

XXVI SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS

UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA PREVISÃO DE NÍVEL EM POSTOS TELEMÉTRICOS

Helen Graciane Ruela Machado¹; Josielton da Silva Santos; Milena Pereira Dantas; Rafael Grinberg Chasles; Diego Freitas de Souza; Diego Monteiro; Brisa Maria Fregonesi; Vitória Amélia Lemes Gonçalves; André Luis Navarro; Anderson Barboza Esteves

Abstract: Floods in the State of São Paulo cause significant impacts on the population and can occur following extreme events of riverbank overflow due to intense rainfall. Many rivers in the state are equipped with telemetric stations that record water levels in real time. These stations are monitored by the São Paulo State Water Agency (SP-Águas), but the monitoring is limited to the current water level only. Studies involving Artificial Neural Networks (ANNs) have increasingly contributed to flood prediction based on rainfall, streamflow, and physical characteristics of the watershed, demonstrating strong model performance in forecasting complex events. In this context, the aim of this study was to evaluate the performance of ANNs in predicting river levels at the telemetric stations monitored by SP-Águas, based solely on rainfall behavior and streamflow records. The model showed good performance in learning rainfall patterns during overflow events and in predicting water level values that were very close to the actual measurements at some stations. It achieved a Mean Absolute Percentage Error (MAPE) of 0.5% and an R^2 of 0.94, indicating that the model was able to explain 94% of the data variability, with 6% still requiring improvement.

Resumo: As inundações no Estado de São Paulo causam grandes impactos para a população e podem ocorrer após eventos extremos de extravasamentos da calha do rio em consequência a grandes chuvas. Muitos rios no Estado de São Paulo possuem postos telemétricos que registram os seus níveis em tempo real, os postos são monitorados pela Agência de Água do Estado de São Paulo (SP-Águas), porém o monitoramento ocorre apenas para o estado atual do nível. Trabalhos envolvendo *Artificial Neural Networks* (ANN) ou em português Redes Neurais Artificiais (RNA), têm ajudado cada vez mais a prever eventos de inundação com base em dados pluviométricos, fluviométricos e características físicas da bacia hidrográfica, mostrando um grande desempenho do modelo para prever eventos complexos. Nesse sentido, o objetivo deste trabalho foi analisar o desempenho da ANN para prever os níveis nos postos telemétricos monitorados pela SP-Águas com base apenas no comportamento da chuva e os registros fluviométricos. O modelo apresentou um bom desempenho em aprender os padrões de chuva em eventos de extravasamento e a prever valores de nível muito próximos do real para algumas estações, apresentando o MAP de 0.5% e R^2 de 0,94% indicando que o modelo conseguiu explicar 94% da variabilidade dos dados e 6% ainda precisam ser melhor trabalhados e explicados.

Palavras-Chave – Redes Neurais Artificiais; Previsão de Inundação; Monitoramento de nível do rio.

1) Agência de Águas do Estado de São Paulo (SP-ÁGUAS). Rua Boa Vista, 175-Sé – São Paulo/SP. CEP 01014-000. Telefone: (11)3293- 8200, hgrmachado.hcc3@spaguas.sp.gov.br.

