



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS

VINICIUS ARAUJO

MODELO BASEADO EM DADOS PARA PREDIÇÃO DE INUNDAÇÃO EM BACIA
URBANA DE CAMPINAS VISANDO A CONSTRUÇÃO DE UM SISTEMA DE ALERTA
ANTECIPADO

CAMPINAS

2022

VINICIUS ARAUJO

MODELO BASEADO EM DADOS PARA PREDIÇÃO DE INUNDAÇÃO EM BACIA
URBANA DE CAMPINAS VISANDO A CONSTRUÇÃO DE UM SISTEMA DE ALERTA
ANTECIPADO

DISSERTAÇÃO APRESENTADA AO INSTITUTO
DE GEOCIÊNCIAS DA UNIVERSIDADE
ESTADUAL DE CAMPINAS PARA OBTENÇÃO
DO TÍTULO DE MESTRE EM GEOCIÊNCIAS NA
ÁREA DE GEOLOGIA E RECURSOS NATURAIS

ORIENTADORA: PROFA. DRA. ANA ELISA SILVA DE ABREU

ESTE EXEMPLAR CORRESPONDE À VERSÃO
FINAL DA DISSERTAÇÃO DEFENDIDA PELO
ALUNO VINICIUS ARAUJO E ORIENTADA
PELA PROFA. DRA. ANA ELISA SILVA DE
ABREU

CAMPINAS

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Geociências
Marta dos Santos - CRB 8/5892

Ar15m Araujo, Vinicius, 1995-
Modelo baseado em dados para predição de inundação em bacia urbana de Campinas visando a construção de um sistema de alerta antecipado / Vinicius Araujo. – Campinas, SP : [s.n.], 2022.

Orientador: Ana Elisa Silva de Abreu.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Geociências.

1. Aprendizado de máquina. 2. Hidrologia urbana. 3. Avaliação de riscos. 4. Inundações. I. Abreu, Ana Elisa Silva de, 1973-. II. Universidade Estadual de Campinas. Instituto de Geociências. III. Título.

Informações Complementares

Título em outro idioma: Data-driven model for flood temporal prediction aiming at building an early warning system for an urban watershed in Campinas

Palavras-chave em inglês:

Machine learning

Urban hidrology

Risk assessment

Floods

Área de concentração: Geologia e Recursos Naturais

Titulação: Mestre em Geociências

Banca examinadora:

Ana Elisa Silva de Abreu [Orientador]

Filipe Antonio Marques Falcetta

Paula Dornhofer Paro Costa

Data de defesa: 20-10-2022

Programa de Pós-Graduação: Geociências

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-5709-3143>

- Currículo Lattes do autor: <https://lattes.cnpq.br/2190016904408565>



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS

AUTOR: Vinicius Araujo

MODELO BASEADO EM DADOS PARA PREDIÇÃO DE INUNDAÇÃO EM BACIA URBANA DE
CAMPINAS VISANDO A CONSTRUÇÃO DE UM SISTEMA DE ALERTA ANTECIPADO

ORIENTADORA: Prof^a. Dr^a. Ana Elisa Silva de Abreu

Aprovado em: 20 / 10 / 2022

EXAMINADORES:

Prof^a. Dr^a. Ana Elisa Silva de Abreu - Presidenta

Prof^a. Dr^a. Paula Dornhofer Paro Costa

Prof. Dr. Filipe Antonio Marques Falcetta

*A Ata de Defesa assinada pelos membros da Comissão Examinadora consta no processo de
vida acadêmica do aluno.*

Campinas, 20 de outubro de 2022.

BIOGRAFIA

Vinicius Araujo, nascido em Araraquara, São Paulo, em 20 de março de 1995. Filho de Valdivino Geraldo Araujo e Rosangela Maria Calvo Araujo, cursou o ensino fundamental na escola “Prefeito Francisco Metidieri” em Nova Europa – SP, ensino médio na escola “ETEC – Professora Ana de Oliveira Ferraz” em Araraquara – SP e, posteriormente, em 2014 ingressou no curso superior de Geologia na Universidade Estadual de Campinas (UNICAMP) – SP. Em seus estudos, desenvolveu pesquisas de iniciação científica na área de petróleo e gás natural. Em 2018, estagiou em uma empresa de exploração mineral onde vivenciou as principais atividades do ramo. No mesmo ano formou-se entre os melhores alunos da turma. Em outubro de 2020, ingressou no programa de pós-graduação em Geociências no instituto de Geociências da Universidade Estadual de Campinas, onde desenvolveu um trabalho multidisciplinar voltado a previsão de inundações utilizando algoritmos de machine learning, em parceria com o Instituto de Pesquisa Tecnológica de São Paulo, submetendo-se a defesa da dissertação no dia 20 de outubro de 2022.

AGRADECIMENTO

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processos nº. 2017/50343-2 e nº. 2020/00058-2.

RESUMO

Diversas medidas de gerenciamento de inundações vêm sendo tomadas para mitigar os danos destes eventos no mundo todo, dentre elas podemos citar a geração de modelos hidrológicos baseados em dados para subsidiar sistemas de alerta antecipados. Tais modelos não necessitam da caracterização de diversos parâmetros físicos envolvidos no problema e requerem uma menor quantidade de dados quando comparados com modelos mais tradicionais. Neste contexto, o presente trabalho visa desenvolver um modelo baseado em dados para prever níveis d'água futuros na bacia do ribeirão Proença, localizada no município de Campinas (SP), que apresenta historicamente inundações que ocasionam riscos à população e danos ao patrimônio. Para tal, foram utilizados dados de nível d'água sub-horário e dados de radar meteorológico referentes ao período de novembro de 2014 até junho de 2019, além de registros de ocorrência de inundações coletados em meios digitais e no banco de dados da Defesa Civil do município. A partir da análise do cotograma gerado, três limiares foram aplicados (2,30m, 1,96m e 1,58m), com base no nível excedido ou igualado em 3, 6 e 10% do tempo analisado, respectivamente. A análise dos 145 eventos selecionados pelo limiar mais abrangente (1,58 metros) registrou a presença de 17 eventos de inundação na bacia, além de constatar que os níveis de 2,30 metros e 3,00 metros podem ser utilizados como níveis de atenção e alerta para a bacia, respectivamente. Para o desenvolvimento da pesquisa foi organizado um banco de dados multimodal a partir de dados reais, que foi disponibilizado no Repositório de Dados de Pesquisa da Unicamp. No que tange à criação do modelo baseado em dados para previsão de níveis d'água 30 minutos no futuro, foi utilizado o algoritmo XGBoost. Tal modelo apresentou para os dados de teste um RMSE de 0,284 metros e um NSE de 0,614, sendo assim classificado como um modelo satisfatório. Para os 3 eventos de inundação pertencentes aos dados de teste, o modelo foi capaz de prever níveis acima do nível de atenção em 2 deles, embora não tenha previsto níveis acima do nível de alerta. Deve-se destacar que, nas modelagens para diferentes tempos futuros, à medida que diminuimos o tempo futuro, melhoramos a previsão realizada pelo modelo. Analisando-se o *lag-time* entre a precipitação e a subida do nível d'água para os 145 eventos supracitados, verificou-se a presença de dois grupos distintos de valores (30 e 70 minutos), sendo o primeiro correlacionado a eventos de maior intensidade pluviométrica, enquanto o segundo relacionado a eventos de menor intensidade pluviométrica. Este resultado, atrelado ao desempenho do modelo, sugere que há uma elevada variabilidade intrínseca ao problema que não está sendo completamente capturada pelo modelo baseado em dados gerados pela presente pesquisa.

Palavras-chave: Aprendizado de máquina; hidrologia urbana; risco geológico; inundações bruscas.

ABSTRACT

Several flood management measures have been taken to mitigate the damages of flash-flood events worldwide. Among them we can mention the generation of data-driven hydrological models to support early warning systems, since such models do not have the need to characterize the several physical parameters involved in the modelling and require a smaller amount of data when compared to more traditional models. In this context, the present work aims at developing a data-driven model to predict future water levels in the Proença watershed, located in Campinas city, São Paulo state, Brazil, which historically presents floods that pose risks to the population and damage to property. To this end, sub-hourly water level data and weather radar data were used for the period from November 2014 to June 2019, in addition to flood occurrence records collected in digital media and in the Civil Defense database. From the analysis of the generated cotagram, three thresholds were applied (2.30m, 1.96m and 1.58m), based on the level exceeded or equalled in 3, 6 and 10% of the analysed time, respectively. The analysis of the 145 events selected by the most comprehensive threshold (1.58 meters) recorded the presence of 17 flood events in the watershed, in addition to verifying that the levels of 2.30 meters and 3.00 meters can be used as levels of attention and alert for the basin, respectively. For the development of the research, a multimodal dataset was organized from real data, which was made available at Unicamp's research data repository. XGBoost algorithm was used to model water levels with 30 minutes lead time using this dataset. This model presents for the test data an RMSE of 0.284 meters and an NSE of 0.614 and can be classified as a satisfactory model. For the 3 flood events belonging to the test data, the model was able to predict levels above the attention level for 2 events, but none above the alert level. Modeling other lead times revealed that as the lead time is decreased, the prediction quality improves. Analyzing the lag-time between precipitation and water level for the 145 events mentioned above, the presence of two distinct groups of values emerged: one with average lag time of 30 minutes and the other with average lag time of 70 minutes. The first one is related to events of greater rainfall intensity, while the second one is related to events of lower rainfall intensity. This result and the performance of the data-driven model, suggests that there is high variability intrinsic to the problem and that it has been only partially captured by the data-driven model developed in the present research.

Keywords: Machine Learning; urban hydrology; geological risk; flash-floods.

LISTA DE ILUSTRAÇÕES

Figura 3.1: Modelo esquemático dos processos de inundação, enchente e alagamento em um canal (Adaptado de: Carvalho et al. 2007)	19
Figura 3.2: Os quatro pilares para criação de um efetivo sistema de alerta antecipado (Adaptado de: Practical Action, 2016).....	27
Figura 4.1: Localização da bacia do ribeirão Proença em relação a cidade de Campinas (SP).....	39
Figura 4.2: Bacia do ribeirão Anhumas subdividida em sete sub-bacias (Modificado de: Adami et al. 2006).....	40
Figura 4.3: Mapa geológico da bacia do ribeirão Proença (Modificado de: Instituto Geológico, 2009).....	42
Figura 4.4: Mapa pedológico da bacia do ribeirão Proença (Adaptado de: Coelho et al. 2006).....	43
Figura 4.5: Mapa de declividade da bacia do ribeirão Proença.....	44
Figura 4.6: Pontos críticos de enchente/alagamentos na bacia do ribeirão Anhumas (Jacomazzi, 2015).....	47
Figura 4.7: Mapa de suscetibilidade a inundações e alagamentos na bacia do ribeirão Proença (Modificado de: Plano Diretor do Município de Campinas, 2018).....	48
Figura 5.1: Localização do sensor de nível d'água 4D-038 na bacia hidrográfica do ribeirão Proença.....	50
Figura 5.2: Dados de radar referente ao evento de 25/11/2018 às 16:50.....	52
Figura 5.3: Fluxograma representando os principais processos presentes no processamento de informação (Adaptado de: Fayaad et al. 1996).....	53
Figura 5.4: Sensor de nível d'água instalado na foz do ribeirão Proença.....	53
Figura 5.5: Fluxograma representando os principais processos realizados na etapa de tratamento e análise dos dados fluviométricos.....	54
Figura 5.6: Principais etapas realizadas no processo de criação do modelo final da pesquisa.....	64
Figura 6.1: Análise da quantidade de falhas nos dados de nível d'água registrados pelo sensor 4D-038.....	70
Figura 6.2: Nível médio mensal comparado com o nível médio geral.....	71
Figura 6.3: Cotagrama dos níveis máximos diários registrados pelo sensor instalado na foz do ribeirão Proença.....	74

Figura 6.4: Distribuição do número de dias e suas respectivas porcentagens em cada classe para cada um dos limiares aplicados.....	75
Figura 6.5: Lag-time versus nível d'água para cada um dos 145 eventos analisados.....	78
Figura 6.6: Precipitação média (azul) versus nível d'água registrado (vermelho).....	80
Figura 6.7: Números de eventos classificados como eventos de menor ou de maior intensidade pluviométrica e seus respectivos lag-time.....	81
Figura 6.8: Segmentação do banco de dados em dados de treino, validação e teste.....	87
Figura 6.9: Predições de nível d'água 30 minutos no futuro realizadas pelo modelo XGBoost 15VAR para os dados de treinamento, validação e teste.....	90
Figura 6.10: Dados reais versus dados previstos pelo modelo XGBoost 15VAR.....	91
Figura 6.11: Importância em porcentagem de cada uma das 15 variáveis utilizadas no modelo.....	92

LISTA DE TABELAS

Tabela 4.1: Principais parâmetros morfométricos calculados para a bacia do ribeirão Proença.....	45
Tabela 5.1: Principais características do sensor fluviométrico 4D-038.....	50
Tabela 5.2: Avaliação do desempenho do modelo com base nos valores de NSE obtidos (Adaptado de: Moriasi <i>et al.</i> 2007).....	61
Tabela 5.3: Procedimento realizado no cálculo da taxa de precipitação média.....	65
Tabela 5.4: Variáveis adicionadas ao modelo denominado de XGBoost 15VAR.....	65
Tabela 5.5: Dados hipotéticos para ilustrar a técnica empregada para geração da variável alvo.....	67
Tabela 6.1: Principais parâmetros estatísticos dos dados de nível corrigidos.....	72
Tabela 6.2: Principais parâmetros estatísticos dos dados de nível máximo diários.....	73
Tabela 6.3: Número de dias selecionados por cada um dos limiares empregados.....	74
Tabela 6.4: Data dos 145 dias selecionados pelo Limiar 1.....	77
Tabela 6.5: Tempo de concentração da bacia do ribeirão Proença calculado para diferentes equações.....	81
Tabela 6.6: Estrutura da base de dados unificada.....	83
Tabela 6.7: Resultado da avaliação do modelo XGBoost 15VAR para os dados de treino e validação.....	85
Tabela 6.8: Resultado da avaliação dos modelos gerados a partir da adição de novas variáveis defasadas no tempo.....	86
Tabela 6.9: Avaliação do modelo final XGBoost 15VAR aplicado aos dados de teste.....	88
Tabela 6.10: Desempenho de diferentes modelos para a previsão de distintos níveis d'água futuros.....	94

LISTA DE ABREVIATURAS E SIGLAS

DAEE – Departamento de Águas e Energia Elétrica

EM-DAT – *Emergency Events Database*

MAPE – *Mean Absolute Percentual Error*

NSE – *Nash-Sutcliffe Efficiency*

NaN – *Not a Number*

R² – Coeficiente de Determinação

RMSE – *Root Mean Square Error*

RNA – Rede Neural Artificial

SUMÁRIO

1. INTRODUÇÃO.....	15
2. OBJETIVOS	18
3. CONCEITOS BÁSICOS E REVISÃO BIBLIOGRÁFICA	19
3.1 Enchentes, inundações, enxurradas e alagamentos	19
3.1.1 Enchentes.....	19
3.1.2 Inundação	20
3.1.3 Enxurradas e alagamentos	24
3.2 Soluções para mitigar os eventos de inundação	24
3.3 Sistemas de alerta antecipado de inundações	26
3.3.1 Sistemas de alerta antecipado de inundações bruscas	29
3.4 Tipos de sistema de alerta antecipado de inundação	30
3.4.1 Sistemas de alerta antecipado de inundações baseados em estimativas de precipitação oriundos de radares meteorológicos	32
3.4.2 Sistemas de alerta antecipado de inundações, modelos hidrológicos e aprendizado de máquina	33
4. CARACTERIZAÇÃO DA ÁREA DE ESTUDO.....	38
4.1 Localização.....	38
4.2 Contextualização regional	38
4.2.1 Alto Anhumas.....	41
4.3 Clima e precipitação	41
4.4 Geologia e Pedologia.....	41
4.5 Declividade.....	44
4.6 Parâmetros morfométricos.....	44
4.7 Inundações e enchentes no Alto Anhumas	46
5. MATERIAIS E MÉTODOS	49
5.1 Materiais	49

5.1.1 Dados cartográficos	49
5.1.2 Dados fluviométricos.....	49
5.1.3 Dados de radar meteorológico	51
5.1.4 Notícias de inundações em meios digitais.....	52
5.2 Métodos	52
5.2.1 Tratamento e análise dos dados fluviométricos.....	52
5.2.2 Tratamento e análise dos dados de radar	55
5.2.3 Unificação dos bancos de dados	56
5.2.4 Variáveis auxiliares	56
5.2.5 Cálculo do tempo de concentração da bacia.....	57
5.2.6 Correlação cruzada entre as estimativas de precipitação obtidos a partir de radar meteorológico e os dados de nível d'água.....	58
5.2.7 Modelagem hidrológica: aprendizado de máquina, XGBoost e análise de desempenho	59
6. RESULTADOS	69
6.1 Nível d'água no ribeirão Proença – estabelecimento de níveis de alerta e de atenção.....	69
6.2 Estimativa do valor do <i>lag-time</i>	76
6.3 Estimativa do tempo de concentração da bacia	81
6.4 Base de dados unificada (BDU)	82
6.5 Modelagem XGBoost.....	84
6.6 Avaliação do Modelo Final (XGBoost 15VAR)	88
6.7 Modelagem para diferentes tempos futuros.....	93
7. CONCLUSÃO.....	95
8. REFERÊNCIAS BIBLIOGRÁFICAS	98

1. INTRODUÇÃO

A ocorrência de desastres naturais intensificou-se nos últimos anos e os mesmos estão cada vez mais frequentes no cotidiano da população, independente delas habitarem ou não áreas de risco.

De acordo com dados disponibilizados pela *Emergency Events Database* (EM-DAT), sintetizados pelo CRED (2020), entre os anos de 1980-1999 o número de desastres naturais reportados foi de 4212, enquanto entre os anos de 2000-2019 ocorreram mais de 7300 desastres, um aumento de aproximadamente 73%. Somente no ano de 2020 ocorreram um total de 389 desastres naturais, afetando mais de 98 milhões de pessoas e acarretando perdas econômicas de mais de 170 bilhões de dólares no mundo todo (CRED, 2021).

Embora grande parte da população associe o termo desastres naturais apenas a eventos de grande escala, tais como terremotos, tsunamis, ciclones e furações, o termo também abrange eventos mais localizados, dentre eles inundações, deslizamentos, erosão, entre outros (Tominaga *et al.* 2009).

Responsáveis por expressivos danos e perdas em âmbito social, econômico e ambiental, as inundações e enchentes estão cada vez mais recorrentes e os impactos ocasionados mais intensos. Segundo CRED (2021), do total de 389 desastres naturais registrados no ano de 2020, 201 eventos estão relacionados às inundações, um valor 23% acima da média do número de eventos obtido entre o período de 2000-2019.

No contexto do gerenciamento de inundações, uma das principais ações, dentre o contexto de medidas não estruturais, é a realização de uma acurada modelagem hidrológica, após calibração realizada com a utilização de dados medidos *in loco*, buscando entender o comportamento do rio e sua respectiva bacia hidrográfica, e utilizar esta ferramenta para fornecer os dados de entrada necessários para diversas aplicações, dentre elas podemos citar a organização de um sistema de alerta antecipado de inundações (Borga *et al.* 2011; Monte *et al.* 2016).

