Realização PUC CAMPINAS USP
Workshop Internacional
SUSTENTARE & WIPIS 2025
Sustentabilidade, Indicadores e Gestão de Recursos Hídricos
Apoio Institucional CAPES CNPq

PREVISÃO DE INUNDAÇÕES URBANAS UTILIZANDO INTELIGÊNCIA ARTIFICIAL

Marcelo Gomide Silveira, Ana Elisa Silva de Abreu
Instituto de Geociências, Universidade Estadual de Campinas - UNICAMP, Campinas, SP, Brasil
marcelo.gomide@gmail.com, aeabreu@unicamp.br

Resumo: A urbanização intensiva, aliada à crescente frequência de eventos climáticos extremos, tem agravado significativamente a ocorrência de inundações bruscas em centros urbanos, gerando significativos riscos sociais, econômicos e ambientais. Este trabalho tem como objetivo principal avaliar a aplicação do algoritmo de Machine Learning para a modelagem hidrológica da bacia do Ribeirão Proença, em Campinas-SP, visando a previsão do nível de água em tempos 30 minutos à frente, a partir de dados de radar meteorológico. Foram utilizados dados oriundos do radar meteorológico de dupla polarização de banda X que operou na região de Campinas entre os anos de 2016 e 2018. A metodologia aplicada envolveu a organização de um banco de dados multimodal, o tratamento e a análise da correlação entre a precipitação e o comportamento do nível do ribeirão para estimar o lag-time da bacia. Para a predição do nível de água, foi empregado o algoritmo Extreme Gradient Boosting (XGBoost). Os resultados obtidos a partir da análise de 248 dias, indicam um desempenho satisfatório do modelo, com um Erro Quadrático Médio (RMSE) de 0,064 metros e um Coeficiente de Eficiência de Nash-Sutcliffe (NSE) de 0,580 para os dados de teste. A análise revelou variabilidade significativa no tempo de resposta da bacia, com lag-times variando entre 10 e 150 minutos. A abordagem aplicada demonstrou potencial para aplicação em Sistemas de Alerta Antecipado (SAA) em microbacias urbanas com características semelhantes, contribuindo para a mitigação dos impactos das inundações, principalmente as bruscas. O estudo reforça a viabilidade do uso de inteligência artificial em contextos hidrológicos urbanos, especialmente em áreas com alta impermeabilização.

Palavras-chave: Modelagem Hidrológica, Aprendizado de Máquina, Inundações Bruscas, Radares Meteorológicos, Sistemas de Alerta Antecipado.

1. INTRODUÇÃO

A ocorrência de inundações urbanas tem se tornado cada vez mais comum, principalmente devido à urbanização, que frequentemente é considerada uma força motriz para o progresso socioeconômico de uma cidade ou país [1]. A urbanização provoca mudanças significativas no uso da terra e na infraestrutura, com o objetivo de melhorar a qualidade de vida [2]. As inundações bruscas, que ocorrem rapidamente em áreas densamente povoadas, representam riscos significativos para a infraestrutura, segurança publica e o desenvolvimento sustentável, e prever esses eventos vem se tornando cada vez mais importante para construir resiliência e mitigar seus efeitos prejudiciais [3].

Nesse contexto, a alta resolução espacial e temporal das observações fornecidas por radares meteorológicos tem se mostrado essencial tanto para a geração de estimativas quantitativas de precipitação quanto para previsões mais precisas de eventos extremos, sendo que essas informações são fundamentais para o aprimoramento da eficácia dos Sistemas de Alerta Antecipado (SAA) voltados à alertas de risco de inundações bruscas. Tratando desses sistemas, modelos que utilizam o nível de água dos rios como variável de monitoramento, juntamente com os dados de precipitação extraídos de radares meteorológicos, são cada vez mais comuns [4].

Dentre os modelos de inteligência artificial (IA) aplicáveis à modelagem hidrológica, os de *machine learning* (aprendizado de máquina) vem se tornando cada vez mais comuns, devido à sua robustez, flexibilidade e eficácia, atrelado à redução do custo das tecnologias de processamento computacional e acesso a grandes volumes de dados [5].