INTRODUÇÃO

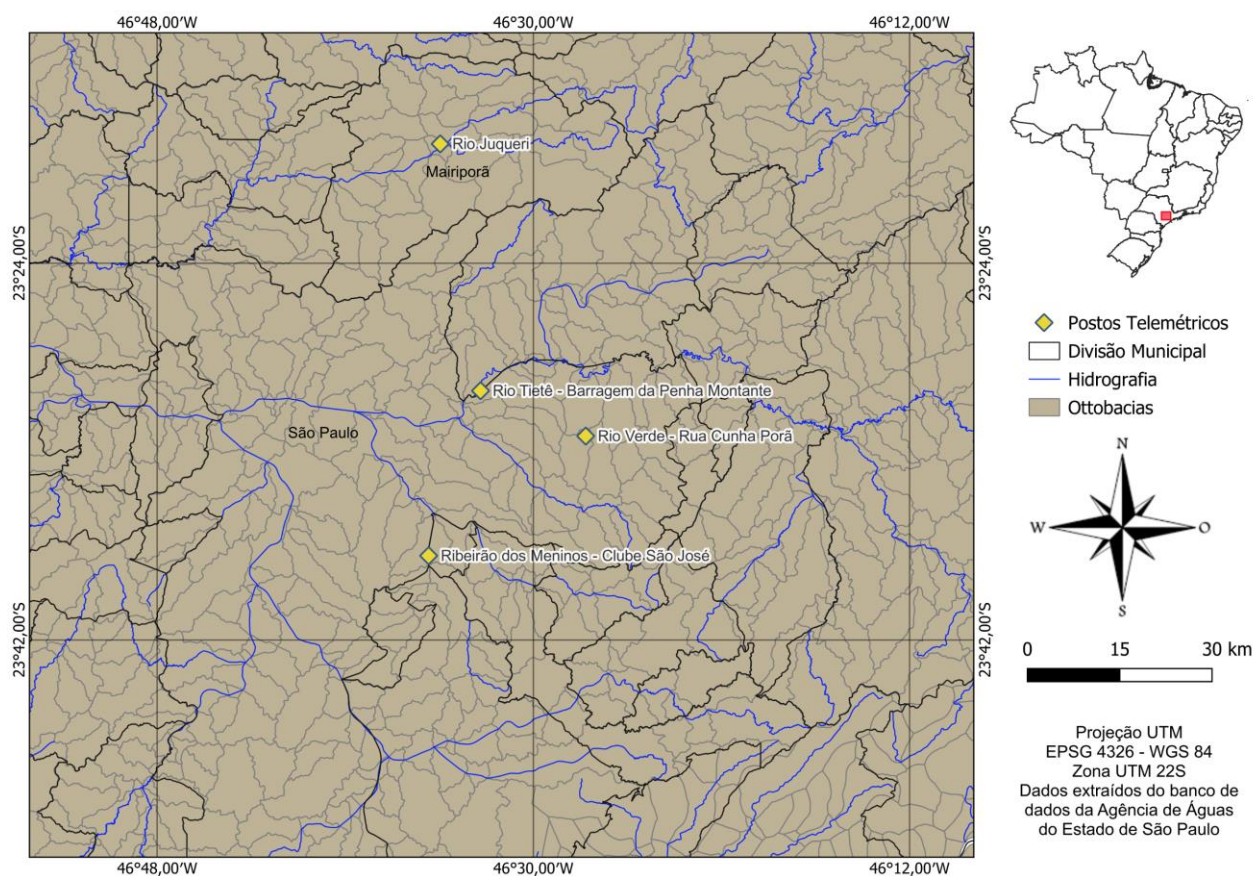
Os eventuais impactos causados por alterações de nível nos sistemas de drenagem em decorrência de eventos extremos causam danos massivos para a vida humana, infraestrutura, agricultura e sistemas socioeconômicos (Mosavi *et al.*, 2018). O aumento na precipitação e em suas anomalias, influenciados por fatores climáticos, evidenciam a susceptibilidade da região a eventos de inundação (Mantovani *et al.*, 2025). Os eventos extremos de chuvas podem ocasionar nas inundações, que são provenientes do processo de extravasamento das águas do canal de drenagem para as áreas marginais (planície de inundação, várzea ou leito maior do rio) quando a enchente atinge cota acima do nível máximo da calha principal do rio (Tucci, 2007).