Tratando especificamente dos sistemas de alerta antecipado de inundações, são cada vez mais numerosos aqueles que utilizam o nível d'água dos rios como variável de monitoramento (Dotson e Peters, 1990; Corcoran *et al.* 2004; Gautam e Dulal, 2013; Nepal, 2014, Smith *et al.* 2017). Segundo Smith *et al.* (2016), a partir de um modelo hidrológico acurado é possível realizar previsões dos níveis d'água futuros com mais horas de antecedência e assim gerar alertas mais cedo. Entretanto, o comportamento dos sistemas

fluviais e do fluxo d'água não são facilmente calculáveis, principalmente pela complexidade dos processos físicos envolvidos e as variabilidades naturais associadas aos sistemas de fluxo do rio (Cheng *et al.* 2016; Yaseen *et al.* 2019).

De acordo com Adnan *et al.* (2019), os processos de fluxos apresentam uma relação não linear com a bacia hidrográfica e são afetados por diversos parâmetros, dentre eles: precipitação, uso e ocupação do solo e temperatura, tornando-os assim extremamente complexos e difíceis de prever. Desta forma, não há um modelo universal que seja efetivamente utilizado em todas as bacias existentes, uma vez que as características hidrológicas variam (Yaseen *et al.* 2019).

Devido a atual necessidade de compreender melhor os sistemas de fluxo de rio e os processos hidrológicos, ambos vitais no gerenciamento de recursos hídricos e consequentemente na criação de sistema de alerta antecipado, aliado ao crescente avanço tecnológico, novas pesquisas são necessárias para explorar e desenvolver modelos mais robustos, flexíveis e eficazes (Yaseen *et al.* 2019).

No contexto de criação de modelos mais robustos, flexíveis e eficazes, atrelado à sensível redução do custo das tecnologias de processamento computacional e acesso a grandes volumes de dados dos sensores nos últimos anos, os modelos gerados em dados tornaram-se comuns e eficientemente aplicáveis (Alizamir *et al.* 2018; Fotovatikhah *et al.* 2018; Yaseen *et al.* 2018; Yaseen *et al.* 2019).

Nesta pesquisa partiu-se da hipótese de que seria possível criar um modelo baseado em dados para prever as variações do nível do rio e gerar bases para implantação futura de um sistema de alerta antecipado em uma bacia urbana em que se dispunham de dados precários para a realização de modelagem física robusta. Desta forma, pretendia-se verificar se um modelo baseado em dados seria capaz de captar as características intrínsecas da bacia hidrográfica, de forma sintética, evitando-se assim a necessidade de definição dos diversos parâmetros que são necessários para a modelagem física.

A bacia hidrográfica do ribeirão Proença, objeto de estudo desta pesquisa, é uma sub-bacia do ribeirão Anhumas, localizado na cidade de Campinas (SP), que apresenta historicamente inundações que ocasionam riscos à população, danos ao patrimônio e possui sensor de nível do rio já instalado.

Foram utilizados dados reais de radares meteorológicos e de níveis de rio para compor o banco de dados do projeto. Estes dados serviram de base para caracterizar e modelar o comportamento hidrológico da bacia a partir de algoritmos de aprendizado de máquina.

Caracterizou-se diretamente a correlação dos níveis d'água com a pluviosidade, sem a necessidade de se estimar as vazões, uma vez que os algoritmos de aprendizado de máquina possuem a capacidade de captar as inter-relações existentes no sistema, não necessitando entender completamente os mecanismos de interação. Além disso, foram utilizados os níveis de rios, já que podem ser medidos facilmente e são mais úteis que a descarga na previsão de enchentes, uma vez que as transformações de vazão em nível d'água comumente apresentam erros e incertezas e as autoridades utilizam diretamente os níveis como parâmetro nos alertas de inundações (Yu *et al.* 2006).

São contribuições deste trabalho: a criação de códigos para unificação de dados hidrológicos brasileiros; a criação de uma base de dados unificada multimodal para a área de estudo contendo dados sub-horários e a criação de um modelo de referência para a área de estudo. O presente projeto está inserido na pesquisa denominada “Plano de desenvolvimento institucional na área de transformação digital: manufatura avançada e cidades inteligentes e sustentáveis” que contou com recursos da FAPESP (projeto 2017/50343-2 e 2020/00058-2) e apoio técnico do Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT).

2. OBJETIVOS

O objetivo desta pesquisa é desenvolver um modelo baseado em dados para prever níveis d'água futuros em uma pequena bacia hidrográfica intensamente urbanizada, a partir de um estudo de caso realizado na bacia do ribeirão Proença, formador do ribeirão Anhumas, na cidade de Campinas, SP.

Além disso, pretende-se gerar bases teóricas e técnicas para o desenvolvimento de um sistema de alerta antecipado de inundações na área de estudo a partir dos níveis previstos pelos algoritmos.

3. CONCEITOS BÁSICOS E REVISÃO BIBLIOGRÁFICA

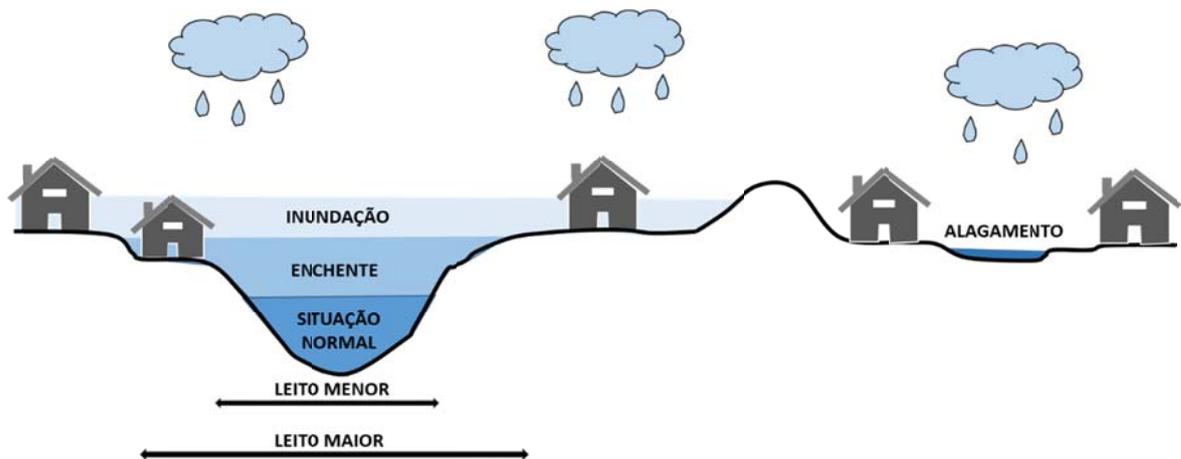
3.1 Enchentes, inundações, enxurradas e alagamentos

Segundo Goerl e Kobiyama (2005), há diversos termos associados às inundações no Brasil, dentre eles: cheia, enchente, enxurrada, alagamento, inundações, inundações urbanas, entre outros. Segundo os autores, devido à diversidade dos termos, fruto de traduções incertas e adaptações equivocadas, muitas vezes os eventos ocorridos são caracterizados erroneamente, prejudicando assim seus estudos. Abaixo iremos detalhar os principais termos empregados no Brasil, dando enfoque nas inundações, objeto de estudo da presente pesquisa.

3.1.1 Enchentes

Tucci *et al.* (1995) define as cheias ou enchentes como eventos hidrológicos naturais no qual o rio escoar, principalmente, para seu leito maior, possuindo uma probabilidade de ocorrência definida. Neste mesmo sentido, trabalhos como Goerl e Kobiyama (2005) e Carvalho *et al.* (2007) definem enchentes ou cheias como a elevação do nível d'água no canal até a altura de suas margens, sem que ocorra o transbordamento da água para as áreas adjacentes (Figura 3.1).

Figura 3.1: Modelo esquemático dos processos de inundação, enchente e alagamento em um canal (Adaptado de: Carvalho *et al.* 2007).



3.1.2 Inundação

Com relação a inundações, há diversos trabalhos que conceituam o evento mencionado (Oliveira, 1998; Tucci, 1999; Kron, 2002; Goerl & Kobiyama, 2005; Carvalho *et al.* 2007; Amaral & Ribeiro, 2009; Tang *et al.* 2020; Ma *et al.* 2021).

Segundo Oliveira (1998) as inundações são fenômenos da própria dinâmica fluvial no qual há o extravasamento das águas do curso d'água principal para as áreas marginais (Figura 3.1). Carvalho *et al.* (2007) define que as inundações consistem no transbordamento em um curso d'água para a planície de inundação adjacente e são fenômenos naturais e recorrentes que estão intimamente ligados a quantidade e intensidade da precipitação.

Trabalhos como o de Amaral & Ribeiro (2009), Tang *et al.* (2020) e Ma *et al.* (2021) complementam ainda a definição de Carvalho *et al.* (2007) e afirmam que tais fenômenos são comumente desencadeados por precipitações fortes e rápidas ou precipitações de longa duração e por eventos climáticos, tais como furacões e tornados e são intensificados pelas alterações ambientais e intenso processo de urbanização.

Aprofundando na caracterização das inundações e correlacionando-as com os processos de ocupação territorial, Tucci (1999) as classifica em três tipos: inundações ribeirinhas, urbanas e localizadas.

Segundo o autor, as inundações ribeirinhas são fenômenos relacionados a própria dinâmica natural da drenagem e não são condicionadas por fatores antrópicos. Já as inundações urbanas são ocasionadas devido ao processo de impermeabilização do solo fruto da urbanização na bacia, que ocasiona maiores volumes de água escoantes na superfície e consequentemente nos rios urbanos (Tucci, 1999).

Por fim, as inundações localizadas são localizadas no espaço e são comumente provocadas por fatores ligados à engenharia, tais como: estrangulamento de seções do rio por pontes e bueiros, assoreamento de drenagens por sedimentos, remanso, entre outros (Tucci, 1999).

Já Kron (2002) aborda os diferentes tipos de inundações que ocorrem após o extravasamento da água para fora do canal, sendo os tipos mais comuns: inundações costeiras, graduais e bruscas.

Segundo o autor, as inundações costeiras estão relacionadas aos processos que ocorrem em áreas costeiras, caracterizada quando as águas costeiras excedem a altura normal da maré alta (Kron, 2002; Goerl & Kobiyama, 2005).

Trabalhos como Vitousek *et al.* (2017) e Silva *et al.* (2020) abordam os impactos ocasionados pela subida do nível do mar na ocorrência de inundações costeiras. Segundo Vitousek *et al.* (2017), um aumento de 10 a 20 centímetros do nível do mar pode dobrar a frequência destes eventos extremos.

Em contrapartida, enquanto as inundações costeiras estão associadas aos processos exclusivos de áreas costeiras, as inundações graduais e bruscas estão relacionadas a velocidade do fenômeno de inundação, não possuindo assim um local específico de ocorrência (Kron, 2002; Goerl & Kobiyama, 2005).

Trabalhos mais recentes que abordam as inundações como temática central (Apel *et al.* 2016, Muthusamy *et al.* 2021, Tanaka *et al.* 2020) distinguem dois termos principais na classificação das inundações: inundações fluviais e pluviais.

Inundações fluviais ocorrem quando o nível d'água de um rio, lago ou canal excede sua capacidade e ultrapassa sua cota máxima, extravasando para as áreas adjacentes, fruto de precipitações excessivas por um longo período de tempo ou devido ao derretimento de neve (Maddox, 2014). Tal definição é similar as definições previamente mencionadas, contudo, há dois fenômenos principais denominados de inundações graduais e inundações bruscas.

Já o evento denominado de inundações pluviais ocorre quando a precipitação gera um processo de inundação que independe da água que transborda do canal, ou seja, este processo pode ocorrer em qualquer localização, não estando necessariamente associado a um canal ou corpo hídrico (Breinl, 2012; Maddox, 2014). Segundo Breinl (2012), as inundações pluviais são mais que inundações ocasionadas pela precipitação antes que a água entre nos rios. Por fim, Tanaka *et al.* (2020) evidencia que tais eventos ocorrem quando a intensidade da precipitação excede a capacidade de infiltração do solo e dos sistemas de drenagem urbana.

Os termos inundações graduais e bruscas serão detalhadas a seguir, com base no compilado gerado por Goerl & Kobiyama (2005) que reúne definições utilizadas por diferentes instituições e pesquisadores. Tais definições serão tratadas separadamente, uma vez que constituem a temática central da presente pesquisa.

3.1.2.1 Inundações graduais

Segundo Goerl & Kobiyama (2005), as inundações graduais ocorrem quando há um extravasamento das águas para a planície de inundação, quando há uma elevação gradual do nível d'água no canal. Além disso, são comumente ocasionadas por precipitações intensas

e/ou persistentes (Kron, 2002; Tucci & Bertoni, 2003; Choudhury *et al.* 2004; Goerl & Kobiyama, 2005).

Segundo Castro (2003), as inundações graduais estão correlacionadas as enchentes, uma vez que a água se eleva de forma lenta e previsível até a borda do canal e escoar-se gradualmente para as planícies. Segundo o autor, as inundações graduais são típicas de grandes rios e bacias hidrográficas, tais como o Amazonas e o Nilo, e são cíclicas e sazonais. O autor discute, ainda, que tais inundações estão mais relacionadas com períodos contínuos de precipitações do que com precipitações intensas e concentradas.

Como exemplo de estudos de inundações graduais temos o trabalho de Righi (2016), onde o autor, com base nos dados disponibilizados pelo Centro Universitário de Estudos e Pesquisas sobre Desastres (CEPED), destaca que entre os anos de 1991-2010 as inundações graduais afetaram aproximadamente 730 mil pessoas em todo o Estado do Rio Grande do Sul.

Neste mesmo contexto podemos citar o trabalho da PRACTICAL ACTION (2016) que apresenta as experiências obtidas na criação de alertas antecipados no Nepal para eventos de inundações graduais. Segundo os autores, no começo dos anos 2000 os monitoramentos das inundações ocorriam através de um sistema denominado de *Watch-and-Warn*, onde pessoas treinadas analisavam diariamente e visualmente o nível do rio a montante através de uma torre. Desta forma, apenas as comunidades que viviam a jusante possuíam um tempo de resposta frente ao evento. Contudo, é nítido que tal método funcione apenas em eventos no qual o nível d'água não suba de forma repentina.

3.1.2.2 Inundações bruscas

As inundações bruscas serão detalhadas neste tópico dando ênfase aos processos relacionados as áreas urbanas, temática central da pesquisa.

Tais inundações são o extremo oposto das inundações graduais, uma vez que ocorrem subitamente e são caracterizadas por possuir fluxos extremamente rápidos de água em áreas secas, ou ainda, rápidas ascensões do nível d'água em um canal, possuindo assim um curto tempo para declarar um estado de alerta para a área de ocorrência (Goerl & Kobiyama, 2005; Barbarotto Junior & Zuffo, 2017) e apresentando uma elevada variabilidade no tempo e espaço (Adnan *et al.* 2019a).

Segundo Saharia *et al.* (2015) e Trigo *et al.* (2016), as inundações bruscas estão intimamente correlacionadas a eventos de intensa precipitação e são reconhecidas

mundialmente como um dos eventos naturais que mais trazem danos à população. Tais inundações são internacionalmente conhecidas como *Flash Flood*.

FEMA (1981) define as inundações bruscas como inundações que apresentam uma rápida subida do nível d'água com velocidades anormais e são comumente resultados da combinação da intensa precipitação ocorrida em uma pequena e íngreme bacia de drenagem com elevada taxa de impermeabilização do solo.

Kron (2002) descreve as inundações bruscas como inundações com elevado potencial destrutivo que ocorrem em pequenas áreas e em poucas horas, produzidas por intensas precipitações sobre uma pequena área. Além disso, NWS/NOAA (2005) complementa a definição de Kron (2002), definindo um intervalo de tempo de 6 horas após a precipitação ou após uma quebra de barragem ou reservatório para ocorrência de inundações bruscas. Portanto, as inundações bruscas ocorrem no tempo próximas ao momento do evento que as ocasionou e assim, dificultam a emissão de alertas e alarmes (Goerl & Kobiyama, 2005).

Entretanto, não há um consenso entre Castro (2003) e Kron (2002). Enquanto Castro (2003) destaca que as inundações bruscas são típicas de relevo acidentado, Kron (2002) explicita que tais inundações não são exclusivas de terrenos íngremes, pois também ocorrem em áreas planas. Segundo Goerl & Kobiyama (2005), devido ao intenso processo de urbanização ocorrido nos últimos anos, as cidades de médio a grande porte, independente da declividade, apresentam locais com ocorrências de inundações com velocidades maiores, ou seja, são típicas de áreas urbanas.

Cabe destacar que trabalhos como o de Goerl & Kobiyama (2005) ressaltam a existência de outros fatores que devem ser considerados na definição da inundação brusca ou gradual, e a adoção do intervalo de tempo de 6 horas definida por NWS/NOAA (2004) pode variar para cada bacia, uma vez que cada qual possui uma resposta hidrológica para a quantidade de precipitação ocorrida.

Em sua definição para as inundações bruscas, Georgakakos (1986) sugere que, uma vez que as inundações bruscas ocorrem de maneira rápida e imprevisível, elas devem ser alertadas por centros locais de previsão, ou seja, devem possuir um sistema de monitoramento, previsão e alerta de ocorrência no local onde há possibilidade de ocorrer. Assim, com base no trabalho de Georgakakos (1986), Goerl & Kobiyama (2005) ressalta que um dos aspectos para diferenciar as inundações bruscas das graduais é a necessidade de centros de monitoramento, previsão e alerta locais no primeiro caso. Portanto, fica nítida a necessidade de um monitoramento especial e robusto em eventos de inundação brusca, uma

vez que a velocidade de subida do nível d'água é abrupta, diferentemente das velocidades observadas em eventos de inundações graduais.

3.1.3 Enxurradas e alagamentos

Além das enchentes e inundações, no Brasil são comumente adotados também os termos enxurradas e alagamentos, ambos usualmente empregados em áreas urbanas (Carvalho *et al.* 2007).

Segundo o Carvalho *et al.* (2007), as enxurradas são fenômenos associados a precipitações intensas e concentradas, caracterizadas pela rápida subida da quantidade de água que passa pelo canal. Contudo, apesar de possuírem conotação similar as inundações bruscas, as enxurradas podem ou não estar associadas a área onde os processos fluviais são predominantes, enquanto as inundações bruscas estão necessariamente associadas a um canal fluvial ou corpo hídrico.

Por fim, os alagamentos são caracterizados pelo acúmulo de água em determinados locais, comumente nas ruas e perímetros urbanos, ocasionados por sistemas de drenagem deficientes (Castro, 2003; Carvalho *et al.* 2007). Segundo Castro (2003), os alagamentos sofrem mais influência dos sistemas de drenagem deficientes do que das precipitações locais.

3.2 Soluções para mitigar os eventos de inundação

Segundo Müller (2011), diversos países, incluindo o Brasil, comumente adotam medidas de caráter localizado para solucionar os problemas de inundação, não observando assim o problema de maneira mais ampla. Como exemplo podemos citar o processo de ampliação dos canais, uma das soluções estruturais mais utilizadas nesta temática, pois reduzem consideravelmente os prejuízos locais das áreas afetadas. Contudo, a solução localizada de problemas, como no caso da ampliação dos canais nas áreas mais afetadas, acaba por gerar uma transferência de vazões, agravando assim as inundações mais a jusante (Canholi, 2005).

Considerando o fato que os desastres naturais afetam todos os continentes e que há um aumento dos impactos econômicos e sociais ocasionados pelos mesmos ao longo dos anos, diversos países buscaram soluções para prevenir, preparar e mitigar os riscos à sociedade (Alfieri *et al.* 2012). Tais soluções são agrupadas em dois grupos denominados de medidas estruturais e não estruturais (Tucci & Bertoni, 2003; Carvalho *et al.* 2007, Tominaga, 2015).

As medidas estruturais são aquelas onde se aplicam obras de engenharia, em geral de elevado impacto financeiro, social e ambiental, tais como obras de contenção de taludes, implantação de sistemas de drenagens, entre outros (Tucci & Bertoni, 2003; Carvalho *et al.* 2007; Tominaga, 2015). Segundo Carvalho *et al.* (2007), as medidas estruturais mais comuns relacionadas às inundações são aquelas que modificam o sistema fluvial, tais como canalizações, construção de diques e reservatórios.