Este trabalho se insere nesse contexto e busca desenvolver uma modelagem hidrológica capaz de prever o extravasamento de um rio situado na área urbana de Campinas, onde já foram registradas diversas inundações bruscas, em tempo hábil para emissão de alerta antecipado a inundação.

2. ESTUDOS ANTERIORES

Araujo [6] desenvolveu um trabalho de modelagem baseada em dados para a bacia do rio Proença, me Campinas, que é a mesma área de estudo desta pesquisa. O objetivo principal do estudo de Araujo [6] foi criar um modelo de predição de níveis de água futuros na bacia do ribeirão Proença, visando gerar uma base técnica para futura implementação de um sistema de alerta antecipado. Para a construção do modelo, foram utilizados dados de nível de água obtidos pelo sensor 4D-038, situado na confluência do ribeirão Proença com o córrego Serafim, juntamente com dados de precipitação provenientes do radar meteorológico de banda S localizado no munícipio de Salesópolis no estado de São Paulo. Os dados utilizados por Araujo [6] compreenderam um período entre novembro de 2014 a junho de 2019. O estudo de Araujo [6] utilizou registros de ocorrências de inundações da Defesa Civil, além de meios digitais para validar os eventos hidrológicos. Uma das contribuições centrais do trabalho foi a organização de um banco de dados multimodal unificando as diferentes fontes de informação.

Para a identificação das datas com eventos de inundação na bacia alvo do estudo Araujo realizou uma análise estatística dos níveis máximos diários registrados definindo assim limiares de criticidade a partir da frequência com que os níveis de água eram atingidos ou superados, sendo estabelecidos três principais limiares:

- Limiar 1 (Q10): Corresponde ao nível de 1,58 metros, que é igualado ou excedido em 10% do tempo analisado. É o limiar mais abrangente, utilizado para selecionar os eventos alvo do estudo.
- Limiar 2 (Q6): Corresponde ao nível de 1,96 metros, igualado ou excedido em 6% do tempo analisado
- Limiar 3 (Q3): Corresponde ao nível de 2,30 metros, igualado ou excedido em 3% do tempo analisado.

Com base na correlação desses limiares com registros históricos, o nível de 2,30 metros (Limiar 3) foi sugerido por Araujo [6] como uma cota de atenção para potenciais inundações, enquanto o nível de 3,00 metros foi estabelecido como cota de alerta. Araujo desenvolveu um modelo hidrológico com o algoritmo de *Machine Learning XGBoost* com o objetivo de prever o nível de água com 30 minutos de antecedência. Os resultados para os dados de teste indicaram um desempenho classificado como “satisfatório”, com um erro médio (RMSE) de 0,284 metros e um índice de eficiência Nash-Sutcliffe (NSE) de 0,614.

Um resultado fundamental da pesquisa de Araujo foi a identificação de dois grupos distintos de *lag-time*: um de 30 minutos, associado a chuvas mais intensas, e outro de 70 minutos, para chuvas de menor intensidade, evidenciando a alta variabilidade intrínseca do comportamento da bacia.

3. MATERIAIS E MÉTODOS

A Figura 1 apresenta a localização da bacia do rio Proença no município de Campinas, SP, que foi objeto desse estudo.