De acordo com os dados de monitoramento da Agência de Águas do Estado de São Paulo (SP-Águas), foram registrados entre 2010 e o início de 2025 aproximadamente 1900 extravasamentos nos postos telemétricos e durante o mesmo período foram registrados pela Defesa Civil 921 (São Paulo, 2025) casos de inundações no estado de São Paulo, causando danos e perdas principalmente nas áreas urbanas. Projetos de mitigação a inundações em escala local ou regional têm sido construídos cada vez mais com base no monitoramento da rede hidrográfica e na construção de modelos de previsão (Mosavi *et al.* 2021; Meng *et al.* 2025). No entanto, a previsão de cheias apresenta diversos desafios técnicos e científicos, especialmente em relação ao tempo de antecedência (*lead time*) e à localização da ocorrência, devido à complexidade dinâmica do clima e do comportamento das bacias hidrográficas (Mosavi *et al.* 2018).

A precipitação pluvial é uma variável chave e com maior influência nos estudos de previsão e modelagem nos riscos de inundação (Lafdani *et al.* 2013; Mosavi *et al.* 2018) em conjunto com outras variáveis como vazão do rio, umidade do solo, precipitação-vazão e entre outros (Maier *et al.* 2010). Muitos trabalhos que adotaram o uso de Machine Learning (ML), apontam para a grande eficiência do modelo *Artificial Neural Networks* (ANN), em português Redes Neurais Artificiais (RNA), para previsão de eventos de curto prazo que ocasionam em inundações (Pereira Filho *et al.* 2005; Maier *et al.* 2010; Meng *et al.* 2025). As ANNs possuem versatilidade e eficiência na modelagem de processos complexos de inundação em que são frequentemente utilizados registros históricos dos eventos, além de dados cumulativos em tempo real de vários pluviômetros ou dados de reanálise, em escala de tempo horária, diária e/ou mensal (Mosavi *et al.* 2018; Meng *et al.* 2025). Assumindo a importância dos modelos de previsão para eventos extremos e que as inundações são decorrentes de extravasamentos no canal fluvial, o objetivo deste trabalho é analisar o desempenho da ANN para a previsão de nível em eventos de extravasamentos nos postos telemétricos, levando em consideração os dados de precipitação.

Os postos utilizados para análise deste trabalho estão localizados nos municípios de São Paulo e Mairiporã conforme figura 1, foram selecionados postos telemétricos que possuem níveis similares e cotas de extravasamento definidas. Os postos selecionados levam os nomes dos rios em que estão instalados, sendo eles o posto Rio Tietê - Barragem da Penha Montante, Rio Verde - Rua Cunha Porã e Ribeirão dos Meninos - Clube São José no município de São Paulo e Rio Juqueri no município de Mairiporã, todos os postos são operados pelo Sistema de Alerta a Inundações de São Paulo (SAISP) e monitorados pelo SP-Águas.

Figura 1: Localização da área de estudo. Postos telemétricos SAISP monitorados pelo SP-Águas



MATERIAIS E MÉTODOS

A metodologia utilizada se baseia em Maier *et al.* (2010) que descreve os passos para o desenvolvimento das ANNs aplicadas à previsão de variáveis hidrológicas: o primeiro passo consiste na escolha das variáveis a serem previstas e as variáveis de entrada do modelo e no processamento dos dados, em seguida é feito o desenvolvimento do modelo baseado na seguinte função:

$$Y = f(X, W) + \varepsilon \quad (1)$$

onde Y é o vetor de saídas do modelo; X, o vetor de entradas do modelo; W, vetor de parâmetros do modelo (pesos das conexões); $f()$, relação funcional entre saídas, entradas e parâmetros do modelo; ε , vetor de erros do modelo. Em conjunto com a construção do modelo é necessário a divisão dos dados de treino, teste e validação, avaliação da performance e calibração do modelo.