Já as medidas não estruturais são aquelas relacionadas às políticas públicas, ao gerenciamento de inundações, ao planejamento urbano, às legislações, à educação ambiental e ao plano de defesa civil, buscando reduzir os danos ocasionados por tais eventos (Tucci, 2003; Sun *et al.* 2012; UNISDR, 2016).

De modo geral, as medidas não estruturais são consideradas menos onerosas e apresentam bons resultados, principalmente na prevenção de desastres naturais (Tucci, 2003; Kobiyama *et al.* 2006). Uma vez que as medidas estruturais são frequentemente onerosas e não proporcionam total eficácia na proteção contra as inundações (Faisal *et al.* 1999), as medidas não estruturais tornaram-se cada vez mais importantes e amplamente estudadas na última década (HILL *et al.* 2010; Dawson *et al.* 2011; Sun *et al.* 2012; Barcellos *et al.* 2017; Kundzewicz *et al.* 2019; Coutinho *et al.* 2020).

Dawson *et al.* (2011) descreveu um método para avaliar a efetividade e vantagem da aplicação de diversas medidas não estruturais em termos de redução de riscos ocasionados pelas inundações no estuário do rio Tâmis, localizado no Reino Unido. A partir da análise de diversos cenários hipotéticos futuros, os autores concluíram que a efetividade das medidas não estruturais está correlacionada com mudanças socioeconômicas e políticas de governança, e os riscos diminuíram consideravelmente devido a mudanças nas políticas de planejamento e na construção de propriedades resilientes na planície de inundação, ambas mudanças consideradas não estruturais.

Já Sun *et al.* (2012) expõem o projeto nacional criado pela China que visa prevenir os desastres ocasionados pelos eventos de *Flash Floods* e teve como passos iniciais a implantação de medidas não estruturais em um projeto piloto composto por 103 condados. Os autores detalham, ainda, oito medidas não estruturais aplicadas em todos os municípios do projeto para a prevenir e mitigar os riscos de inundação, dentre elas: planos de evacuação, monitoramento e instalação de alertas, treinamentos, entre outros. Graças ao sucesso obtido na aplicação das medidas supracitadas, o governo chinês expandiu a aplicação das medidas adotadas para mais de 1800 condados.

Barcellos *et al.* (2017) analisou se os métodos não estruturais empregados pela defesa civil do município de Duque de Caxias, Brasil, uma das cidades mais avançadas à época em termos de implementação de medidas não estruturais, estão alinhados com as principais ações de prevenção sugeridas pelo *Flash Flood Early Warning System Reference Guide* (FFEWSRG) (HILL *et al.* 2010), com base nas respostas de cada profissional entrevistado. Segundo o autor, mesmo sem o conhecimento prévio das práticas propostas pelo FFEWSRG a defesa civil do município já aplicava tais medidas.

Kundzewicz *et al.* (2019) destaca as principais medidas estruturais e não estruturais adotadas na China nos últimos anos, visando assim a prevenção e redução dos riscos ocasionados pelos fenômenos de inundações. Os autores destacam que, apesar das medidas estruturais serem tradicionais e frequentemente implantadas na China, há um crescente avanço na implementação de medidas não estruturais graças aos benefícios associados a tais medidas, principalmente os atrelados a criação de sistemas de alertas antecipados a inundações.

3.3 Sistemas de alerta antecipado de inundações

No contexto do gerenciamento de inundações, concentrando-se em medidas não-estruturais, uma das principais tarefas é a realização de uma acurada modelagem hidrológica, após calibração realizada com a utilização de dados medidos *in loco*, buscando entender o comportamento do rio e sua respectiva bacia hidrográfica, e utilizar esta ferramenta para fornecer os dados de entrada necessários para execução de um sistema de alerta antecipado de inundações (Borga *et al.* 2011; Monte *et al.* 2016). Segundo Solomatine & Ostfeld (2008), um dos focos da modelagem hidrológica é realizar uma previsão acurada das vazões, níveis d'água dos rios e precipitações.

De modo simples, os sistemas de alerta antecipado voltados para inundações possuem a função de analisar e disseminar as informações hidrológicas, preferencialmente em tempo real, com o objetivo de gerar alertas antecipados precisos acerca de um evento de inundação iminente (UNISDR, 2018; Li *et al.* 2021).

Durante a Segunda Conferência Mundial de Redução de Desastres em Kobe, Hyogo, Japão, no ano de 2005, um acordo assinado por diversos países denominado de “*Hyogo Framework for Action 2005-2015: Building the Resilience of Nations and Communities to Disasters*” foi estabelecido (UNISDR, 2005). Neste acordo, os países participantes se comprometeram a tomar medidas mais sólidas na prevenção e preparação frente aos desastres naturais, dando ênfase, principalmente, nas etapas de identificação,

avaliação e monitoramento dos riscos de desastres e na criação e melhoria de sistemas de alerta antecipado (UNISDR, 2006; NOAA, 2010).

Em 2009, especialistas do mundo todo se reuniram no MHEWS-II (*Second Expert's symposium on Multi-Hazard Early Warning Systems*) para discutir as boas práticas necessárias para construção de um eficiente sistema de alerta antecipado. Neste evento, os participantes destacaram a importância dos quatro pilares essenciais para criação de um sistema de alerta antecipado efetivo, conforme proposto pela UNISDR (2006), sendo eles: avaliação e conhecimento do risco; monitoramento e alerta de risco; informação, comunicação e disseminação do risco; e capacitação para resposta (Figura 3.2) (UNISDR, 2006; UNISDR, 2009; NOAA, 2010; Practical Action, 2016).

Figura 3.2: Os quatro pilares para criação de um efetivo sistema de alerta antecipado (Adaptado de: Practical Action, 2016).



Deve-se destacar que ambos os pilares são inter-relacionados e a falha em um dos componentes pode gerar uma falha no sistema como um todo (UNISDR, 2017, Brown *et al.* 2019). Além disso, todos os setores envolvidos deverão atuar na criação e atualização do sistema (Brown *et al.* 2019).

Segundo NOAA (2010), após o acordo de Hyogo, diversos esforços foram tomados para incorporar os sistemas de alerta antecipados nas estratégias de gestão de risco de desastres. Dentre tais esforços podemos destacar o denominado *Sendai Framework* de 2015, onde foram definidos sete objetivos globais para redução dos riscos de desastres, dentre

eles: “aumentar substancialmente a disponibilidade e o acesso de sistemas de alerta antecipado para as pessoas até 2030” (UNISDR, 2015; Javelle *et al.* 2019).

Neste cenário, merecem destaque os esforços da comunidade científica para criação e aplicação dos sistemas de alertas antecipados na previsão e mitigação de desastres relacionados a enchentes e inundações, uma vez que tais fenômenos estão cada vez mais recorrentes no mundo todo (CRED, 2021). Tais esforços podem ser evidenciados em trabalhos como Sun *et al.* (2012), Alfieri *et al.* (2013), Thielen-del Pozo *et al.* (2015).

Alfieri *et al.* (2013) avaliou o estágio inicial da aplicação e funcionamento de um sistema de previsão de inundações em escala global, aplicado em grandes bacias mundiais, denominado de GloFAS (*Global Flood Awareness System*). Neste estudo, o autor exalta que embora o sistema funcione bem apenas para grandes bacias, o mesmo conseguiu detectar e prever eventos de inundação com mais de 10 dias de antecedência. O autor destaca, ainda, a eficiência do sistema em prever as inundações ocorridas na bacia do rio Indus, Paquistão, em 2010. Segundo o autor, em 28 de julho de 2010, o sistema previu um pico de vazão e conseqüente subida do nível d’água com probabilidade de 100% de exceder o nível de alerta severo no dia 10 de agosto de 2010, treze dias antes do evento.

Já Thielen-del Pozo *et al.* (2015) detalha o desenvolvimento do *European Flood Alert System* (EFAS), um sistema de alerta desenvolvido para diminuir os riscos ocasionados por inundações em grandes bacias transnacionais da Europa. Segundo o autor, o EFAS foi capaz de prever inundações continentais e gerar benefícios para todas as esferas envolvidas no gerenciamento dos eventos.

No que tange ao Brasil, pode-se destacar o desenvolvimento de diversos projetos em grande escala, dentre eles: SAISP (Sistema de Alerta a Inundações em São Paulo), SISMADEN (Sistema de Monitoramento e Alerta de Desastres Naturais), CEMADEN (Centro Nacional de Monitoramento e Alertas de Desastres Naturais) e por fim o SACE (Sistema de Alerta de Eventos Críticos).

Entretanto, o Brasil ainda apresenta diversas dificuldades no que tange o desenvolvimento efetivo de sistemas de alerta, principalmente para pequenas bacias urbanas. Problemas como falta de técnicos para coletar e avaliar a qualidade dos dados, existência de postos não telemétricos, perdas de dados, não padronização de registros de dados e a inexistência de um banco de dados unificados, dificultam a criação de um sistema de alerta.

3.3.1 Sistemas de alerta antecipado de inundações bruscas

No contexto de desenvolvimento dos alertas antecipados para inundações, os eventos de inundações bruscas se tornaram um desafio global, uma vez que tais eventos possuem uma pequena escala espacial e temporal, suas previsões são tarefas difíceis (Sun *et al.* 2012) e estes eventos são os mais recorrentes em todo o mundo, afetando principalmente as áreas urbanas (Han *et al.* 2015). Segundo Sun *et al.* (2012), até mesmo os alertas mais sofisticados possuem a incapacidade de identificar a localização no espaço e tempo de tais eventos em pequenas bacias.

Desta forma, nos últimos anos diversos trabalhos buscaram gerar bases para criação de alertas que funcionem adequadamente em eventos de inundações bruscas, principalmente em áreas urbanas (Georgakakos, 1986; Marchi *et al.* 2010; Roy *et al.* 2019, Javelle *et al.* 2019).

Georgakakos (1986) sugere que, uma vez que as inundações bruscas ocorrem de maneira rápida, elas devem ser alertadas por centros locais de previsão, ou seja, devem possuir um sistema de monitoramento, previsão e alerta de ocorrência no local onde há possibilidade de ocorrer.

Já Marchi *et al.* (2010), no estudo dos 25 maiores eventos de inundações bruscas ocorridos na Europa nos últimos 20 anos, constatou que menos da metade dos eventos foram devidamente registrados a partir dos métodos de medidas convencionais, reforçando assim as sugestões de Georgakakos (1986).

Roy *et al.* (2019) avaliou a frequência dos eventos de inundações bruscas na região de Haor, Bangladesh, buscando determinar um novo nível de alerta para tais eventos. Segundo os autores, os níveis previamente estabelecidos para a área foram determinados com foco nos eventos extremos de inundações que ocorrem nos períodos de monções e os mesmos não servem para os eventos de inundações bruscas, eventos recorrentes na região nos demais períodos e que ocasionam severos danos à agricultura local. Com base nos resultados obtidos, os autores concluem que o nível de alerta para os eventos supracitados é muito menor, média de 1,5 m abaixo dos utilizados para os eventos extremos nos períodos de monções, podendo chegar a mais de 5 m em algumas estações.

Ademais, os autores apontam que uma subtração de aproximadamente dois terços do valor atual do nível para monções é suficiente para utilizar como nível de alerta para os eventos de inundações bruscas.

Apesar das dificuldades encontradas na criação dos sistemas de alerta para eventos de inundações bruscas no passado, trabalhos mais recentes destacam a sua efetividade

atual na previsão, fruto do avanço tecnológico, atrelado a crescente coleta e disponibilidade de dados (Liu *et al.* 2018, Kundzewicz *et al.* 2019; Kron & Cheng, 2019; Li *et al.* 2021, Ritter *et al.* 2021).

Liu *et al.* (2018) destacam que, embora o número de eventos de inundações bruscas tenha aumentado ao longo dos anos, o número de mortes ocasionadas vem diminuindo, evidenciando assim o papel que as medidas não estruturais, principalmente a criação de alertas antecipados, possuem na proteção da população.

Kundzewicz *et al.* (2019) destacam que entre os anos de 2009-2015 foram implementados sistemas de monitoramento e de alerta antecipado em mais de 2000 condados da China, condados que sofrem comumente eventos de inundações bruscas, com base nos excelentes resultados obtidos em trabalhos anteriores.

Já Li *et al.* (2021) relatam a importância do sistema de alerta de inundações FAS4 (*Flood Alert System*) durante os eventos de inundações ocasionados pelo furacão Harvey em 2017. Segundo os autores, uma vez que sistema foi capaz de prever com considerável antecedência os eventos de inundação, ações foram tomadas para mitigar os possíveis danos.

Ritter *et al.* (2021) avaliaram dois eventos recentes de inundações bruscas que causaram diversas fatalidades na Europa e concluíram que o método empregado pelos autores, denominado de ReAFFINE, associado ao sistema funcional já existente na Europa (ERICHA), seria capaz de prever os eventos e identificar as áreas impactadas.

3.4 Tipos de sistema de alerta antecipado de inundação

Segundo Li *et al.* (2021) os sistemas de alerta antecipado de inundação possuem a capacidade de auxiliar diferentes clientes, regiões e nacionalidades, uma vez que variam desde o básico monitoramento dos níveis d'água do rio até os sistemas de previsão mais complexos que utilizam dados de radares e sofisticadas técnicas de modelagem computacional para gerarem tempos de respostas cada vez maiores.

Dentre os diferentes tipos de sistemas de alerta antecipado para inundações, há duas abordagens comuns na bibliografia: as que utilizam apenas a leitura visual do nível d'água para emissão do alerta e as mais modernas que utilizam diversas outras fontes de dados integradas, tais como descarga, previsão meteorológica, umidade relativa do solo, dentre outras.

Para contrastar as duas abordagens, podemos citar o trabalho de Smith *et al.* (2017), que resume como era o sistema de alerta no passado no Nepal e quais os benefícios e desafios observados na implementação de um sistema mais sofisticado. Segundo os autores, o

alerta antecipado no passado era realizado com base na leitura visual de nível d'água por pessoas treinadas em um medidor a 30 quilômetros de distância da comunidade em estudo, fornecendo assim 2 a 3 horas de tempo de resposta frente aos eventos. Após alguns anos, a mesma comunidade implementou um sistema de alerta baseado em um modelo hidrológico específico para a área e avaliou dois eventos de inundação.

Os autores destacam que, com a implementação do sistema, foi observado um tempo de resposta adicional de 4 a 5 horas, totalizando assim 6 a 7 horas de previsão antecedente do evento. Apesar de parecer pequeno o tempo adicional, ele é extremamente significativo na mitigação dos riscos à vida e aos meios de subsistência (Smith *et al.* 2017).

Contudo, devemos destacar que as abordagens visuais, muitas vezes, possuem baixo tempo de resposta e conseqüentemente não conseguem prever eventos de inundações bruscas, dificultando assim seu emprego principalmente em áreas urbanas.

Desta forma, no contexto atual do gerenciamento de inundações dos municípios, cada vez mais deixa-se de lado a medição visual dos eventos e empregam-se sistemas mais sofisticados que utilizam diversas fontes de dados integradas.

Segundo Solomatine *et al.* (2008), um dos focos do gerenciamento de inundação é a geração de modelos hidrológicos para prever acuradamente as vazões, níveis d'água dos rios e as precipitações, fornecendo assim os dados necessários para execução dos sistemas de alerta. Smith *et al.* (2017) complementam, ainda, que a partir de um modelo hidrológico acurado é possível realizar previsões dos níveis d'água futuros com mais horas de antecedência e assim gerar alertas mais cedo.

Tratando especificamente dos sistemas de alerta antecipado de inundações, são cada vez mais numerosos os sistemas que utilizam o nível d'água dos rios e a pluviosidade como variável de monitoramento e entrada nos modelos hidrológicos (Gautam & Dulal, 2013; Adhikari *et al.* 2014; Smith *et al.* 2017, Roy *et al.* 2019).

Gautam & Dulal (2013) determinaram níveis para emissão de alertas nos seis maiores rios do Nepal, gerando assim bases para a instalação de sistemas de alertas antecipados no país. Os autores concluem que a utilização de tais níveis, associados as informações em tempo real de precipitação e nível d'água nos rios, permitem a previsão de inundações com 6 até 20 horas de antecedência.

Adhikari *et al.* (2014) determinaram níveis de aviso e perigo para dois principais rios localizados no Nepal com base em dados de descarga, nível d'água e precipitação aplicados a um modelo hidráulico denominado HEC-RAS.

3.4.1 Sistemas de alerta antecipado de inundações baseados em estimativas de precipitação oriundos de radares meteorológicos

Sabe-se que a precipitação medida através de pluviômetros é consideravelmente acurada, porém sua área de cobertura é relativamente pequena, restringindo suas medidas próximas a área do sensor (Rocha Filho, 2010). Além disso, a comum baixa densidade de estações nas bacias hidrográficas em geral e as relativas incertezas presentes, tanto no instrumento quanto na interpolação dos sensores, dificultam a predição de enchentes e inundações (Ly *et al.* 2013, Li *et al.* 2021), necessitando assim de novas fontes de dados.

De modo geral, as bacias urbanas comumente apresentam um elevado grau de impermeabilização e conseqüentemente menor tempo de concentração, gerando assim inundações e enchentes com baixo tempo de resposta. Creutin & Borga (2003, apud Rocha Filho *et al.* 2013) relatam que o baixo tempo de resposta frequentemente observado em bacias urbanas requer uma elevada resolução temporal e espacial dos dados obtidos.

Segundo Li *et al.* (2021), as previsões de tempo meteorológico possuem auxílio de técnicas de radar há mais de 50 anos e nas últimas duas décadas passaram a ser empregadas amplamente nas aplicações hidrológicas. No que tange ao Brasil, com enfoque na região metropolitana da cidade de São Paulo, o monitoramento atmosférico começou a ser efetuado pelo Radar Meteorológico de São Paulo (RSP) em 1988 por meio da aquisição de um radar meteorológico banda S (Rocha Filho *et al.* 2011).

No contexto de inundações urbanas, comparando os sistemas de alerta de inundações que utilizam sensores instalados nos rios urbanos e os baseados em dados de radares, este último possui maior resolução espacial e temporal, gerando assim um maior tempo de previsão (Li *et al.* 2021) e conseqüentemente melhores resultados na previsão dos fenômenos.

São inúmeros os trabalhos que utilizam estimativas de precipitação oriundos de radares meteorológicos na previsão de inundações e enchentes (Rocha Filho *et al.* 2011; Li *et al.* 2021; Rapant & Kolejka, 2021)

Li *et al.* (2021) destacam o sistema de alerta de inundações desenvolvido pelo *Texas Medical Center* e a *Rice University* para uma sub-bacia intensamente urbanizada do Texas, a partir da aplicação de dados de precipitação obtidos pelo radar NEXRAD nos modelos hidrológicos disponíveis para a área. Conforme os autores, o sistema foi posto em prova em diversos eventos de inundação, dentre eles o furacão Harvey em 2017, e foi capaz de fornecer informações de alertas confiáveis, rápidos e precisos.

Rapant & Kolečka (2021) desenvolveram um método que transforma a distribuição da precipitação sobre a área monitorada, obtida a partir de radares meteorológicos, em seu potencial escoamento ao longo de toda a rede de drenagem local. De acordo com os autores, o método foi aplicado na República Checa em escala local e nacional e obteve excelentes resultados na previsão dos possíveis eventos.

3.4.2 Sistemas de alerta antecipado de inundações, modelos hidrológicos e aprendizado de máquina

A maioria dos sistemas de alerta antecipado de inundações modernos possuem como peça central um modelo hidrológico aplicado para a bacia em estudo, capaz de prever os fenômenos de inundação com base nos dados de entrada, que são normalmente informações sobre precipitação.