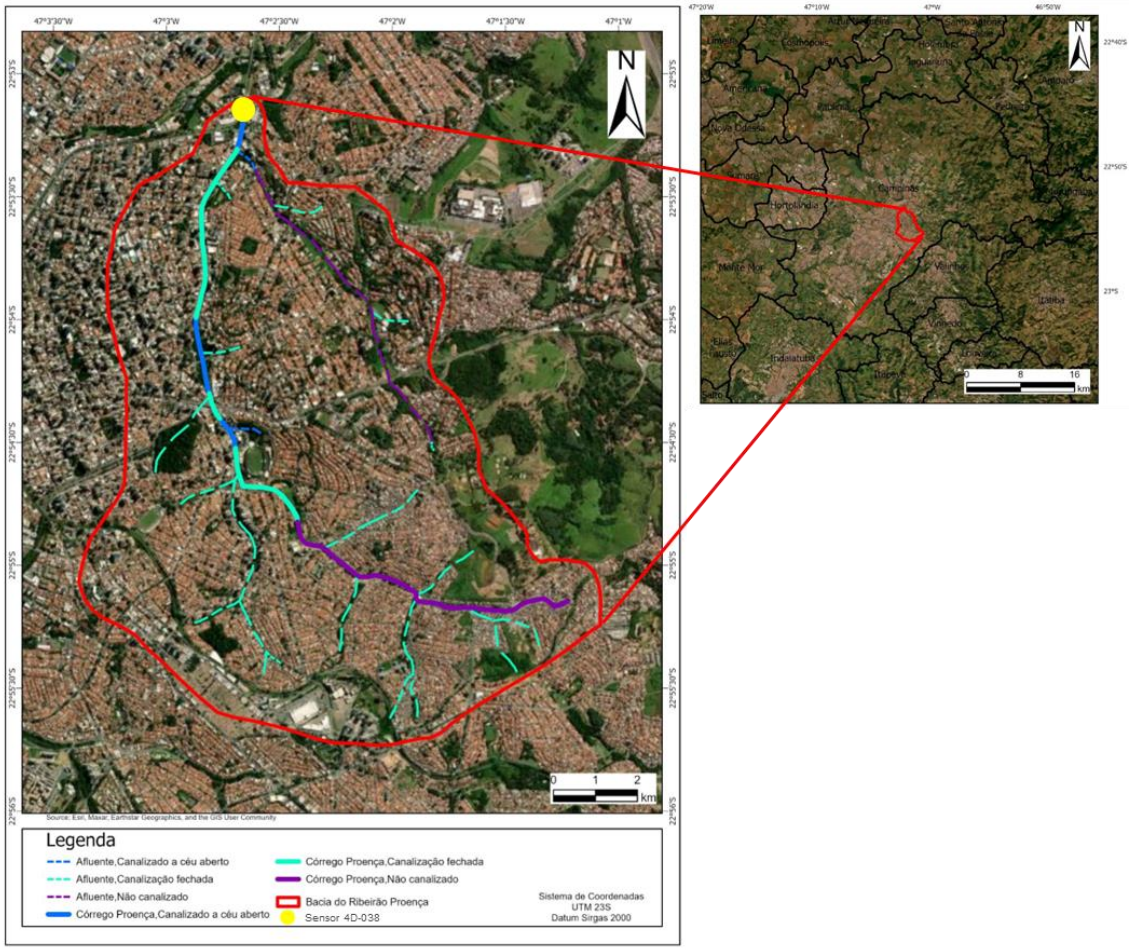


Figura 1: a) Localização da área de estudo com destaque para o sensor 4D-038 em amarelo.

Dados utilizados:

- Níveis de água oriundos do sensor 4D-038 (DAEE) em intervalos de tempo de 2, 5 e 10 minutos, fornecidos pelo DAEE-SP.
- Dados referentes ao radar meteorológico de banda X, operado na UNICAMP entre os anos de 2016 e 2018 durante o programa SOS-CHUVAS em conjunto com o INPE, com espaçamento temporal de 10 minutos, extraído de INPE [7].
- Banco de Dados de inundações no município de Campinas [6].

Processamento dos dados:

- Transformação dos dados do sensor 4D-038 em tabelas com data e hora da medição, nível de água.
- Transformação dos dados de radar meteorológico de reflectância para precipitação através de linguagem computacional (Python), fornecido pelo INPE [7] e georreferenciamento de imagens através do software ArcGis. Nessa etapa foi extraída a quantidade de chuva em cada pixel localizado no interior da bacia, totalizando 79 pixels.
- Criação de um banco de dados unificado, correlacionando os dados de radar meteorológicos e níveis de água de acordo com a Data e Hora, contendo 83 variáveis (Data, Hora, Nível de

Água, quantidade de chuva em cada um dos 79 pixels da imagem de radar banda X localizados dentro da bacia, média de precipitação) através de linguagem computacional (Python)

- Cálculo do *lag-time* da microbacia através de correlação cruzada, realizada em ambiente Python.
- Modelagem *XGBOOST* para previsão de nível de água 30 minutos à frente, realizada em ambiente Python.