Os vetores de saída (Y) e os de entrada (X) para o modelo de previsão de extravasamento foram utilizados dados históricos das redes telemétricas monitoradas pela Agência de Águas do Estado de São Paulo, considerando a série histórica de 2010 a 2025 de 4 postos públicos com cotas de extravasamento definidas. O pré-processamento e tratamento de dados consistiu no agrupamento das medições fluviométricas das redes telemétricas em valores diários, considerando-se o nível máximo registrado no dia durante o agrupamento como valor principal. Para o tratamento dos dados a metodologia utilizada se baseou em Meng *et al.* (2025), adotando os dados de reanálise para a chuva acumulada. Os dados pluviométricos de reanálise utilizados foram do modelo MERGE do Instituto Nacional de Pesquisa Espaciais (INPE) que consiste na combinação da precipitação observada com

estimativa de precipitação por satélite (Rozante *et al.* 2010), disponível em: <https://ftp.cptec.inpe.br/modelos/tempo/MERGE>.

Foi realizado a estatística zonal do MERGE para se obter as médias e máximas de chuva acumulada por área (Rob e Yanan, 1996), as áreas utilizadas como zona de referência foram as Bacias Hidrográficas Ottocodificadas (BHO) de nível 7, baseado na codificação de Otto Pfafstetter para representar a rede hidrográfica em trechos entre os pontos de confluência dos cursos d'água de forma unifilar, cada trecho é associado a uma superfície de drenagem denominada ottobacia e representam corretamente o fluxo hidrológico dos rios, por meio de trechos conectados e com sentido de fluxo (BRASIL, [s.d.]). Os dados das ottobacias do Estado de São Paulo podem ser acessados em <https://geodados.daee.sp.gov.br/catalogue/#/dataset/428>. Para cada data entre 2010 e o início de 2025 foi utilizada como valor principal a cota máxima registrada no dia e o acumulado de chuva do dia (d0), do dia seguinte (d+1), de um dia anterior (d-1) e de dois dias anteriores (d-2).

Seguindo a estrutura de Maier *et al.* (2010) foi utilizada a arquitetura de rede *multilayer perceptron* (MLP) como relação funcional $f()$. O MLP é um tipo de rede neural *feed-forward* (FFNN) que emprega aprendizado supervisionado por *backpropagation* (BP) para treinar uma rede de nós interconectados em múltiplas camadas (Rumelhart *et al.* 1986) e em avaliações comparativas entre diferentes classes de ANNs para modelagem de inundações, os MLPs foram escolhidos por apresentarem maior eficiência e melhor capacidade de generalização (Mosavi *et al.* 2018). A geometria do MLP utilizada no modelo possui 2 camadas ocultas (64 e 32 nós) e função de ativação ReLU (para ocultas) e linear (para saída). Seguindo a equação baseada em Maier *et al.* (2010) o modelo foi estruturado da seguinte forma:

X: Vetor de entradas (dados numéricos de reanálise de chuva [d+1, d0, d-1, d-2] + *embeddings*)

W: Pesos aprendidos (incluindo os das camadas densas e *embeddings*).

$f()$: A função representada pelo MLP.

Y: Saída (previsão do nível)

Os dados utilizados para a previsão do modelo seguiram o mesmo processamento dos dados de treino. Foi selecionada uma data em que 2 dos 4 postos utilizados para treinamento registraram extravasamento, para essa mesma data foram feitas as estatísticas zonais dos dados de reanálise de chuva no intervalo de d+1, d0, d-1 e d-2. O modelo foi construído na linguagem de programação *python* e as principais bibliotecas utilizadas para o modelo foram o *tensorflow*, *keras* e *sklearn*. Foi utilizado na divisão dos dados de treino e teste a função *GroupShuffleSplit* da biblioteca *sklearn*, para evitar vazamento de dados entre as estações, foram considerados 20% dos dados para teste e 80% para treino. Também foi utilizada a função *EarlyStoppin* para evitar *overfitting* e garantir que o modelo aprendesse os padrões nos dados históricos. As métricas de treino utilizadas foram a *loss*, indicando o erro total (função de perda) nos dados de treino, *mae* (Mean Absolute Error) indicando o erro absoluto médio e o *mse* (Mean Squared Error) indicando o erro quadrático médio. As métricas de validação do modelo foram o *val_loss* (erro total nos dados de validação), *val_mae* e *val_mse* (versões validadas das métricas anteriores). Para a avaliação global do modelo foi utilizado o R^2 .