Embora diversas propostas de sistemas de alerta antecipado para inundações tenham surgido nos últimos anos, vide os exemplos supracitados, suas aplicações são, muitas vezes, efetivas apenas para o local no qual o sistema foi desenvolvido, principalmente pelo fato que os processos de fluxo d'água apresentam uma relação não linear com a bacia hidrográfica e são afetados por diversos parâmetros locais, dentre eles: precipitação, uso e ocupação do solo, temperatura (Adnan *et al.* 2019). Desta forma, não há um modelo universal que seja efetivamente utilizado em todas as bacias existentes, uma vez que as características hidrológicas variam (Yaseen *et al.* 2019).

Além disso, nos últimos anos, a criação de modelos hidrológicos para centros urbanos recebeu significativa atenção dos cientistas, principalmente pelo fato de ser um grande desafio gerar modelos de previsão de níveis d'água acurados para pequenos espaços temporais, fator essencial para tais localidades (Nguyen *et al.* 2021a).

Em geral, a metodologia atual de análise do fluxo de canal se subdivide em: modelos baseados em processos físicos (*Physically based models*) e modelos orientados a dados (*Data-driven models*) (Solomatine *et al.* 2008; Adnan *et al.* 2019; Mehdizadeh *et al.* 2019).

O primeiro necessita de diversos parâmetros de entrada, tais como: intensidade e distribuição da precipitação, características fisiográficas da bacia hidrográfica, uso e ocupação de solo, entre outros, e um robusto banco de dados hidroclimáticos para caracterizar os processos hidrológicos locais (Ochoa-Tocachi *et al.* 2016; Adnan *et al.* 2019). A dificuldade em se obter os parâmetros supracitados geralmente limita a geração de um modelo apropriado

e preciso para a área em estudo, restringindo assim a sua utilização (Ragetti *et al.* 2014; Adnan *et al.* 2019; Mehdizadeh *et al.* 2019).

Ademais, tais modelos são frequentemente limitados pela falta de dados existentes e confiáveis (Yaseen *et al.* 2019) e geralmente usam premissas simplificadas para os processos hidrológicos (Nguyen *et al.* 2021a).

Já o segundo são modelos matemáticos construídos com base nas observações de séries temporais históricas (Solomatine, 2005; Mehdizadeh *et al.* 2019) e comumente utilizam apenas as informações hidrológicas mais relevantes do passado para realizar as previsões hidrológicas necessárias (Nguyen *et al.* 2021a). Tais modelos são amplamente utilizados e relativamente fáceis de aplicar em diversas áreas e não carecem de um banco de dados complexo e robusto (Mehdizadeh *et al.* 2019). Ademais, como não necessitam caracterizar diversos parâmetros difíceis de obter, tais modelos estão se tornando progressivamente mais populares (Solomatine *et al.* 2008).

Nos últimos anos, com a sensível redução dos custos das tecnologias de processamento computacional e o acesso a grandes volumes de dados de sensores, os modelos hidrológicos baseados a dados gerados a partir de algoritmos de aprendizado de máquina tornaram-se comuns e eficientemente aplicáveis, principalmente pelo fato de possuírem a capacidade de aprender os detalhes mais complexos envolvidos nos processos hidrológicos (Nguyen *et al.* 2021a).

Em termos genéricos, aprendizado de máquina é um conjunto de algoritmos capazes de identificar padrões em uma elevada quantidade de dados e gerar previsões, realizando assim tarefas específicas de forma autônoma (CETAX, 2022).

Conforme Yaseen *et al.* (2019), tratando especificamente do emprego de algoritmos de aprendizado de máquina aplicados a modelagem hidrológica, tais algoritmos possuem a capacidade de identificar as inter-relações complexas e intrínsecas entre os parâmetros que controlam a transformação de precipitação em vazão, não necessitando assim entender completamente os mecanismos de interação.

Podemos citar como exemplo inúmeros trabalhos onde diferentes algoritmos de aprendizado de máquina obtiveram desempenho confiável em termos de previsão de níveis d'água e vazões (Yaseen *et al.* 2018; Meng *et al.* 2019; Lima & Scofiel, 2021; Nguyet *et al.* 2021; Nguyen *et al.* 2021b; Li *et al.* 2021).

Yaseen *et al.* (2018) criaram um modelo que emprega o algoritmo de *Extreme Learning Machine* (ELM) denominado de *Wavelet Extreme Learning Machine* (WA-ELM) para prever o fluxo de um rio em um ambiente semiárido. Segundos os autores, além de ser

capaz de prever o fluxo com até um mês de antecedência, o modelo apresentou melhores resultados na previsão, redução de 65% no erro quadrático médio, quando comparado com o simples emprego do algoritmo ELM.

Meng *et al.* (2019) desenvolveram modelos para previsão de vazões baseado em informações oriundas de imagens de radar, aplicados a diferentes tipos de algoritmo de *Support Vector Machine* (SVM). Conforme os autores, os algoritmos empregados obtiveram excelentes resultados na previsão de vazões futuras e desempenharam melhores resultados que os obtidos por meio de algoritmos de Rede Neural Artificial (RNA).

Lima & Scofiel (2021) relatam um estudo realizado em uma pequena bacia hidrográfica intensamente urbanizada na cidade de Campos do Jordão (SP), Brasil, no qual um modelo hidrológico baseado em RNA é empregado para prever, com uma hora de antecedência, se o nível d'água irá ultrapassar o nível de atenção local. Para tal, utilizaram dados de precipitação e nível disponibilizados pelo CEMADEN referentes aos anos de 2015 a 2019 e concluíram que, apesar da ocorrência de diversos falsos positivos, a metodologia empregada tem potencial para ser empregada como uma ferramenta que auxilia nas tomadas de decisões.

Nguyen *et al.* (2021b) utilizaram estimativas de precipitação oriundas de radares meteorológicos para prever a precipitação em treze pluviômetros instalados em cinco cidades localizadas na Coréia do Sul, a partir de um modelo gerado com base em um algoritmo denominado de *Long Short-Term Memory* (LSTM). Segundo os autores, apesar do modelo subestimar os eventos extremos de precipitação, ele apresentou excelentes resultados e conseqüentemente é passível de ser empregado para previsões de precipitação. Ademais, ele apresentou melhores resultados quando comparados a modelos mais tradicionais, dentre eles: *Multiple Linear Regression* (MLR), *Multivariate Adaptive Regression Splines* (MARS), *Multi-layer Perceptron* (MLP) e RNA.

Nguyet *et al.* (2021) compararam diversos algoritmos de *Deep Learning* aplicados a séries temporais sequenciais de precipitação com um modelo conceitual denominado de HBV, a fim de verificar a precisão dos modelos de *Deep Learning* na previsão das vazões. Além disso, o autor buscou avaliar, ainda, se há melhores resultados na previsão quando há, além das séries temporais de precipitação, a utilização de dados de vazões como dados de entrada do algoritmo. Segundo os autores, ao comparar os resultados obtidos pelos diferentes modelos, constatou-se que eles tiveram uma ótima performance na previsão das vazões e que o emprego apenas dos dados de precipitação obteve os melhores resultados.

Li *et al.* (2021) relatam a melhoria no desempenho observado ao empregar algoritmos de aprendizado de máquina em um sistema de alerta antecipado para uma bacia urbanizada localizada na cidade de Houston, Texas. Segundo os autores, diferentes tempos de previsão foram testados, onde a previsão de 30 minutos a frente foi o melhor resultado estatístico observado, incrementando assim o tempo de resposta para as autoridades. Os autores complementam, ainda, que para aumentar o tempo de previsão foram combinados diversos métodos de aprendizado de máquina em um sistema abrangente, onde cada qual atua simultaneamente e o menor tempo de resposta obtido com considerável acurácia, será o valor de previsão observado.

Dentre os diferentes modelos baseados em algoritmos de aprendizado e máquina, merecem destaque os modelos baseados em Árvores de Decisão (*Tree-based Models*), uma vez que os mesmos são computacionalmente mais baratos para processar e não requerem uma distribuição estatística específica. Segundo Nguyen *et al.* (2021a), enquanto os modelos baseados em redes neurais (diversos supracitados) apresentam apenas os resultados, sem retornar uma clara explicação dos processos que os geraram, os modelos baseados em árvores fornecem uma interpretação de cada etapa executada até o referido resultado.

No contexto dos modelos baseados em árvores de decisão, um dos algoritmos mais empregados atualmente para previsão de níveis d'água e/ou vazão é denominado de *Extreme Gradient Boosting* (XGBoost). Trabalhos como: Venkatesan & Mahindrakar (2019), Nguyen *et al.* (2021b) e Sanders *et al.* (2022), destacam o potencial e a eficácia do algoritmo XGBoost.

Venkatesan & Mahindrakar (2019) desenvolveram um modelo precipitações - vazão baseado no algoritmo XGBoost para previsão de inundações com 5 horas de antecedência, a partir de dados de precipitação e vazão horários. Tais previsões foram comparadas com as previsões gerados por outros dois algoritmos a saber: *Random Forest* (RF) e *Support Vector Machine* (SVM). Segundo os autores, enquanto os modelos gerados por RF e SVM apresentaram, na etapa de validação, uma raiz do erro quadrático médio (*Root Mean Square Error* – RMSE), um dos parâmetros comumente empregados para avaliar o quanto o modelo erra para mais ou para menos, de aproximadamente 114 m³/s e 78 m³/s, respectivamente, para previsão de 5 horas de antecedência, o modelo gerado por XGBoost apresentou um valor de 18,9 m³/s, evidenciando assim a superioridade deste modelo.

Nguyen *et al.* (2021b) desenvolveram dois modelos híbridos baseados em XGBoost para previsão de até 6 horas de antecedência de níveis d'água em uma bacia hidrográfica urbana localizada na Coreia do Sul, a partir de dados horários de precipitação e

nível d'água. Os resultados obtidos foram comparados com os resultados obtidos por dois outros modelos: *Classification and Regression Tree* (CART) e *Random Forest* (RF). Conforme os autores, os dois modelos híbridos gerados apresentaram resultados similares quando comparados entre si e resultados superiores quando comparados aos outros dois modelos utilizados. Deve-se destacar que, para todos os modelos empregados, à medida que os modelos previam valores com mais horas de antecedência, os erros aumentavam. Entretanto, o aumento do valor do erro foi menor nos dois modelos baseados em XGBoost.

Sanders *et al.* (2022) demonstram a eficácia do modelo baseado em XGBoost, na previsão de níveis d'água para diversos cenários futuros em intervalos de 5 minutos, totalizando uma previsão de até 2 horas de antecedência. Para validar a eficácia do modelo, os autores avaliaram a capacidade do mesmo de prever dois eventos históricos de inundação ocorridos na bacia.

Além disso, os autores compararam os resultados gerados a partir de dois modelos com diferentes dados de entrada, sendo que o primeiro utilizou apenas dados de sensores localizados no interior da área de estudo e o segundo utilizou, além dos sensores na área de estudo, dados de sensores localizados fora da área. A partir dos resultados obtidos, os autores notaram que para previsões próximas de até 30 minutos, o modelo obteve ótima performance, enquanto para previsões intermediárias, por volta de 60 minutos, o modelo apresentou uma boa performance. Entretanto, para previsões acima de 60 minutos o modelo tendeu a subestimar os picos de níveis d'água, além de apresentar um atraso no momento de subida do nível. Ademais, assim como no trabalho de Nguyen *et al.* (2021), à medida que o tempo de previsão aumentava, os erros também aumentavam. Por fim, deve-se destacar que, ao adicionar dados de sensores localizados fora da bacia em estudo, o modelo não apresentou melhorias significativas.

4. CARACTERIZAÇÃO DA ÁREA DE ESTUDO

4.1 Localização

Delimitada pelas coordenadas aproximadas 7.462.900 a 7.467.700 mN e 289.100 a 293.000 mE (UTM, Zona 23), a bacia do ribeirão Proença, objeto do presente de estudo, possui uma área de 11,5 km² e está localizada na região central da cidade de Campinas, estado de São Paulo (Figura 4.1, A).

Note que a bacia se encontra na região central do município supracitado, possuindo intensa taxa de urbanização, poucas áreas verdes e grande parte do Ribeirão Proença encontra-se canalizado.

Segundo Jacomazzi (2015), o ribeirão Proença recebe água de grande parte dos bairros centrais do município e a maioria do seu trecho encontra-se tamponada, uma vez que o ribeirão atravessa partes centrais e densamente urbanizadas da cidade de Campinas (Figura 4.1A, B).

Ademais, em algumas regiões as faixas de Áreas de Preservação Permanente (APP's) foram transformadas em vias expressas marginais, gerando assim trechos nos quais as inundações são frequentes (Jacomazzi, 2015).

4.2 Contextualização regional

Em escala regional, o ribeirão Proença deságua no ribeirão Anhumas, e está inserido no contexto das grandes bacias dos rios Piracicaba, Capivari e Jundiá, pertencentes assim a Unidade de Gerenciamento de Recursos Hídricos 5 (UGRHI 5).

Com o intuito de melhor caracterizar espacialmente a bacia do ribeirão Proença no âmbito da bacia do ribeirão Anhumas, iremos adotar a subdivisão proposta por Adami *et al.* (2006), que subdividiu a bacia do ribeirão Anhumas nas seguintes sub-bacias, de montante para jusante: Mato Dentro, Alto Anhumas, São Quirino, Taquaral, Médio Anhumas, Monte D'este e Rio das Pedras (Figura 4.2).

Desta forma, será abordado detalhadamente a porção denominada de Alto Anhumas, porção na qual está inserida a sub-bacia do ribeirão Proença.

Figura 4.1: (A) Localização da bacia do Ribeirão Proença (área de estudo) em relação a cidade de Campinas (SP) e (B) *zoom* na bacia. Destaque para os sensores já instalados na bacia (estrela vermelha) (modificado de Ribeiro & Abreu, 2021).

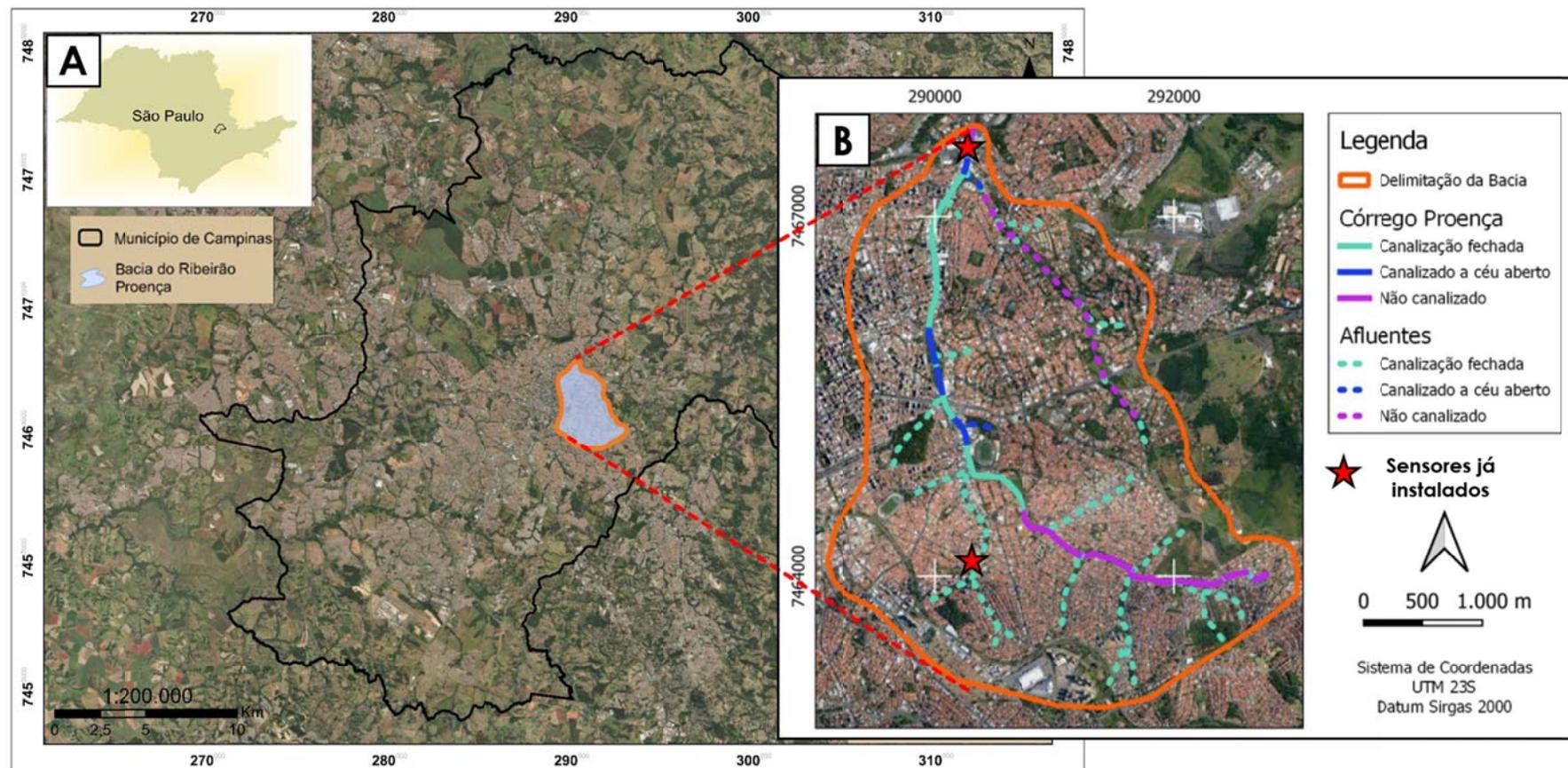
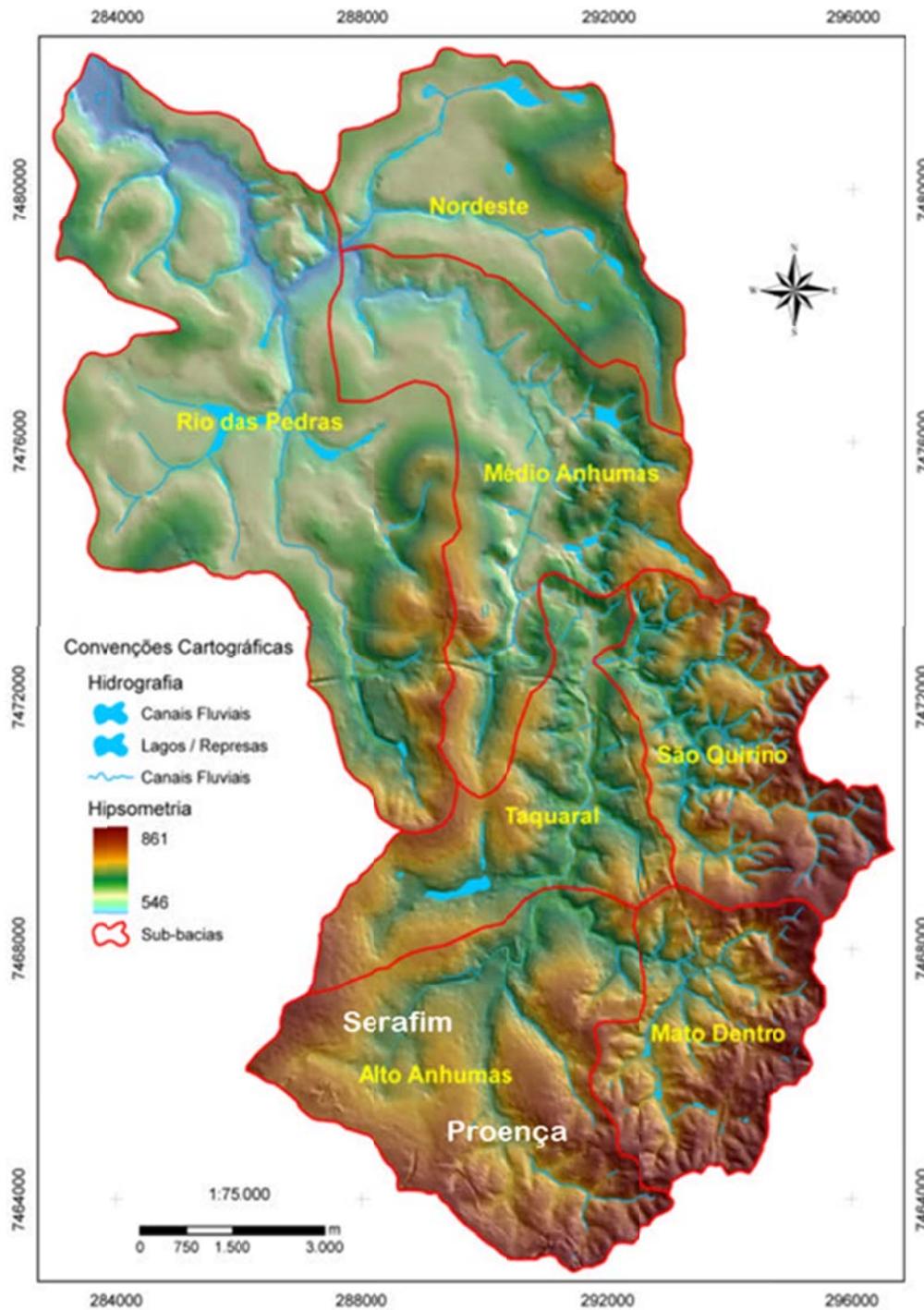


Figura 4.2: Bacia do ribeirão Anhumas subdividida em sete sub-bacias: Rio das Pedras, Nordeste, Médio Anhumas, Taquaral, São Quirino, Alto Anhumas e Mato Dentro. Destaque para os afluentes Serafim e Proença, localizados na sub-bacia do Alto Anhumas (Adaptado de: Adami *et al.* 2006).