No total, foram selecionados os dados referentes a 248 dias entre novembro de 2016 e março de 2018. Nesse período, 20 vezes o Limiar 1 (1,56 m) foi ultrapassado, 4 vezes o Limiar 2 (1,96) foi ultrapassado e 4 vezes o Limiar 3 (2,30 m) foi ultrapassado.

4. RESULTADOS E DISCUSSÃO

Ao total foram selecionados 43.343 dados para aplicação do algoritmo de aprendizado de máquinas. Esses dados compreendem 248 dias diferentes, sendo que desses 248 dias, 2 dias tiveram eventos de inundação na bacia, 2 dias eventos de alagamentos na bacia, 10 dias com eventos registrados fora da bacia e 234 dias sem informação de eventos na bacia. No total, foram 20 dias em que ao menos o Limiar 1 foi atingido.

4.1. Estimativa do *lag-time* da bacia

A Figura 2 abaixo podemos observar que primeiramente temos a chuva, e posteriormente o aumento do nível de água do ribeirão Proença, no dia 18/11/2017, onde através de correlações cruzadas para verificar a correlação entre Nível de água e precipitação, temos o *lag-time*.

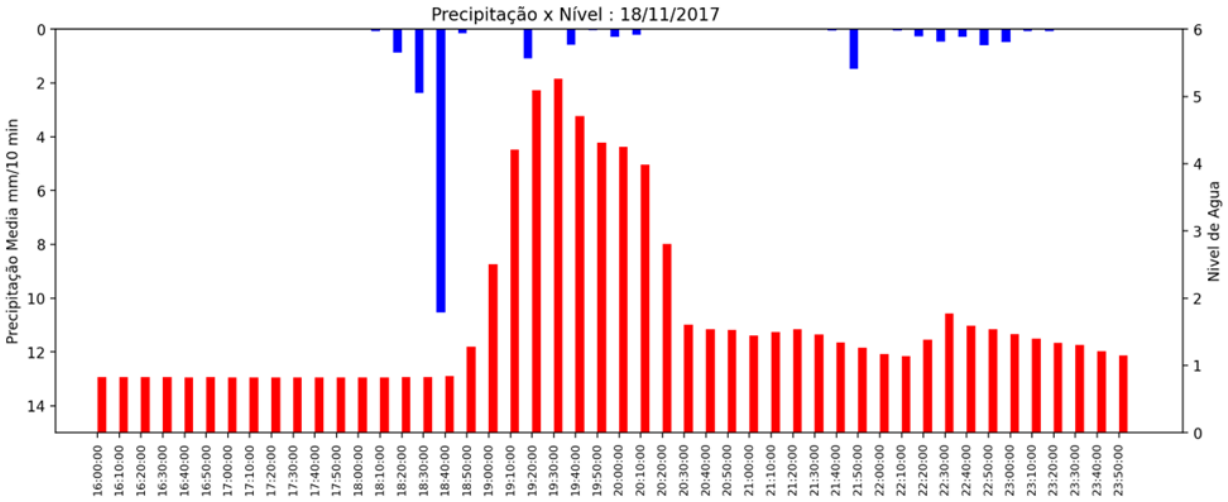


Figura 2: Nível de água (em vermelho – eixo secundário) e precipitação (em azul – eixo primário)) em função do tempo (eixo X) para o evento de inundação do dia 18/11/2017.

A Figura 3 mostra a correlação entre o *lag-time* e o nível de água para cada um dos eventos que ultrapassaram o Limiar 1, com destaque para o evento do dia 18/11/2017.