RESULTADOS E DISCUSSÕES

Foram geradas 19 épocas (*Epoch*) de passagem completa do algoritmo de treinamento por todos os dados, ao final das épocas ocorre um aumento no *val_loss* e diminuição do *loss*, sugerindo que o modelo começou a sofrer *overfitting*, mas logo foi parado pela função *EarlyStoppin* o que reforça a sua importância para evitar e tratar *overfitting*, a tabela 1 apresenta uma descrição e análise das épocas de treinamento. O modelo de rede neural apresentou desempenho satisfatório na previsão dos níveis

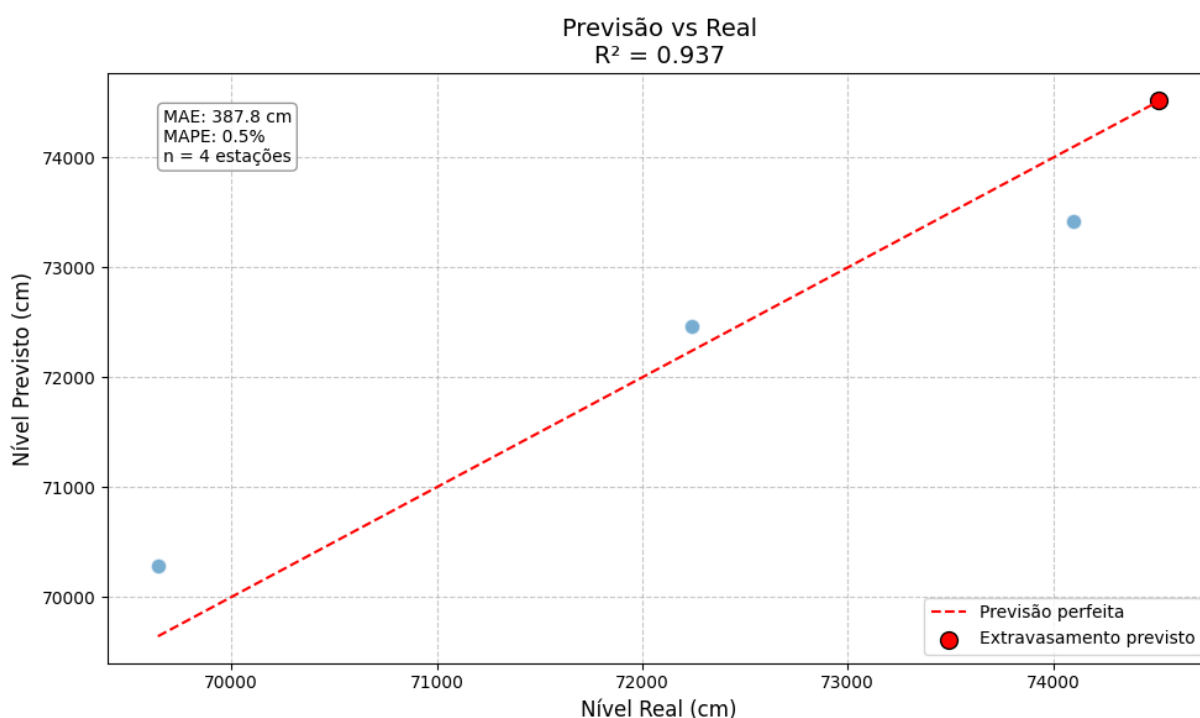
diários, com um MAPE de 0,5% indicando boa assertividade do modelo, porém o MAE de 387.4cm na escala original sugere um erro médio de aproximadamente 4m, mostrando que o modelo aprendeu bem os padrões de algumas estações enquanto ruídos e outliers podem ter influenciado na elevada taxa de erro de outras estações.