4.2.1 Alto Anhumas

O Alto Anhumas está localizado na região montante da bacia do ribeirão Anhumas, abrangendo a área central e densamente urbanizada da cidade de Campinas, compreendendo as bacias do ribeirão Proença, área de estudo, e do canal Saneamento (Córrego Serafim ou Orosimbo Maia) (Jacomazzi, 2015) (Figura 4.2).

Segundo Jacomazzi (2015), a bacia do ribeirão Proença e do canal Saneamento possuem características similares: densa área urbanizada (85,1 e 94,6%, respectivamente); áreas verdes e APPs intensamente suprimidas; e consolidada infraestrutura com forte sobrecarga na macrodrenagem implantada.

Tais fatores condicionam a ocorrência de inundações e enchentes frequentemente relatadas em diversos trabalhos (Briguenti, 2005; Adami *et al.*, 2006; Vilela Filho, 2006; Jacomazzi, 2015; Castellano, 2016; Ignácio, 2020), ocasionando assim danos à população local.

4.3 Clima e precipitação

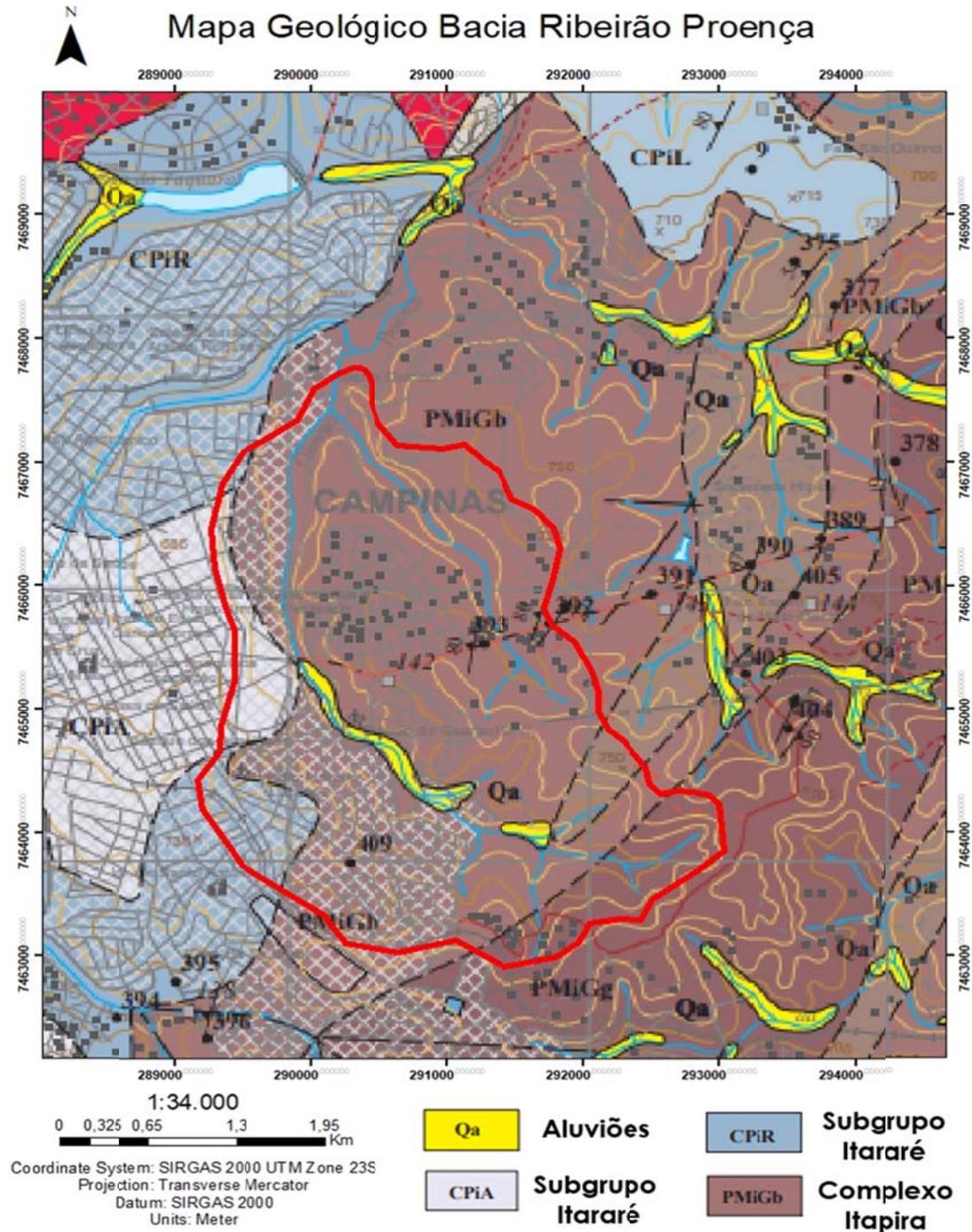
A região no qual a bacia do ribeirão Proença se insere, com base na classificação de Köppen, corresponde ao tipo climático Cwa, clima subtropical úmido com invernos amenos e de baixo volume de precipitações e verões quentes e mais úmidos (Castellano, 2016; Ignácio, 2020). Segundo as autoras supracitadas, o município de Campinas sofre influência da Massa Tropical Atlântica, Massa Polar Atlântica e da Zona de Convergência do Atlântico Sul, este último aumentando o volume de precipitação no município e arredores.

Conforme Ignácio (2020), baseada nos dados climatológicos do CEPAGRI referentes ao ano de 2018, a temperatura média de Campinas é de 21,4°C e as precipitações médias se aproximam de 1372 mm, sendo grande parte deste volume precipitado entre os meses de dezembro a março.

4.4 Geologia e Pedologia

No que tange a geologia da área de estudo, baseado no mapa geológico do município de Campinas (Instituto Geológico, 2009), a bacia do ribeirão Proença está inserida majoritariamente nas rochas gnáissicas bandadas do Complexo Itapira (PMiGb) e em menores porções a leste ocorrem rochas areníticas e ritmitos do Subgrupo Itararé (Figura 4.3). Além disso, são mapeados também aluviões próximos aos canais de drenagem.

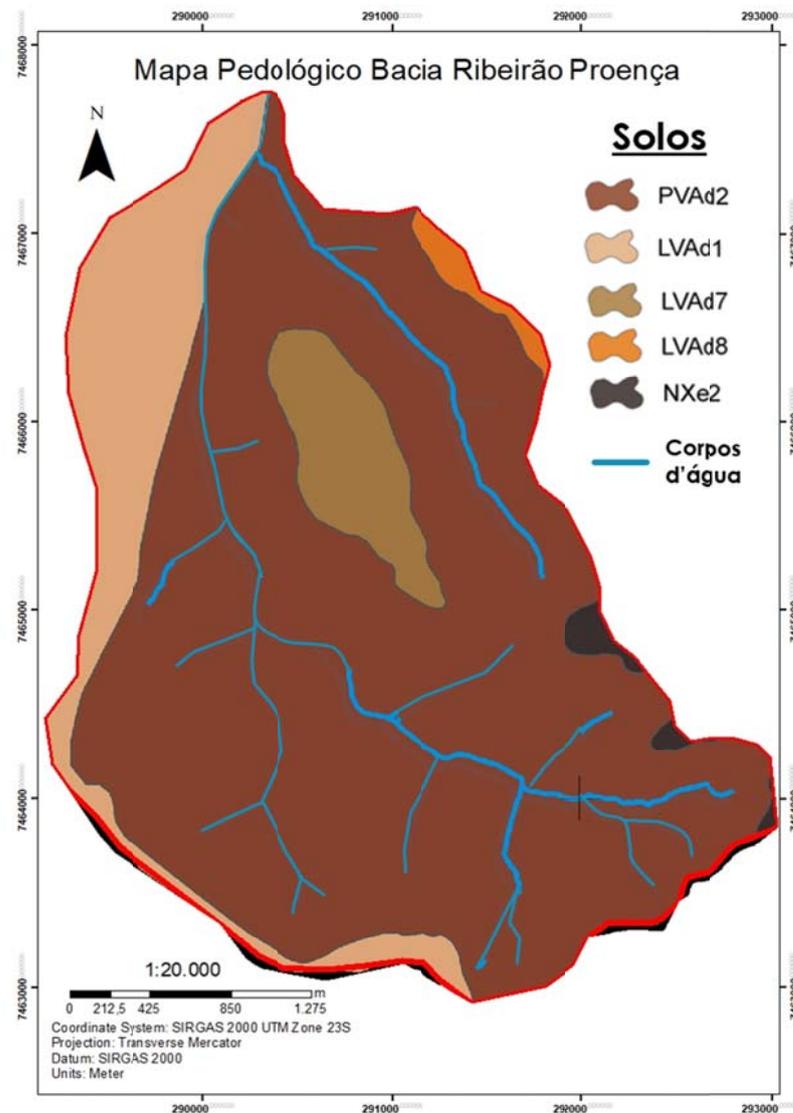
Figura 4.3: Mapa Geológico da bacia do Ribeirão Proença (Adaptado de: Instituto Geológico, 2009).



Deve-se destacar que, grande parte das rochas supracitadas não afloram, uma vez que a bacia em estudo se encontra em uma das áreas mais urbanizadas e antropizadas do município.

O solo predominante na bacia hidrográfica em estudo, baseado nos levantamentos realizados por Coelho *et al.* (2006), é o Argissolo vermelho-amarelo (PVA_d2), seguido pelo Latossolo vermelho-amarelo (LVA_d1) e pelo Latossolo vermelho-amarelo e vermelho, sendo ambos os solos supracitados distróficos e de textura argilosa a média (Figura 4.4).

Figura 4.4: Mapa Pedológico da bacia do Ribeirão Proença, baseado no levantamento realizado pelo Projeto Anhumas (Adaptado de: Coelho *et al.* 2006).



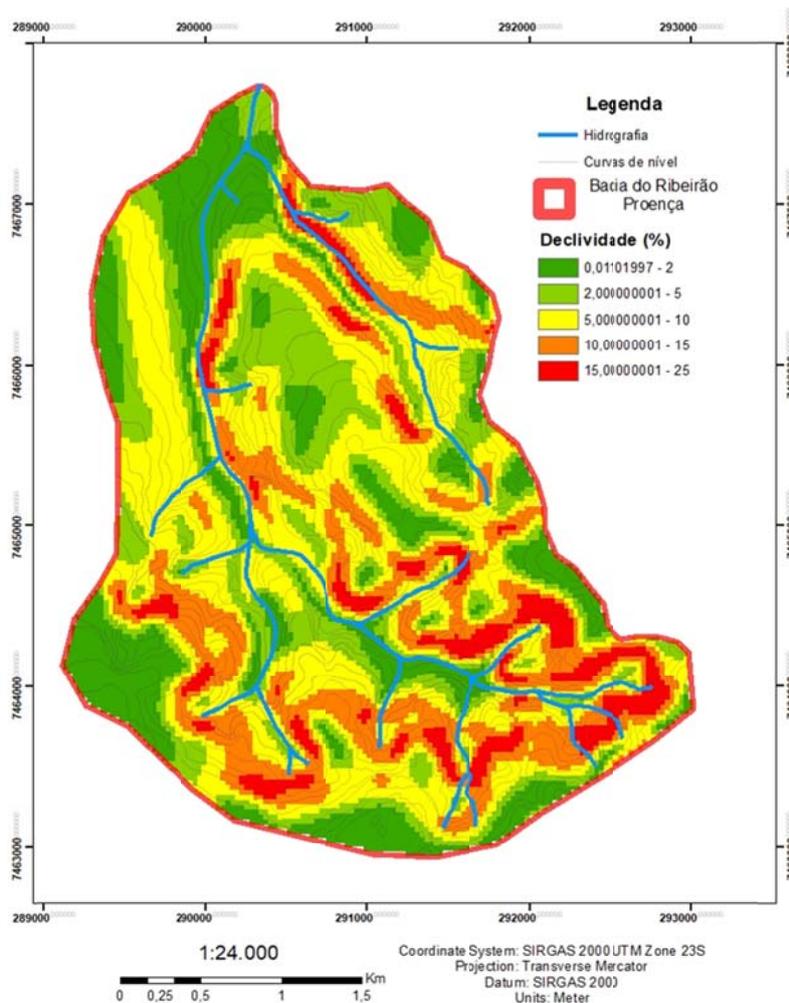
Nota-se também que em menores porções temos o Latossolo vermelho-amarelo associado a Nitossolos Háplicos (LVA_d8) e Nitossolo Háplico eutrófico (NXe2) (Figura 4.4).

4.5 Declividade

A Figura 4.5 ilustra o mapa de declividade da área de estudo. Os subgrupos de declividade foram subdivididos em cinco, conforme a tendência de ocorrência de inundações e enchentes proposta por Prina & Tretin (2014). Segundo os autores, declividades entre 0 a 2% possuem uma propensão muito alta a ocorrência de inundações, seguida de alta para 2 a 5%, média 5 a 10%, baixa 10 a 15% e por fim muito baixa acima de 15%.

Deve-se destacar que o mapa de declividade foi gerado a partir do Modelo Digital de Elevação (MDE) do Estado de São Paulo em escala 1:50.000, obtido a partir da Coordenadoria de Planejamento Ambiental (Secretaria do Meio Ambiente do Estado de São Paulo, 2013).

Figura 4.5: Mapa de Declividade da bacia do ribeirão Proença.



4.6 Parâmetros morfométricos

A bacia hidrográfica do ribeirão Proença possui uma área de 11,50 km², perímetro de 14,40 km e amplitude altimétrica de 109 metros. Já o canal principal possui comprimento de 5,44 km, sendo seu comprimento vetorial na ordem de 4,62 km. Por fim, o comprimento total dos canais é de 17,83 km.

Tais parâmetros foram obtidos utilizando-se ferramentas do programa ArcGis e a base de dados da plataforma GoogleEarth. A partir desses parâmetros foram calculados a densidade de drenagem, o índice de sinuosidade, o índice de circularidade e o coeficiente de compacidade com base na metodologia empregada em trabalhos como: Christofolletti (1981); Tucci & Clarke (1997); Brubacher *et al.* (2011); Santos *et al.* (2016); e Costa *et al.* (2017). A Tabela 4.1 exibe os parâmetros morfométricos da bacia do ribeirão Proença.

Tabela 4.1: Principais parâmetros morfométricos calculados para a bacia do ribeirão Proença.

Parâmetros	Valor	Referência
Área	11,50 km ²	-
Perímetro	14,40 km	-
Amplitude Altimétrica	109 m	-
Comprimento do Canal Principal	5,44 km	-
Comprimento Total dos Canais	17,83 km	-
Comprimento Vetorial do Canal Principal	4,62 km	-
Densidade de Drenagem	1,55 km/km ²	Brubacher <i>et al.</i> (2011); Santos <i>et al.</i> (2016);
Índice de Sinuosidade	1,27	Santos <i>et al.</i> (2016); Costa <i>et al.</i> (2017)
Índice de Circularidade	0,69	Brubacher <i>et al.</i> (2011); Santos <i>et al.</i> (2016); Costa <i>et al.</i> (2017)
Coeficiente de Compacidade	1,19	Brubacher <i>et al.</i> (2011); Santos <i>et al.</i> (2016); Costa <i>et al.</i> (2017)

Deve-se destacar que os parâmetros supracitados possuem uma íntima correlação com os eventos de inundação. No que tange a área calculada e a baixa densidade de drenagem, a bacia do ribeirão Proença, em uma análise isolada de somente ambos os parâmetros, possui uma baixa suscetibilidade a inundações. Contudo, parâmetros como a distância entre a nascente e a foz (comprimento vetorial do canal), baixa sinuosidade, elevado índice de circularidade e o coeficiente de compacidade, indicam uma susceptibilidade natural elevada da bacia às inundações.

Ademais, além dos valores dos parâmetros morfométricos indicarem, por si só, uma elevada susceptibilidade às inundações, há a intensa ocupação urbana da área, propiciando ainda mais a ocorrência de tais eventos.

4.7 Inundações e enchentes no Alto Anhumas

As inundações e enchentes na região central de Campinas são problemas recorrentes e comuns na história de ocupação urbana do município.

Briguenti (2005) ressalta que os problemas ocasionados pelas inundações no município já eram relatados pela imprensa local no final dos anos de 1980, principalmente em locais como a avenida Princesa D'Oeste, localizada na porção central da bacia do Ribeirão Proença.