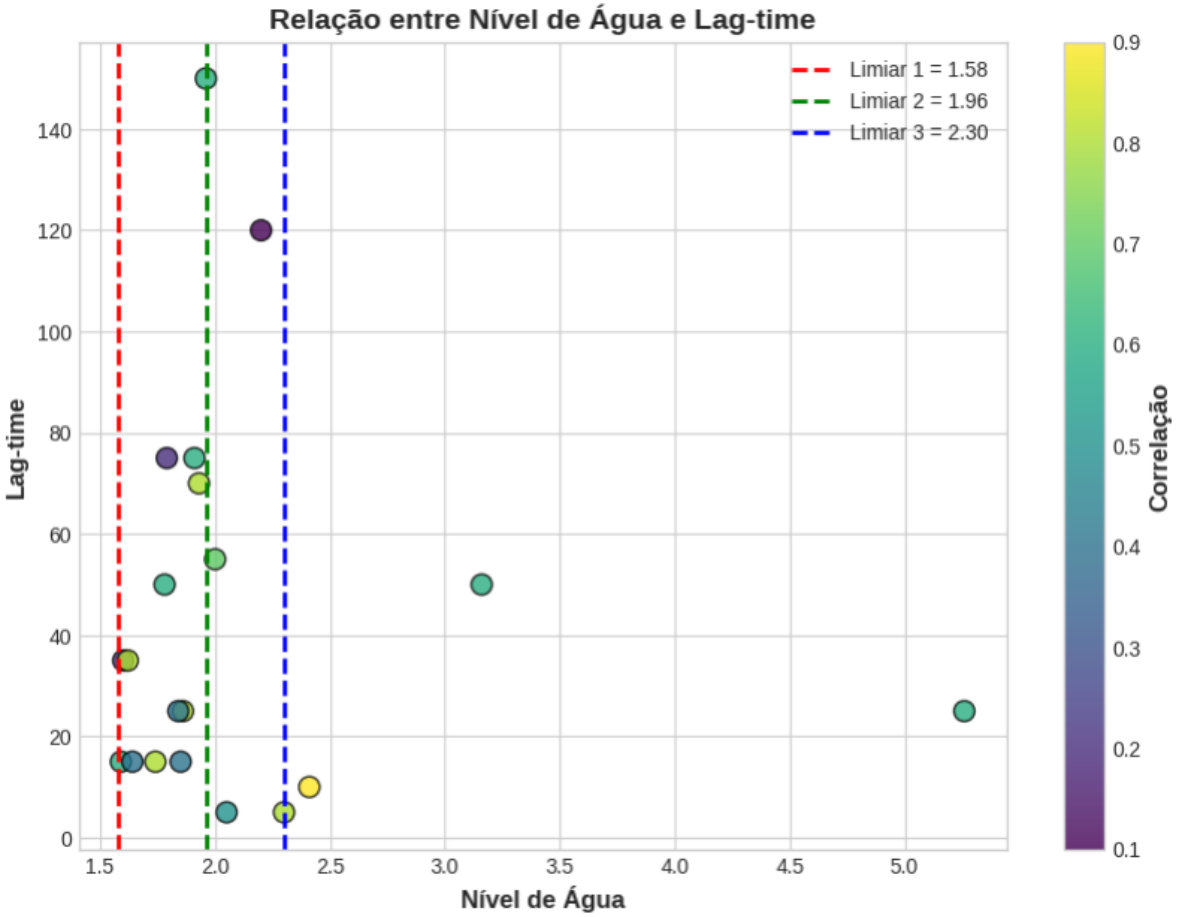


Figura 3: Relação entre Nível de água (eixo X) e *Lag-time* (eixo Y) e sua correlação (mapa de cores) para todos os eventos acima do Limiar 1, com destaque para o evento de 18/11/2017.

4.2. Modelagem *XGBOOST*

Para a modelagem, inicialmente foi criado um "alvo" sendo definido pelo nível de água 30 minutos após o evento. Para o modelo, a planilha de *input* de dados continha 42.321 linhas, divididas em 33.856 linhas para treinamento (80%), 4.231 linhas para validação (10%) e 4.231 linhas para teste (10%), conforme pode ser visualizado na Figura 4.

A Tabela I exibe os resultados de avaliação do modelo, tanto com os dados de treino e validação, como quanto com os dados de teste.

Tabela I: Resultados de avaliação do modelo XGBOSST para os dados de treinamento + validação e teste.

Dados	Métrica	Resultado
Treino + Validação	RMSE (m)	0,066
	NSE	0,578
Teste	RMSE (m)	0,064
	NSE	0,580

O resultado de NSE, tanto dos dados de treino + validação, como de teste indicam que o modelo é satisfatório. A Figura 4 abaixo apresenta a previsão dos níveis de água calculados através do modelo.