Tabela 1 – Análise e descrição das épocas de treinamento

Épocas	Situação	Descrição	Análise
1-5	Aprendizado Inicial	Ocorre queda rápida do loss de 0.379 para 0.102 e val_loss de 0.202 para 0.088	O modelo está aprendendo padrões gerais dos dados.
6-16	Aperfeiçoamento	Melhoria contínua, porém, mais lenta. A época 16 apresentou o val_loss mínimo de 0.0182.	Melhor desempenho em dados não vistos na época 16
17-19	Overfitting	Aumento contínuo de val_loss: 0.0254 na época 17, 0.0170 na época 18 e 0.0237 na época 19, enquanto loss continua caindo.	O modelo pode estar se ajustando demais aos dados de treino. O early stopping interrompeu o treinamento.

A taxa de erro do MAE reflete no coeficiente R^2 , que foi de 0.94, demonstrando que 94% da variabilidade dos dados foi explicada pelo modelo e 6% ainda precisam ser melhor trabalhadas. A figura 2 mostra a dispersão entre os valores previstos e reais, é possível observar forte correlação entre previsões e valores reais, o ponto vermelho no gráfico representa o posto que estaria em extravasamento de acordo com a previsão de nível do modelo.

Figura 2 – Gráfico de Dispersão entre os valores previstos e os valores reais para as estações analisadas



A tabela 2 mostra os valores reais comparados com os níveis previstos. De acordo com os níveis previstos pelo modelo, foi verificado qual estaria em situação de extravasamento comparado com a cota definida em cada posto, o resultado da análise foi que entre os dois postos que extravasaram, apenas um posto estaria em extravasamento pela previsão de nível do modelo.

Tabela 2 – Descrição entre nível previsto pelo modelo e nível real para cada estação

Posto	Cota de Extravasamento	Nível Previsto	Situação Prevista	Nível Real	Situação Real
Rio Tietê - Barragem Da Penha Montante	72560.0	72460.2	-	72243.3	-
Rio Verde - Rua Cunha Porã	74498.0	74521.7	Extravasamento	74515.8	Extravasamento
Ribeirão dos Meninos - Clube São José	74020.0	73415.6	-	74099.4	Extravasamento
Rio Juqueri	70487.0	70287.5	-	69642.8	-

O modelo previu um nível de 6.8 metros abaixo do nível real para o posto Ribeirão dos Meninos - Clube São José e 6.4 metros acima do real para o posto Rio Juqueri, ambos os resultados da previsão podem ser explicados pelo comportamento dos dados históricos, em que, no caso do Rio Juqueri após o tratamento e filtragem de dados apenas um evento de extravasamento foi aproveitado para o treinamento e por esse motivo ele pode ter sido interpretado como um ruído, interferindo nos pesos entre as classes (extravasamento ou não), no treinamento e superestimando o nível previsto. No posto Ribeirão dos Meninos - Clube São José os dados históricos também apresentam grande quantidade de registros com níveis abaixo da média, fazendo com que o modelo resultasse em uma previsão muito abaixo do nível real. O posto Rio Tietê - Barragem Da Penha Montante apresentou a diferença de 2.1 metros acima do nível real, mostrando um equilíbrio maior entre os dados históricos no período analisado e o posto Rio Verde - Rua Cunha Porã teve uma diferença de apenas 5.9 cm acima do nível real, que para o objeto de estudo analisado é um resultado quase perfeito.

CONCLUSÃO

O modelo de Redes Neurais Artificiais desenvolvido neste trabalho apresentou bom desempenho para prever níveis nos postos telemétricos nos municípios de São Paulo e Mairiporã, levando em considerando apenas os dados históricos diários do acumulado de chuva por ottobacia. Porém, o MAE de 387.8cm, equivalente a aproximadamente 4m, ainda é um valor muito alto. Os dados indicam que o modelo pode estar se ajustando demais aos dados de algumas estações e aumentando o erro para outras, isso devido a um possível desbalanceamento dos dados históricos em algumas estações, dificultando a análise de padrões apenas para o padrão de chuva. Os 6% do R^2 não explicados pelo modelo podem ser devido a fatores não incluídos na análise, como as variáveis da vazão do rio, umidade do solo, precipitação-vazão e entre outros indicadores utilizados em modelos de previsão de eventos extremos como mencionado em Maier *et al.* (2010) e nesse sentido, para melhorar a previsão do modelo em trabalhos futuros além de ser adotados as características físicas sugeridas por Maier *et al.* (2010), serão adotadas dados observáveis de chuva registrados nas estações para comparação com os dados de reanálise.