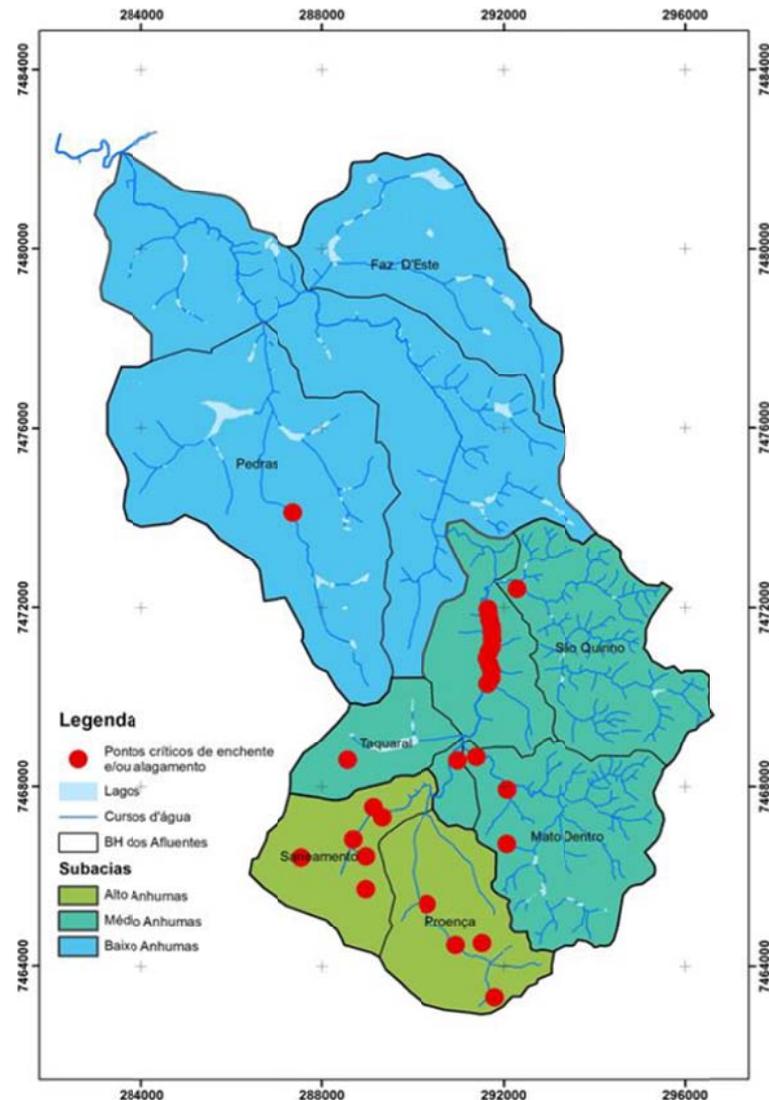
Já Vilela Filho (2006) destaca que os eventos de inundações e enchentes se intensificaram a partir da década de 1970. Segundo o autor, a implantação do modelo urbanístico sanitário adotado para a região central de Campinas levou a construção de obras de drenagem, canalizações e retificações nos canais, favorecendo assim a ocupação de terras destinadas aos cursos d'água e à inundação. Tal abordagem levou a um aumento dos riscos ocasionados pelas enchentes e escorregamentos à infraestrutura urbana e a população de Campinas.

Coelho (2006) e Ferreira *et al.* (2019) evidenciam que as principais avenidas localizadas na região central de Campinas, dentre elas a avenida Princesa D'Oeste, inseridas no contexto da bacia do Ribeirão Proença, possuem suscetibilidade alta às inundações. Conforme Ferreira *et al.* (2019) entre os anos de 2003 e 2019 sete óbitos foram registrados na área de estudo devido às inundações.

Jacomazzi (2015) relata que as enchentes do Alto Anhumas, especialmente a do ribeirão Proença, possuem características típicas das inundações ocasionadas pela excessiva impermeabilização do solo fruto da intensa urbanização na região central do município.

Segundo o autor, a sobrecarga da infraestrutura de drenagem leva ao extravasamento da água para as margens, inundando assim as edificações localizadas na planície de inundação. O autor destaca, ainda, os principais trechos críticos para inundações na região (Figura 4.6).

Figura 4.6: Pontos críticos de enchentes/alagamentos na bacia do ribeirão Anhumas. Note as diversas ocorrências de enchentes/alagamentos (em vermelho) ao longo do ribeirão Proença (Jacomazzi, 2015).

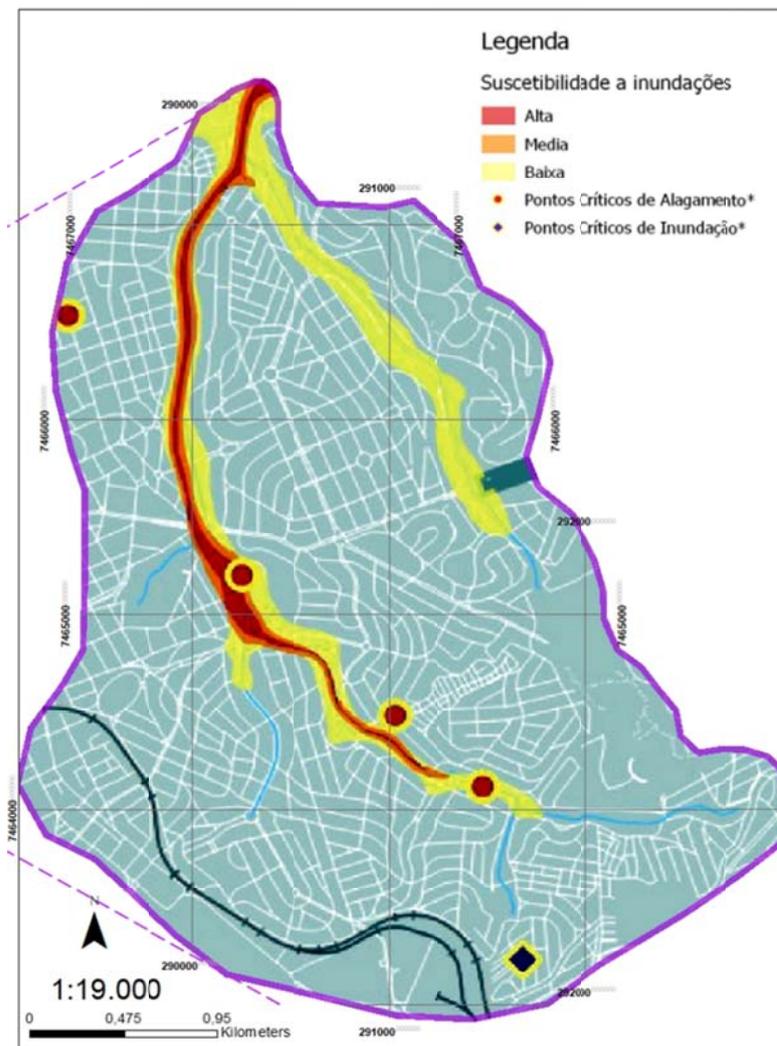


Levantamentos realizados pelo Plano Diretor do Município de Campinas (2018) destacam que, dos 14 pontos críticos de alagamentos mapeados para o município de Campinas, 8 estão inseridos na sub-bacia do Alto Anhumas (Figura 4.7). Já com relação aos

pontos críticos de inundações, o município possui no total 28 pontos mapeados, 5 destes pertencentes a bacia do ribeirão Proença (Figura 4.7).

Segundo o plano diretor supracitado, os problemas mencionados acima possuem duas causas principais: obsoleto sistema de drenagem e ocupação indiscriminada dos fundos de vale e áreas de inundação.

Figura 4.7: Mapa de suscetibilidade a inundações e alagamentos na bacia do ribeirão Proença. As manchas em cores quentes representam as áreas suscetíveis aos fenômenos mencionados e os pontos em vermelho e azul representam locais críticos de alagamento e inundação, respectivamente (retirado e modificado de Plano Diretor do Município de Campinas, 2018).



5. MATERIAIS E MÉTODOS

5.1 Materiais

Neste tópico serão descritas as principais fontes de dados utilizadas, os tipos de dados obtidos em cada fonte e suas principais características.

5.1.1 Dados cartográficos

A primeira etapa do presente trabalho foi a obtenção de dados referentes a delimitação da bacia hidrográfica do ribeirão Proença. Para tal, foram utilizados os produtos gerados por Ribeiro & Abreu (2021).

Com os arquivos de delimitação da bacia hidrográfica em mãos, foi possível determinar a geologia, pedologia e declividade da área de estudo. Tal procedimento foi realizado no *software* ArcGis 10.8 e os dados de entrada foram, respectivamente: mapa geológico do município de Campinas (Instituto Geológico, 2009), levantamento pedológico realizado pelo Projeto Anhumas (Coelho *et al.* 2006) e Modelo Digital de Elevação (MDE) do Estado de São Paulo em escala 1:50.000 (São Paulo, 2013).

5.1.2 Dados fluviométricos

Os dados fluviométricos, assim como os dados de temperatura utilizados na presente pesquisa foram fornecidos pelo Departamento de Águas e Energia Elétrica do Estado de São Paulo (DAEE). Trata-se de dados coletados pelo sensor 4D-038 a cada 2, 5 ou 10 minutos, a depender do mês analisado. Este sensor corresponde a um transdutor de pressão instalado na parede lateral do ribeirão Proença. Ele está associado a um sensor barométrico, para correção das variações da pressão atmosférica.

A Figura 5.1 exibe a localização do sensor fluviométrico (código 4D-038) que está instalado na foz da bacia hidrográfica do ribeirão Proença desde o final do ano de 2014. Já a Tabela 5.1 exibe as principais características deste sensor.

Figura 5.1: Localização do sensor de nível d'água 4D-038 (em amarelo) na bacia hidrográfica do ribeirão Proença.

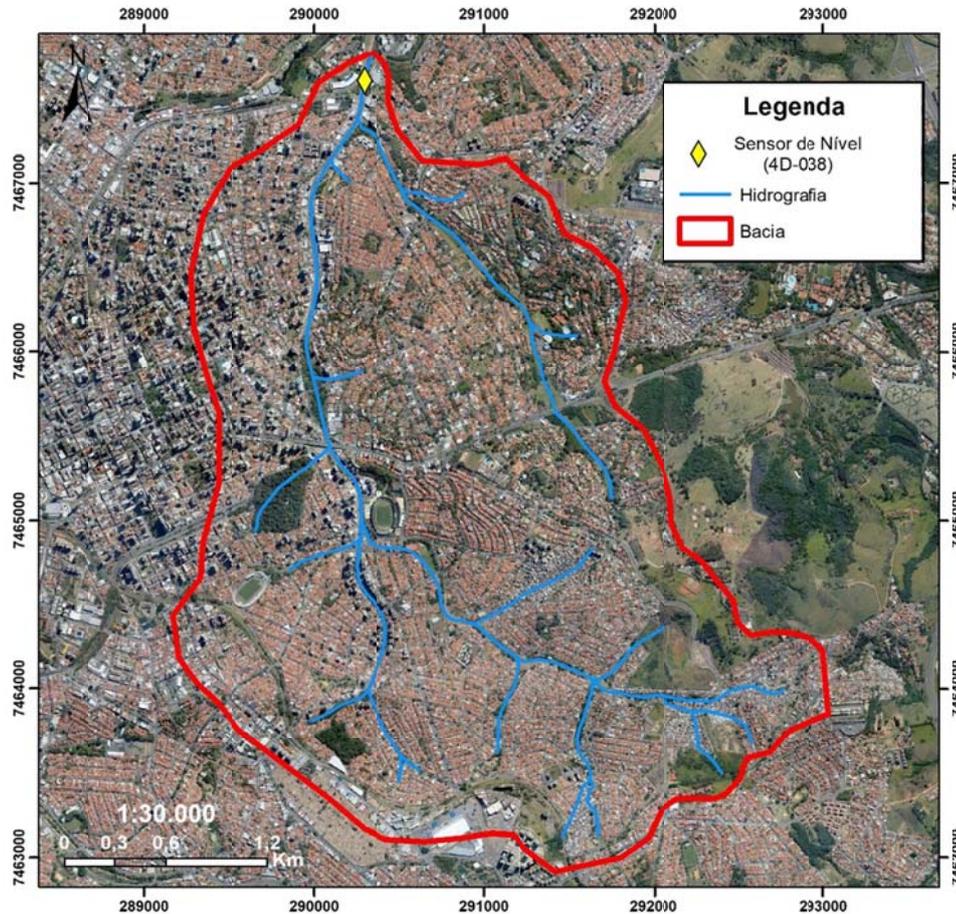


Tabela 5.1: Principais características do sensor fluviométrico 4D-038.

Fonte de Dados	Código	Série Temporal	Taxa de Aquisição
DAEE	4D-038	Nov. 2014 – Jul. 2022	Sub-Horária (10, 5 e 2 min)

Com base na Tabela 5.1, nota-se que a série temporal do sensor 4D-038 abrange dados coletados até julho de 2022, entretanto o presente trabalho irá utilizar apenas dados até junho de 2019, uma vez que os dados posteriores a essa data não são passíveis de serem correlacionados com os dados anteriores. Ademais, os dados de nível d'água coletados até 2019 apresentam valores negativos, uma vez que a régua linimétrica encontrava-se instalada acima do nível d'água, necessitando assim correções. As correções empregadas encontram-se descritas no tópico de métodos (item 5.2 adiante).

Deve-se destacar que, uma vez que na bacia hidrográfica do ribeirão Proença há apenas uma única estação fluviométrica, tal estação serviu como definidora do período de estudo.

5.1.3 Dados de radar meteorológico

As estimativas de precipitação registradas pelo Radar Meteorológico de São Paulo foram fornecidas pelo meteorologista Kleber Lopes da Rocha Filho, funcionário da FCTH e responsável pela operação do radar meteorológico instalado em Salesópolis (SP). Segundo Rocha Filho *et al.* (2015), este radar é do tipo banda S Doppler de dupla polarização e foi instalado no ano de 2014.

Os dados foram fornecidos na forma de imagens em formato .tif. Tratam-se de dados específicos para as três horas que antecedem e sucedem os eventos de subida de nível d'água no ribeirão Proença, identificados a partir da análise dos dados coletados pelo sensor fluviométrico, totalizando 6 horas de cobertura. Foram solicitados apenas dados de radar correspondentes aos dias em que se registraram os 10% mais altos valores de nível d'água no sensor 4D-038 ao longo de período de novembro de 2014 até julho de 2022.

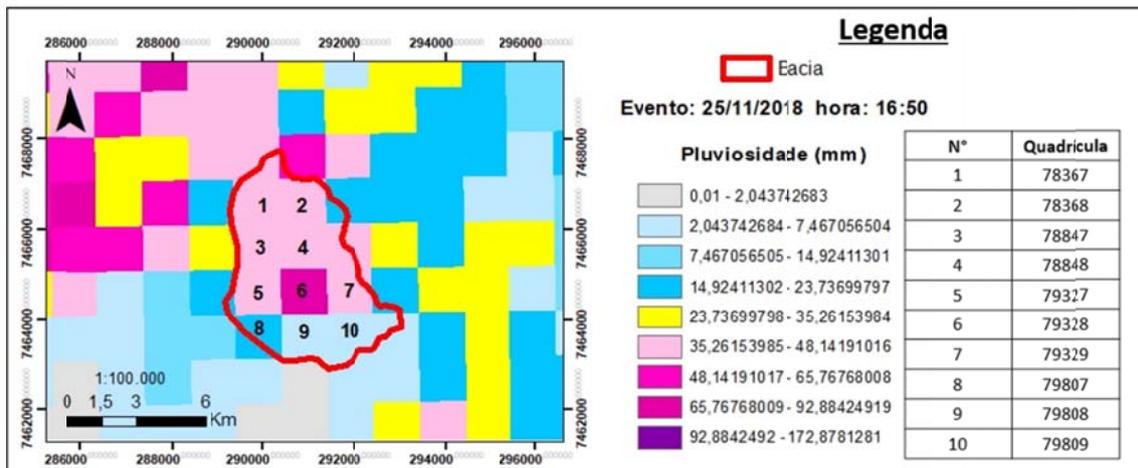
Como tais dados são obtidos a cada cinco minutos, cada conjunto que antecede um evento de subida de nível é composto por 72 imagens subsequentes. Deve-se lembrar que os dados coletados pelo radar são registros de refletividade e são transformados posteriormente em estimativas de precipitação a partir da utilização da relação Z-R de Marshal e Palmer (1948), conforme ilustrado em Rocha Filho *et al.* (2011). Ademais, os dados fornecidos a cada 5 minutos representam a estimativa de precipitação para um total de 60 minutos a frente, logo o valor da precipitação registrada a cada 5 minutos foi dividido por 12.

A Figura 5.2 exemplifica os dados obtidos pelo radar (pós aplicação da relação Z-R) referente a precipitação ocorrida no dia 25 de novembro de 2018, às 16 horas e 50 minutos, tratados no *software ArcGis* 10.8. Note que os dados obtidos pelo radar abrangem uma área maior que a bacia hidrográfica em estudo e a precipitação medida é registrada para cada quadrícula, cada qual com dimensões de 1 x 1 km².

Baseado na Figura 5.2, temos que a bacia do ribeirão Proença possui um total de 10 quadrículas do radar inseridas em seu interior, cada qual contendo um valor de precipitação. Tais quadrículas são numeradas de forma ordenada de oeste pra leste, sendo que

na bacia do ribeirão Proença, as quadrículas do oeste recebem numeração com final 7, as centrais têm numeração com final 8 e aquelas situadas mais a leste possuem numeração terminada em 9.

Figura 5.2: Dados de radar referente ao evento de 25 de novembro de 2018 às 16:50 tratados no *software ArcGis 10.8*, com destaque para as quadrículas pertencentes a bacia hidrográfica do ribeirão Proença e a respectiva pluviosidade registrada em cada uma.



5.1.4 Notícias de inundações em meios digitais

Notícias de inundações, alagamentos e fenômenos correlatos foram pesquisadas na plataforma de pesquisa *Google* e no banco de dados da Defesa Civil de Campinas, utilizando a seguinte estrutura de pesquisa: data e as palavras-chave (inundação, alagamento, enxurrada).

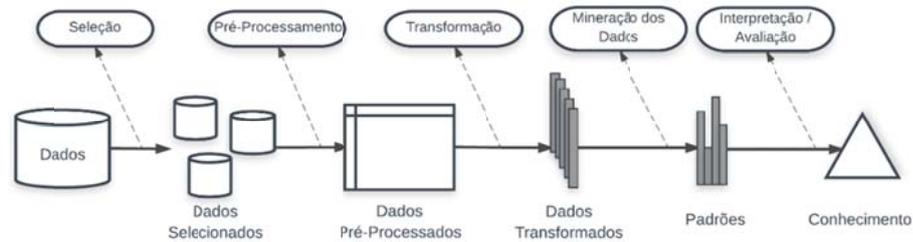
5.2 Métodos

5.2.1 Tratamento e análise dos dados fluviométricos

Em posse dos dados fluviométricos, foi necessário realizar a conversão dos dados para formato *.csv*, uma vez que os dados foram fornecidos em formato *“.lev”*. Para leitura do arquivo e subsequente conversão, foi utilizado o *software Levelogger 4.5.1* da empresa Solinst.

Em posse dos dados fluviométricos convertidos, foi realizado o processo de seleção, pré-processamento, transformação, mineração dos dados e interpretação, com base no fluxograma proposto por Fayaad et al. (1996) apresentado na Figura 5.3.

Figura 5.3: Fluxograma representando os principais processos presentes no processamento de informação (adaptado de Fayaad et al. 1996).



Na etapa de seleção, foram selecionados apenas os dados referentes ao período de novembro de 2014 até junho de 2019, período passível de corrigir os níveis negativos presentes nos dados, fruto da instalação da régua linimétrica acima do nível d'água. Dados coletados posteriores a 2020, apesar de não apresentarem dados de nível negativo, não são passíveis de correlacionar com os dados anteriores a essa data, uma vez que houve uma alteração na referência de nível d'água onde as medidas são realizadas. Além disso, os dados de julho de 2019 até dezembro do mesmo ano estão ausentes. A Figura 5.4 exibe em detalhe o sensor instalado na foz do ribeirão Proença e duas régua linimétricas utilizadas para correção dos valores registrados.

Figura 5.4: Sensor de nível d'água instalado na foz do ribeirão Proença (em vermelho) e escalas antiga e nova. A linha em amarelo correlaciona a escala antiga com a nova.

