgi10020086.

[28] Michael Li and Lily D. Li. A novel method of curve fitting based on optimized extreme learning machine. *Applied Artificial Intelligence*, 34(12):849–865, 2020. doi:10.1080/08839514.2020.1787677.

[29] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 3 edition, 2022.

[30] Ernst Wit, Edwin van den Heuvel, and Jan-Willem Romeijn. 'all models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236, 2012. doi:10.1111/j.1467-9574.2012.00530.x.

[31] G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35:2769–2794, 2007. doi:10.1214/009053607000000505.

[32] Carlos F. O. Mendes and Marcus W. Beims. Distance correlation detecting lyapunov instabilities, noise-induced escape times and mixing. *Physica A: Statistical Mechanics and its Applications*, 512:721–730, 2018. doi:10.1016/j.physa.2018.08.028.

[33] C. F. O. Mendes, R. M. da Silva, and M. W. Beims. Decay of the distance autocorrelation and lyapunov exponents. *Phys. Rev. E*, 99:062206, 2019. doi:10.1103/PhysRevE.99.062206.

[34] C. F. O. Mendes, L. Menon, R. M. da Silva, and M. W. Beims. Detecting correlations in desynchronized chaotic chimera states. *International Journal of Bifurcation and Chaos*, 32(05):2230012, 2022. doi:10.1142/S0218127422300129.

[35] D. Chicco, M. J. Warrens, and G. Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Comput Sci*, 7:e623, 2021. doi:10.7717/peerj-cs.623.

[36] Oliver T. Coomes, Michel Lapointe, Michael Templeton, and Geneva List. Amazon river flow regime and flood recessional agriculture: Flood stage reversals and risk of annual crop loss. *Journal of Hydrology*, 539:214–222, 2016. doi:10.1016/j.jhydrol.2016.05.027.

[37] Fernanda Fonseca, Bruna Worfarth-Couto, Andre Santos, Rogério Marinho, Jean-Michel Martinez, and Naziano Filizola. Hydrological scenarios and malaria incidence in the amazonian context. *Water*, 14(8), 2022. doi:10.3390/w14081283.

Trabalhos em andamento

- Como parte de um projeto atualmente em andamento, também estamos utilizando Redes Neurais Convolucionais (CNNs) para melhorar a gestão de resíduos sólidos nos igarapés e na orla do Rio Negro em Manaus. Este trabalho foca especialmente na identificação de materiais recicláveis, como, por exemplo, garrafas PET a partir de imagens de drone. O objetivo principal é chamar a atenção das autoridades competentes e aumentar a eficiência do processo de reciclagem, contribuindo assim para práticas mais sustentáveis e para a mitigação dos impactos ambientais negativos associados à disposição inadequada de resíduos sólidos nos igarapés e rios;

- Como perspectiva de trabalhos futuros, poderão ser elaboradas redes neurais com arquiteturas de sub-redes utilizando modelos como LSTM (Long Short-Term Memory), RNN (Recurrente Neural Networks), GANs (Generative Adversarial Networks) e modelos híbridos para a previsão de cheias e secas extremas. Além disso, essas sub-redes serão expandidas para outras sub-bacias da Bacia Amazônica, visando aprimorar a modelagem e a previsão de eventos hidrológicos extremos em diferentes regiões da Amazônia. Juntamente com estudos de correlações multivariadas através de dados de monitoramento hidrológicos e climáticos de alta dimensão.