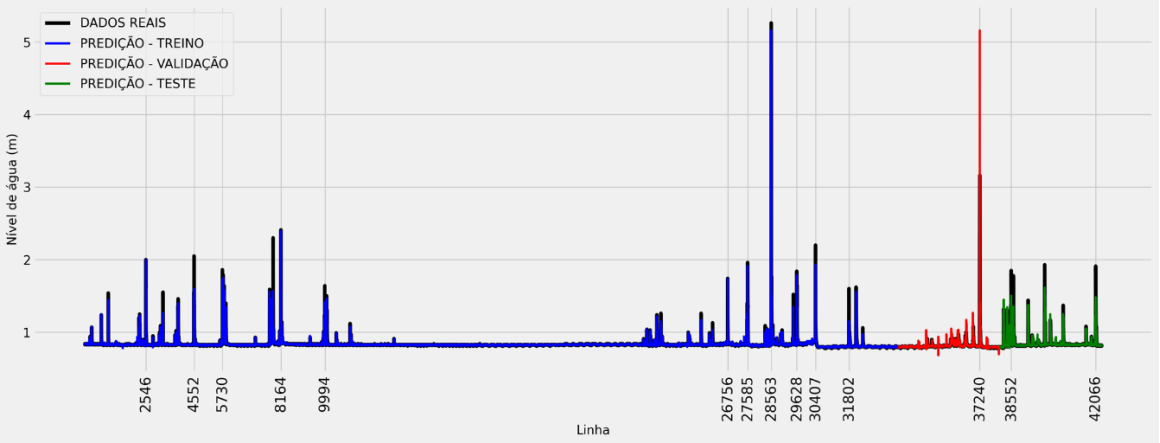


Figura 4: Previsão do nível de água 30 minutos no futuro realizada pelo modelo XGBOOST para os dados de treinamento (azul), validação (vermelho), teste (verde). As linhas em preto representam os níveis reais.

5. CONCLUSÃO

A abordagem aplicada demonstrou potencial para aplicação em Sistemas de Alerta Antecipado (SAA) em microbacias urbanas com características semelhantes, contribuindo para a mitigação dos impactos das inundações, principalmente as bruscas. O estudo reforça a viabilidade do uso de inteligência artificial em contextos hidrológicos urbanos, especialmente em áreas com alta impermeabilização.

Os próximos passos da pesquisa compreendem modelar os níveis d'água para tempos futuros de 10, 50, 60, 90, 120, 150 e 180 minutos, de acordo com os lag-times identificados nesta primeira etapa da pesquisa.

REFERÊNCIAS

[1] KAISER, Z. R. M. A. Failing to attain sustainable development in Bangladesh: A potential comprehensive strategy for sustainability. Sustainable Development, v.31, n.4, 3086–3101, 2023.

[2] STOKES, E. C.; SETO, K. C. Characterizing urban infrastructural transitions for the Sustainable Development Goals using multi-temporal land, population, and nighttime light data. Remote Sensing of Environment, v. 234, a. 111430, 2019.

[3] KAISER, Z.R.M.A.; AKTER, F. From risk to resilience and sustainability: Addressing urban flash floods and waterlogging. Risk Sciences, v. 1, 2025, a. 100011, p. 2950-6298, 2025

[4] SMITH, P. J.; BROWN, S.; DUGAR, S. Community-based early warning systems for flood risk mitigation in Nepal. Natural Hazards and Earth System Sciences, v. 17, n. 3, p. 423-437, 2017.

[5] YASEEN, Z. M.; EBTEHAJ, I.; KIM, S.; SANIKHANI, H.; ASADI, H.; GHAREB, M. I.; BONAKDARI, H.; WAN, W. H. M.; AL-ANSARI, N.; SHAHID, S. Novel Hybrid Data-Intelligence Model for Forecasting Monthly Rainfall with Uncertainty Analysis. Water, v. 11, n. 1, p. 502, 2019

[6] ARAÚJO, V. Modelo baseado em dados para predição de inundação em bacia urbana de Campinas visando a construção de um sistema de alerta antecipado – Campinas, SP. 108p. Instituto de Geociências UNICAMP, Campinas, 2022.

[7] INPE. Instituto Nacional de Pesquisas Espaciais. Disponível em https://ftp.cptec.inpe.br/chuva/read_me/sos_chuva/, acessado em 05/06/2025.