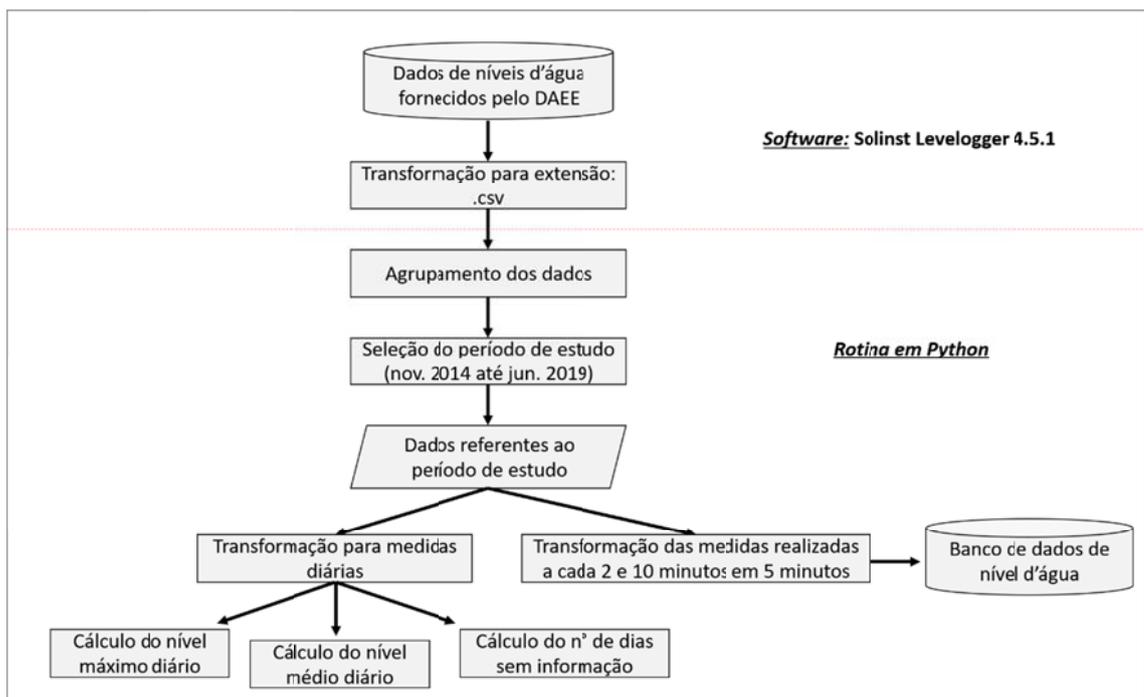


A partir da Figura 5.4, nota-se que a correção dos dados e transformação deles para a escala positiva se deu a partir da adição de 1,18 metros em cada medida. Desta forma, uma vez que não há como correlacionar os dados anteriores a 2020 com os posteriores, na etapa de seleção, optou-se por selecionar apenas os dados de novembro de 2014 até dezembro de 2019. Estes dados foram coletados com diferentes taxas de aquisição ao longo do período de estudo, variando desde aquisição a cada 2 minutos até aquisição a cada 10 minutos. Para homogeneização dos dados foi realizada a padronização em 5 minutos, utilizando-se uma rotina em Python desenvolvida especificamente para isso. O resultado desta etapa é o banco de dados de nível d'água.

Após a criação do banco de dados de nível d'água, a quantidade de dias sem registros e as principais estatísticas descritivas dos dados foram avaliadas. Além disso, foi calculada a média diária do nível d'água na foz do ribeirão Proença e comparada com a média mensal dos níveis d'água, buscando avaliar as flutuações mensais no nível.

Deve-se destacar que todas as etapas supracitadas foram executadas a partir de uma rotina desenvolvida em linguagem *Python*. A Figura 5.5 exibe um fluxograma simplificado que resume a etapa de tratamento e análise dos dados fluviométricos.

Figura 5.5: Fluxograma representando os principais processos realizados na etapa de tratamento e análise dos dados fluviométricos.



Após as análises supracitadas, os maiores valores de nível d'água diários registrados foram selecionados e suas estatísticas descritivas foram calculadas, a fim de avaliar dias prováveis de terem ocorrido inundações na bacia do ribeirão Proença.

A partir das estatísticas, três limiares de nível d'água foram avaliados, buscando determinar qual o limiar que melhor seleciona os eventos de inundação. Tais limiares foram selecionados acima do percentil 90%, partindo do pressuposto que os eventos de inundações ocorrerem com baixa frequência na bacia e estão dentre os 10% maiores valores registrados.

Para testar os limiares supracitados, os dias selecionados em cada um dos limiares foram comparados com a ocorrência ou não de eventos de inundação na bacia e em seu entorno, a partir de dados da Defesa Civil de Campinas e mídias digitais. Tanto para os *sites* informados pelo *Google*, quanto nos registros da Defesa Civil, foi avaliado se a data de ocorrência registrada é coincidente com a data pesquisada, a localização do fenômeno e o tipo de fenômeno registrado.

A partir das informações acima, cada data avaliada foi classificada de acordo com o tipo de ocorrência e sua localização, recebendo as seguintes classificações: ocorrência de inundação na bacia, ocorrência de alagamento na bacia, eventos registrados fora da bacia e sem informações. Por fim, após a comparação, foi possível avaliar a assertividade de cada um dos limiares escolhidos.

5.2.2 Tratamento e análise dos dados de radar

Uma vez que os dados obtidos por radares foram fornecidos em extensão .tif, foi necessário utilizar uma rotina em linguagem R desenvolvida pelo meteorologista Kleber Lopes da Rocha Filho para extração da precipitação ocorrida em cada um dos eventos analisados. Tal rotina, de maneira geral, seleciona apenas as quadrículas do radar inseridas na bacia em estudo e extrai os valores de precipitação em milímetros por hora para cada uma das quadrículas. O arquivo de saída da rotina compreende os dados de pluviosidade horários registrados a cada 5 minutos, três horas antes e depois do nível máximo registrado no dia escolhido.

Com os dados de saída da rotina em mãos, foi criada uma rotina em linguagem *Python* para unir os diversos dados de saída de cada quadrícula em um único banco de dados. Além disso, todos os valores de precipitação registrados em cada quadrícula foram divididos por 12, uma vez que cada hora contém 12 intervalos de 5 minutos e os dados coletados pelo

radar são interpretados como precipitações horárias. Estes dados compõem o banco de dados de radar meteorológico.

5.2.3 Unificação dos bancos de dados

Após a criação do banco de dados de níveis d'água, este contendo todos os dias avaliados pelo sensor de nível d'água, e do banco de dados de radar meteorológico, este contendo apenas os dias selecionados pelo limiar mais abrangente (permanência de 10%), foi desenvolvida uma rotina em linguagem *Python* para unificação dos bancos de dados.

Para tal, utilizou-se a função *merge* presente em uma das mais famosas bibliotecas do *Python*, denominada de *Pandas*. De modo geral, a função *merge* unifica dados a partir da especificação de dois parâmetros: coluna(s) que será comparada e o tipo de unificação dos dados.

No presente trabalho, foram empregadas as colunas *data* e *hora* para comparação dos bancos de dados e unificação do tipo *inner* foi a escolhida, uma vez que desejávamos selecionar apenas os dias e horas onde ambos os bancos de dados possuíam registros, ou seja, selecionar apenas os dias selecionados pelo limiar de 10% de permanência dos níveis na foz do ribeirão Proença.

Além disso, foi criada a variável precipitação média a partir da precipitação registrada em cada quadrícula ponderada pelo inverso da distância ao quadrado do centroide de cada quadrícula até o sensor 4D-038.

Deve-se destacar que o banco de dados supracitado é composto por 15 variáveis.

5.2.4 Variáveis auxiliares

Visando o desenvolvimento do modelo, foram agregadas ao banco de dados mais três variáveis, a saber: taxa de variação da precipitação, taxa de variação do nível d'água e o nível d'água 30 minutos no futuro, calculados como descrito a seguir:

- As taxas de crescimento ou redução da intensidade média da precipitação e do nível d'água, calculadas como:

$$Tx = \frac{P_{(ATUAL)} - P_{(5MINANTES)}}{5} \quad (\text{Eq. 1})$$

Sendo:

T_x a taxa de crescimento ou redução da intensidade média da precipitação (mm/min) ou aumento ou redução do nível d'água (m/min);

P_{ATUAL} a precipitação calculada no tempo analisado (mm) ou o nível d'água no tempo analisado (m);

$P_{5MINANTES}$ a precipitação registrada 5 minutos antes (mm) ou o nível d'água registrado 5 minutos antes (m);

Por fim, para geração da variável nível d'água 30 minutos no futuro o nível d'água foi deslocado 30 minutos no tempo. Maiores detalhes serão discutidos nos próximos tópicos.

5.2.5 Cálculo do tempo de concentração da bacia

Com os parâmetros morfométricos da bacia do ribeirão Proença, foi determinado o tempo de concentração da bacia a partir de duas equações apresentadas pelo manual do DNIT (2005), a saber: Equação de Kirpich Modificada (Equação 2); Equação de Kirpich publicada no “*California Culverts Practice*” (Equação 3) e a partir da equação de Tsuchiya retirada de Santos (2010) (Equação 4).

Equação de Kirpich Modificada

$$Tc = 1,42 \left(\frac{L^3}{H} \right)^{0,385} \quad (\text{Eq. 2})$$

Sendo: Tc = tempo de concentração, em horas;

L = comprimento do curso d'água, em quilômetros;

H = desnível máximo, em metros.

Equação de Kirpich (*California Culverts Practice*)

$$Tc = 0,95 \left(\frac{L^3}{H} \right)^{0,385} \quad (\text{Eq. 3})$$

Onde: Tc = tempo de concentração, em horas;

L = comprimento do curso d'água, em quilômetros;

H = desnível máximo, em metros.

Deve-se destacar que a Fórmula de Kirpich (*California Culverts Practice*) possui uma redução de 50% no valor do tempo de concentração.

Equação de Tsuchiya

$$Tc = 0,83 \left(\frac{L}{S^{0,6}} \right) \quad (\text{Eq. 4})$$

Onde, Tc = Tempo de concentração, em minutos;

L = comprimento do curso d'água, em quilômetros;

S = declividade equivalente, em metro por metro;

Note que, enquanto as duas primeiras equações retornam o valor em horas, a equação de Tsuchiya apresentada no presente trabalho retorna o tempo de concentração em minutos, havendo necessidade de transformação do resultado para horas.

Além das equações supracitadas, será analisado também o tempo de concentração obtido após redução de 40% dos resultados de Kirpich, a fim de adaptar tais fórmulas para cenários urbanos (Chow *et al.* 1988, apud Santos, 2010).

5.2.6 Correlação cruzada entre as estimativas de precipitação obtidos a partir de radar meteorológico e os dados de nível d'água

Buscando analisar comparativamente a precipitação ocorrida na bacia e o comportamento do nível d'água frente ao evento, foram calculadas a correlação de Pearson para as datas selecionadas pelo limiar de permanência de 10%. Após o cálculo da correlação direta de Pearson entre os dois fenômenos, calculou-se o tempo de atraso entre os dois fenômenos supracitados. Deve-se destacar que o intervalo de tempo de atraso entre dois fenômenos é comumente denominado de *lag time* e, segundo Taborga e Torrico (1974, apud

Farias Junior & Botelho, 2011), este possui uma íntima correlação com o tempo de concentração da bacia, dada pela seguinte equação (Equação 5):

$$Tc = \frac{\text{lag-time}}{0,6} \quad (\text{Eq. 5})$$

Onde, Tc = Tempo de concentração, em minutos;

Lag-time = tempo de atraso entre dois sinais (intensidade da precipitação e subida do nível d'água);

5.2.7 Modelagem hidrológica: aprendizado de máquina, XGBoost e análise de desempenho

Nesta etapa, foi desenvolvida uma rotina em linguagem *Python* para o emprego de um algoritmo de aprendizado de máquina denominado de *Extreme Gradient Boost* (XGBoost) para prever níveis d'água futuros. Esse algoritmo é denominado supervisionado, uma vez que fornecemos tanto os dados de entrada como a saída esperada, e o mesmo pode ser empregado para realizar classificações (classificar se ocorrerá ou não inundações) ou regressões (retornar o valor do nível d'água). No presente trabalho, ele foi empregado para realizar regressões, ou seja, prever o nível d'água futuro com base nos dados de entrada. Por fim, merece destaque também o fato de tal algoritmo ser baseado em árvores de decisão, ou seja, ele cria um mapa inúmeros possíveis resultados de uma série de escolhas que possuem uma relação entre si. Além disso, por tal característica, é possível analisar as decisões tomadas pelo algoritmo e os pesos atribuídos a cada uma das variáveis de entrada.

No que tange à análise de desempenho do modelo, dois parâmetros comumente utilizados em trabalhos de hidrogeologia, por exemplo em Venkatesan & Mahindrakar, (2019), Nguyen *et al.* (2021), Sudarno *et al.* (2022), Sanders *et al.* (2022) serão empregados: raiz do erro quadrático médio (*Root Mean Square Error – RMSE*) (Eq. 6) e eficiência de Nash-Sutcliffe (*Nash-Sutcliffe Efficiency - NSE*) (Eq. 7).

a) Root Mean Square Error (RMSE)

Segundo Sanders *et al.* (2022), o RMSE é utilizado para comparar, em termos de valores, no nosso caso em metros, o quanto os valores previstos pelo modelo diferiram em média dos valores observados (reais). Em outras palavras, quanto maior o RMSE, maior será a diferença entre o valor observado e o previsto, e conseqüentemente pior será o modelo. A Equação 6 exibe a fórmula de cálculo do RMSE. Trata-se de um parâmetro que possui a mesma dimensão dos dados de entrada (nível d'água em metros).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y^{obs} - Y^{prev})^2}, RMSE > 0 \quad (\text{Eq. 6})$$

Onde, N é o número de dados avaliados;

Y^{obs} é o valor do nível d'água observado; e

Y^{prev} é o valor do nível d'água previsto pelo modelo.

b) Nash-Sutcliffe Efficiency (NSE)

Segundo Moriasi *et al.* (2007), o NSE compara a quantidade de variância dos valores observados ao longo do tempo entre os valores previstos e o valor médio, conforme descrito pela Equação 7.

$$NSE = 1 - \frac{\sum_i^n (Y^{obs} - Y^{prev})^2}{\sum_i^n (Y^{obs} - Y^{med})^2}, 0 \geq NSE \geq 1 \quad (\text{Eq. 7})$$

Onde:

n = número de dados avaliados;

Y^{obs} é valor do nível d'água observado;

Y^{prev} é valor do nível d'água previsto pelo modelo; e

Y^{med} é valor do nível d'água médio

Cabe destacar que o autor correlaciona os valores obtidos pelo NSE com o desempenho do modelo, conforme exposto na Tabela 5.2.

Tabela 5.2: Avaliação do desempenho do modelo com base nos valores de NSE obtidos.

Adaptado de: Moriasi *et al.* (2007).

Avaliação do desempenho	Valor de NSE
Muito bom	$0,75 < NSE \leq 1,00$
Bom	$0,65 < NSE \leq 0,75$
Satisfatório	$0,50 < NSE \leq 0,65$
Insatisfatório	$NSE \leq 0,50$

c) Coeficiente de Determinação (R^2)

O Coeficiente de Determinação (R^2) é uma medida de ajuste de modelo que calcula a variação que o modelo regressivo gerado é capaz de explicar. Em outras palavras, tal medida expressa a quantidade de variância presente nos dados o modelo é capaz de explicar (Muralidhar, 2021). A Equação 8 expressa o cálculo do R^2 conforme apresentado por Muralidhar (2021).

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}}, \quad 0 \leq R^2 \leq 1 \quad (\text{Eq. 8})$$

Onde:

$SS_{residual}$ = soma dos quadrados da variância não explicada;

SS_{total} = soma dos quadrados da variância total;

Com base na Equação 8, temos que o valor de R^2 varia entre 0 e 1, onde quanto maior o valor de R^2 , melhor é o desempenho do modelo.

d) Coeficiente de Determinação Ajustado (R^2 ajustado)

Já o Coeficiente de Determinação Ajustado (R^2 ajustado), como o nome já diz, é um ajuste do R^2 , uma vez que a inclusão de variáveis quaisquer ao modelo de regressão múltipla leva a um aumento do valor de R^2 (Muralidhar, 2021). A Equação 9 exibe a fórmula de cálculo do R^2 ajustado, onde é possível notar uma penalização do valor de R^2 ao adicionar/utilizar variáveis pouco explicativas ao modelo.

$$R^2_{ajustado} = 1 - \frac{(1-R^2)(n-1)}{(n-1)-p}, \quad 0 \leq R^2_{ajustado} \leq 1 \quad (\text{Eq. 9})$$

Onde:

n = número de dados avaliados;

R^2 é valor do coeficiente de determinação;

p = número de preditores/variáveis;

Note que à medida que aumenta o número de variáveis (p), diminui o valor de R^2 ajustado.

e) *Mean Absolute Percentual Error* (MAPE)

O Erro Absoluto Percentual Médio (MAPE) é uma medida de avaliação da acurácia do modelo regressivo que retorna o valor do erro absoluto médio em termos de porcentagem (Glen, 2022). A Equação 10 expressa o cálculo do MAPE, adaptada de Glen (2022).

$$MAPE = \frac{1}{n} \sum^n \left| \frac{Y^{obs} - Y^{prev}}{Y^{obs}} \right| \cdot 100, \quad 0 \leq MAPE \leq 100\% \quad (\text{Eq. 10})$$

Onde:

n = número de dados avaliados;

Y^{obs} é valor do nível d'água observado;

Y^{prev} é valor do nível d'água previsto pelo modelo;

5.2.7.1 Definição dos hiperparâmetros do modelo e seleção das variáveis de interesse

Nesta etapa, foram utilizados os dados presentes no banco de dados unificado supracitado, composto pelas seguintes variáveis: temperatura na água, nível d'água, taxa de variação do nível d'água, precipitação em cada quadrícula (total de 10), precipitação média e taxa de variação da precipitação média, totalizando 15 variáveis.

O fluxograma da Figura 5.6 exibe de forma simplificada os principais procedimentos executados na etapa de criação do modelo final da presente pesquisa.

Após a importação do banco de dados unificado, contendo as 15 variáveis supracitadas, foi realizado o pré-processamento dos dados, onde foram divididos os dados em 80% para treinamento do modelo, 10% para validação e 10% para fase posterior de teste.

Na etapa de desenvolvimento do XGBoost, os seguintes hiperparâmetros do modelo foram avaliados: *number estimadors*, *max depth*, *learning rate*, *subsample* e *colsample_bytree*. Tais hiperparâmetros correspondem a: número de vezes que o XGBoost tentará aprender, tamanho máximo das árvores de decisão, taxa de aprendizado, proporção de subamostras das instâncias de treinamento e subamostra aleatória de colunas quando uma nova árvore é criada.

Ainda na etapa de desenvolvimento do XGBoost, foi utilizada a função denominada de *GridSearchCV*, pertencente a biblioteca *Scikit-Learn*, que tem como objetivo realizar todas as combinações possíveis de hiperparâmetros que lhes foram passados e avaliar o desempenho de cada uma, para seleção dos melhores hiperparâmetros. Para a avaliação do desempenho de cada combinação de hiperparâmetros foi utilizada a técnica de validação cruzada denominada de K-Fold, onde o número de pastas (*Folds*) utilizadas foi 10, com embaralhamento dos dados e a métrica de avaliação do resultado foi o valor de RMSE.

Após a definição dos melhores hiperparâmetros do modelo, obteve-se o modelo prévio denominado de “Modelo XGBoost 15VAR”.

Com o modelo prévio em mãos, suas respectivas estatísticas associadas (RMSE e NSE) e hiperparâmetros definidos, novas variáveis foram adicionadas ao modelo afim de verificar se modelo com a nova variável apresenta melhor desempenho (melhores resultados de RMSE e NSE). As novas variáveis adicionadas foram geradas a partir do deslocamento de dados do passado para o presente, em um intervalo de 5 até 10 minutos.