REFERÊNCIAS

BRASIL. *Agência Nacional de Águas e Saneamento Básico – ANA*. Posto pluviométrico – Código 24464100 – Ribeirão da Prata. SNIRH – Sistema Nacional de Informações sobre Recursos Hídricos.

Disponível em: <https://metadados.snirh.gov.br/geonetwork/srv/api/records/b228d007-6d68-46e5-b30d-a1e191b2b21f>. Acesso em: 16 jun. 2025.

LAFDANI, E.K.; NIA, A.M.; PAHLAVANRAVI, A.; AHMADI, A.; JAJARMIZADEH, M. (2013) *Research article daily rainfall-runoff prediction and simulation using ANN, ANFIS and conceptual hydrological MIKE11/NAM models*. Int. J. Eng. Technol 1, 32–50.

MAIER, H. R.; JAIN, A.; DANDY, DANDY, C. G.; SUDHEER, K. P. (2010). *Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions*. Environmental Modelling & Software 25, 891 e 909. <https://doi.org/10.1016/j.envsoft.2010.02.003>

MANTOVANI, J. R. et al. (2025). *Unprecedented flooding in Porto Alegre Metropolitan Region (Southern Brazil) in May 2024: Causes, risks, and impacts*. Journal of South American Earth Sciences 160, 105533. <https://doi.org/10.1016/j.jsames.2025.105533>

MENG, X.; LIN, C.; DING, J.; WANG, G.; ZHANG, J.; WANG, H.; CHU, C. (2025). *Spatiotemporal evolution of droughts and floods in the Yellow River Basin: A novel approach combining CMADS-L evaluation, hydroclimatic zonation and CNN-LSTM prediction*. Journal of Hydrology: Regional Studies 58, 102250. <https://doi.org/10.1016/j.ejrh.2025.102250>

MOSAVI, A.; OZTURK, P.; CHAU, K. (2018). *Flood Prediction Using Machine Learning Models: Literature Review*. Water 10, 1536. <https://doi.org/10.3390/w10111536>

PEREIRA FILHO, A. J.; SANTOS, C. C. dos. (2006). *Modeling a densely urbanized watershed with an artificial neural network, weather radar and telemetric data*. Journal of Hydrology 317, 31–48. <https://doi.org/10.1016/j.jhydrol.2005.05.007>

ROB, J. H.; YANAN F. (1996) *Sample Quantiles in Statistical Packages*. The American Statistician, Vol. 50, N. 4, novembro, p. 361-365.

ROZANTE, J. R. et al. (2010). *Combining TRMM and surface observations of precipitation: technique and validation over South America*. Weather and forecasting v. 25, n. 3, p. 885-894, 2010.

TUCCI, CARLOS E. M. (org.). (2007). *Inundações urbanas*. Brasília: Agência Nacional de Água.

RUMELHART, D.E.; HINTON, G.E.; WILLIAMS, R.J. (1986). *Learning representations by back-propagating errors*. Nature, 323, 533.

SÃO PAULO (Estado). SIDECA – *Sistema de Informações para a Decisão Estadual sobre Clima*. Secretaria de Meio Ambiente, Infraestrutura e Logística. Disponível em: <https://www.sidec.sp.gov.br/>. Acesso em: 16 jun. 2025.

AGRADECIMENTOS

Os autores agradecem à SP-ÁGUAS pelo fornecimento dos dados da rede telemétrica que foram a base da análise desta pesquisa e à HCC3 pelo apoio ao longo do desenvolvimento desta pesquisa.