Uma vez que a série temporal empregada na presente pesquisa não é sequencial, ou seja, apresenta lacunas, uma verificação de sequencialidade foi realizada para garantir que dados espaçados mais que 5 minutos no tempo não recebessem dados de 5 minutos no

A Tabela 5.3 ilustra o procedimento descrito acima para o cálculo da taxa de precipitação média com base na Equação 1 a partir de dados hipotéticos. Além da precipitação média, tal procedimento foi empregado na geração da taxa de subida de nível e na geração das novas variáveis.

Tabela 5.3: Procedimento realizado no cálculo da taxa de precipitação média.

Dia e Hora	Precipitação Média (mm/5min)	Taxa de Precipitação Média (mm/5min)
01/01/2018 às 14:00	0,47	NaN
01/01/2018 às 14:05	0,33	-0,028
01/01/2018 às 14:10	0,66	0,066
01/01/2018 às 14:20	0,88	NaN
01/01/2018 às 14:25	1,12	0,048
01/01/2018 às 14:30	0,77	-0,07

A Tabela 5.4 exibe as novas variáveis avaliadas. Ao total foram avaliadas as variáveis nível e precipitação média e 6 novos modelos de teste foram gerados.

Tabela 5.4: Variáveis adicionadas ao modelo denominado de XGBoost 15VAR.

Variável Adicionada	Nome do Modelo de Teste
Nível 5 minutos no passado	XGBoost 15VAR + N5
Nível 10 minutos no passado	XGBoost 15VAR + N10
Nível 5 e 10 minutos no passado	XGBoost 15VAR + N5-10
Precipitação média 5 minutos no passado	XGBoost 15VAR + PM5
Precipitação média 10 minutos no passado	XGBoost 15VAR + PM10
Precipitação média 5 e 10 minutos no passado	XGBoost 15VAR + PM5-10

Para cada um dos modelos de teste gerados, apresentados na Tabela 5.4, novos cálculos de RMSE e NSE foram realizados, a fim de constatar um aumento ou não da performance do novo modelo gerado.

Além disso, cabe ressaltar que cada nova variável gerada foi adicionada ao banco de dados e um novo treinamento e subsequente avaliação dos resultados foi realizado, a fim de verificar quais variáveis adicionavam melhorias ao modelo. Tal procedimento foi realizado de forma individual, ou seja, primeiramente adicionou-se no banco de dados a variável contendo o nível d'água 5 minutos no passado e em seguida o modelo foi treinado e avaliado (Modelo 15VAR + N5). Em seguida, foi avaliada a variável contendo o nível d'água 10 minutos no passado e assim por diante. Ao final, as variáveis que apresentaram melhorias no desempenho do modelo foram mantidas, enquanto as variáveis que não geraram melhorias foram excluídas.

Deve-se destacar que o procedimento utilizado para a criação da variável “nível d'água 10 minutos no passado” foi similar ao exemplificado pela Tabela 5.3, porém o espaço temporal avaliado foi de 10 minutos no passado. Ou seja, caso a data e hora avaliada não possua um nível d'água 10 minutos no passado, o valor recebido será NaN.

Paralelamente aos processos supracitados, foi criada no banco de dados a variável alvo, variável que corresponde ao nível d'água 30 minutos no futuro, resultado que será previsto pelo modelo. A Tabela 5.5 exibe o procedimento adotado para criação da variável alvo.

Tabela 5.5: Dados hipotéticos para ilustrar a técnica empregada para geração da variável alvo. Destaque para a posição do valor em vermelho.

Hora	Nível (m)	Precipitação (mm)	Variável Alvo (Nível 30 minutos no futuro)
15:30	0,8361	0,2779	0,8382
15:35	0,8347	0,1841	0,9827
15:40	0,8336	1,2761	0,9939
15:45	0,8352	8,0630	...
15:50	0,8360	8,0103	...
15:55	0,8392	37,3716	...
16:00	0,8382	82,4168	...
16:05	0,9827	18,3425	...
16:10	0,9939	21,4837	...

Ao final, tanto as variáveis que adicionaram melhorias ao modelo, quanto a variável alvo foram mantidas no banco de dados unificado e este, por sua vez, será utilizado na etapa subsequente que abrange a geração do modelo final.

Deve-se destacar que a criação da variável “nível 30 minutos no futuro” seguiu os mesmos princípios empregados na criação das novas variáveis e das taxas. Entretanto, o espaço temporal avaliado foi de 30 minutos no futuro. Ou seja, caso a data e hora avaliada não possua um nível d’água 30 minutos no futuro, o valor recebido será NaN.

5.2.7.2 Criação do Modelo Final

Na etapa de criação do modelo final da pesquisa, foram utilizados como dados de entrada do modelo as variáveis existentes e as variáveis criadas que geram melhorias ao modelo. Tais variáveis serão utilizadas pelo modelo final para prever a variável alvo, ou seja, prever o nível d’água 30 minutos no futuro.

Nesta etapa, novos hiperparâmetros foram testados, agora calculados com base nas variáveis antigas e novas. Com os melhores hiperparâmetros definidos, prosseguiu-se com a etapa de validação do modelo final e seu respectivo desempenho.

Com o modelo final em mãos, foi verificada a proporção de acertos e erros das previsões realizadas pelo modelo final com base nos dados de teste. Os limiares de aviso e alerta foram considerados nesta etapa.

5.2.7.3 Modelagem do nível d'água para variados tempos futuros

Após a definição e avaliação do modelo para previsão de níveis d'água 30 minutos no futuro, novos modelos foram gerados para prever níveis d'água em diferentes tempos no futuro, a fim de avaliar o desempenho do modelo. Deve-se destacar que o banco de dados utilizado para a criação dos novos modelos e os hiperparâmetros empregados foram os mesmos utilizados na criação do modelo de previsão de níveis d'água 30 minutos no futuro.

Além da comparação do desempenho dos modelos na previsão de diversos tempos futuros, comparou-se também quais variáveis o modelo considerava mais importante na geração da estimativa.

6. RESULTADOS

6.1 Nível d'água no ribeirão Proença – estabelecimento de níveis de alerta e de atenção

Adotando, em um primeiro momento, o intervalo de estudo de 2014 a 2020, foram avaliadas as lacunas presentes no banco de dados de nível d'água do sensor fluviométrico instalado na foz do ribeirão Proença. A Figura 6.1 denota as lacunas encontradas nos dados de nível d'água no período supracitado. As barras cinza da Figura 6.1 representam o número de dias em cada mês do ano, enquanto as barras amarelas representam o número de dias registrados pelo sensor em cada mês. Desta forma, meses no qual a barra cinza está completamente preenchida pela barra amarela são meses em que há registros para todos os dias.

Além disso, os resultados apresentados na Figura 6.1 explicam a necessidade e um novo refinamento na escolha do período abordado pela presente pesquisa, uma vez que não há dados a partir do mês de julho de 2019 e nem ao longo do ano de 2014. Dito isso, o presente trabalho abordará o período de dados que compreende novembro de 2014 até junho de 2019.

Ademais, nota-se que dentre o intervalo que será analisado, o ano de 2016 é o ano que possui maior quantidade de lacunas, principalmente no período menos chuvoso da região. Por outro lado, o ano de 2017 pode ser considerado o ano com maior cobertura de dados, uma vez que todos os dias do ano possuem registros de nível d'água.

Do início de novembro de 2014 até o final de junho de 2019 há 1702 dias e o número de dias no qual há dados disponíveis no sensor é de 1378, logo a porcentagem de lacunas presentes nos dados é de 19%.

Já a Figura 6.2 compara a média geral (média de todos os dias – em cinza) e a média mensal (média para cada mês – em amarelo) dos níveis d'água. Baseado nesta figura, temos que nos períodos de novembro a março o nível d'água encontra-se maior que a média geral, enquanto no período de abril a setembro, o nível encontra-se abaixo da média. Isso se deve ao fato que, o primeiro é o período chuvoso da região, enquanto o último é o período seco.

Figura 6.1: Análise da quantidade de falhas nos dados de nível d'água registrados pelo sensor 4D-038. As barras cinzas representam o número de dias em cada mês e as barras amarelas representam o número de dias registrados pelo sensor.

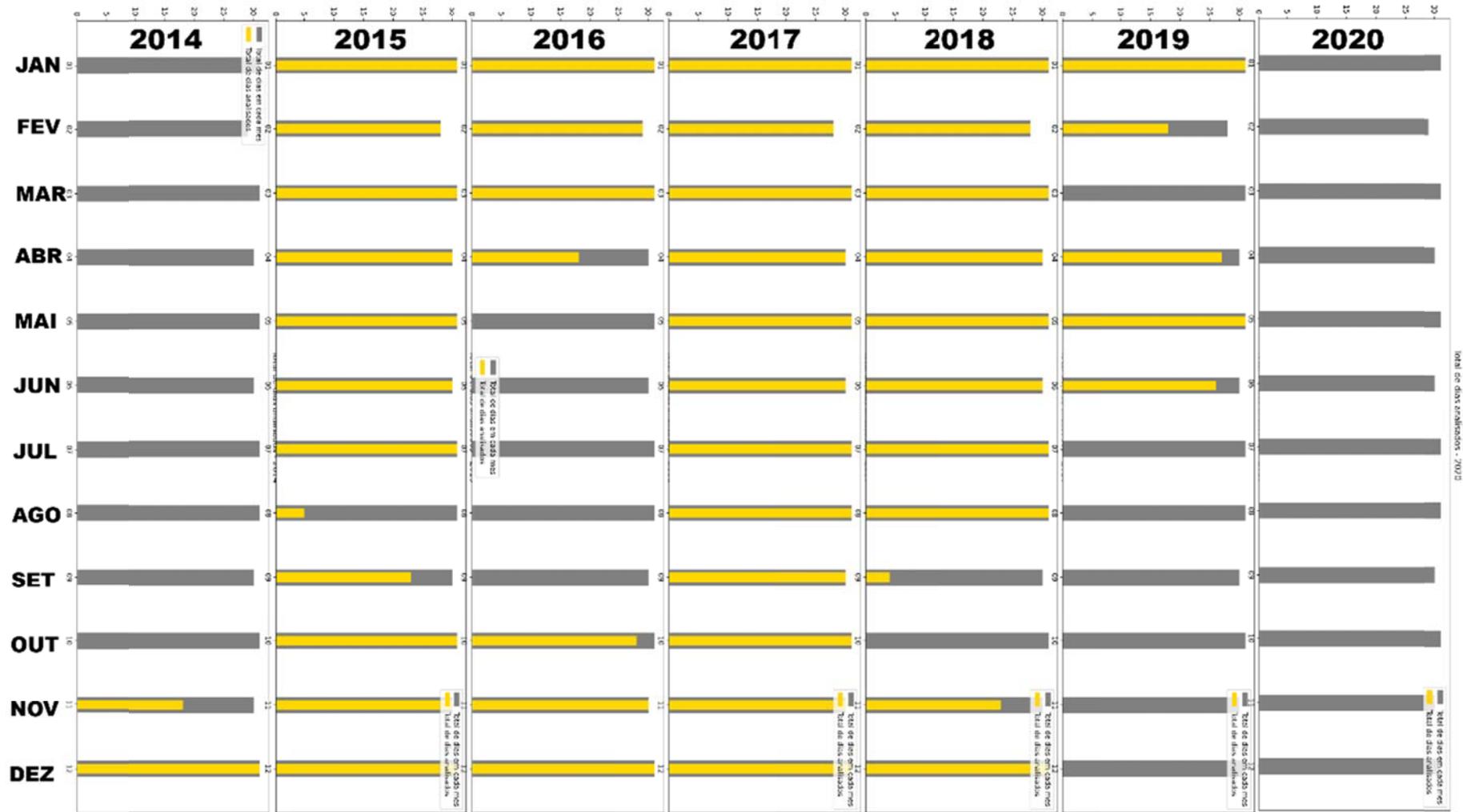
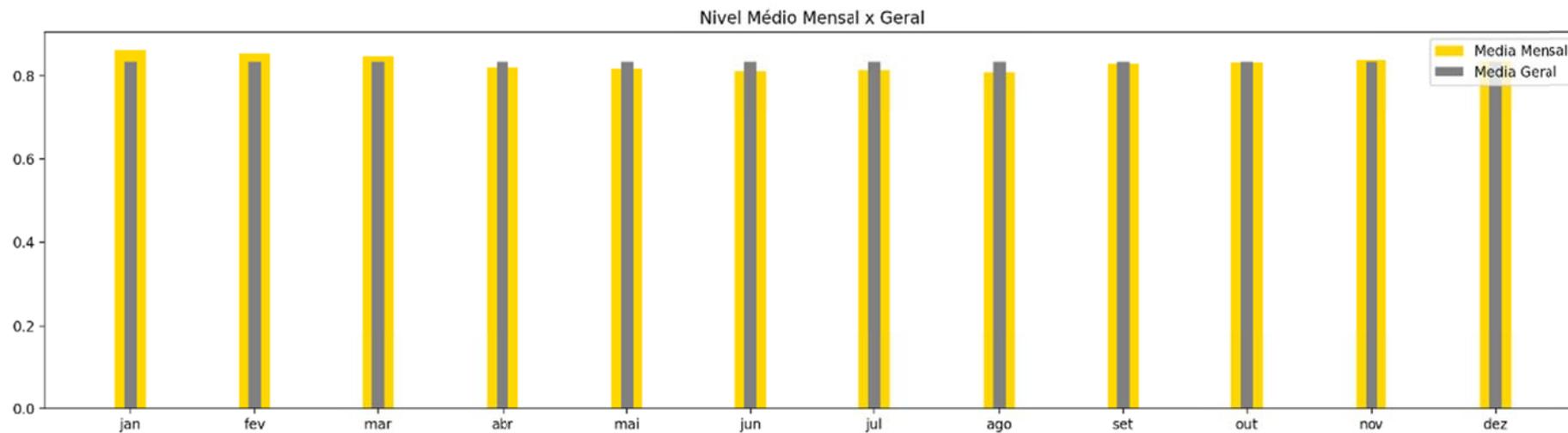


Figura 6.2: Nível médio mensal (amarelo) comparado com o nível médio geral (cinza).



A Tabela 6.1 exibe a média do nível d'água geral e demais estatísticas descritivas dos dados. Com base na Tabela 6.1, temos que o número de dados coletados (medidas limnimétricas) nos 1378 dias é de aproximadamente 385 mil, sendo o nível médio registrado de 0,83 metros, variando conforme a época seca ou chuvosa. Além disso, o nível mínimo registrado foi de 0,57 metros, enquanto o máximo foi de 5,26 metros.

Tabela 6.1: Principais parâmetros estatísticos dos dados de nível corrigidos.

Estatísticas	Nível corrigido
Nº de dados	385.865
Nº de dias	1378
Média	0,83 (m)
Desvio Padrão	0,10 (m)
Mínimo	0,74 (m)
25%	0,81 (m)
50%	0,82 (m)
75%	0,84 (m)
90%	0,86 (m)
95%	0,90 (m)
Máximo	5,26 (m)

A partir dos quartis observados, é possível notar que em 95% das medidas limnimétricas realizadas o nível d'água não excedeu valores maiores que 0,90 metros, evidenciando que níveis acima deste valor são pouco comuns.

Já a Tabela 6.2 exibe os parâmetros estatísticos dos maiores níveis observados diariamente. Note que agora o número de dados é igual ao número de dias avaliados, uma vez que o nível máximo de cada dia foi selecionado.

Com base na Tabela 6.2, temos um incremento na média e no desvio padrão, quando comparada aos dados totais. Note que para 95% dos dias o nível máximo diário não ultrapassou os 2,02 metros, enquanto para 90% dos dias o nível d'água permaneceu abaixo de 1,59 metros.

Tabela 6.2: Principais parâmetros estatísticos dos dados de nível máximo diários.

Estatísticas	Nível corrigido
Nº de dados	385.865
Nº de dias	1378
Média	0,83 (m)
Desvio Padrão	0,10 (m)
Mínimo	0,74 (m)
25%	0,81 (m)
50%	0,82 (m)
75%	0,84 (m)
90%	0,86 (m)
95%	0,90 (m)
Máximo	5,26 (m)

Buscando selecionar apenas os dias nos quais os níveis d'água máximos diários registrados apresentam valores incomuns e que estão possivelmente correlacionados a eventos de inundação, três limiares de níveis d'água foram aplicados.

A Figura 6.3 exibe a posição dos limiares (Q3, Q6 e Q10) aplicados no cotograma composto pelos níveis máximos diários. Deve-se destacar que os limiares 1 (Q10), 2 (Q6) e 3 (Q3) selecionam os níveis maiores ou iguais a 1,58m, 1,96m, 2,30m e ocorrem apenas em 10, 6 e 3% do total de tempo avaliado, respectivamente. Já a Tabela 6.3 exibe a quantidade de dias selecionados por cada um dos limiares empregados.

Com base na Figura 6.3 e na Tabela 6.3, temos que o Limiar 1 (vermelho) é o mais abrangente, selecionando um total de 145 dias nos quais o nível d'água excedeu ou igualou a 1,58 metros, enquanto o Limiar 3 (roxo) é o menos abrangente, selecionando um total de 40 dias em que o nível d'água igualou ou excedeu 2,30 metros. Deve-se destacar que os 40 dias selecionados pelo Limiar 3 estão presentes dentro dos 145 dias selecionado pelo Limiar 1 e assim por diante.

Figura 6.3: Cotograma dos níveis máximos diários registrados pelo sensor instalado na foz do ribeirão Proença. Note a presença dos três limiares: Limiar 1 (em vermelho), Limiar 2 (em verde) e Limiar 3 (em roxo).

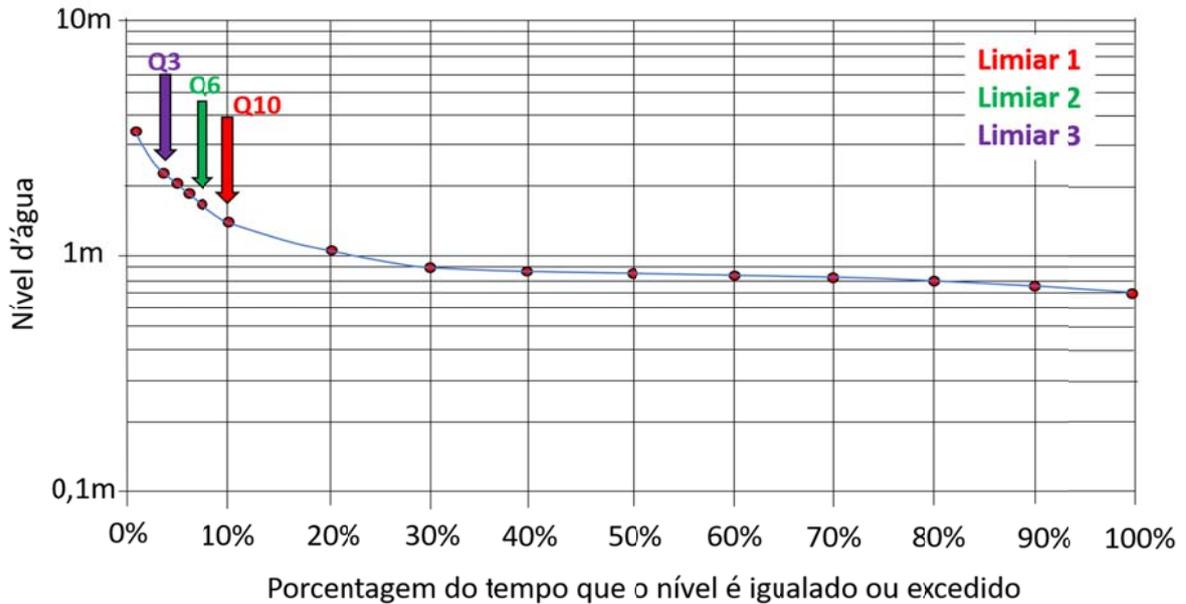


Tabela 6.3: Número de dias selecionados por cada um dos limiares empregados.

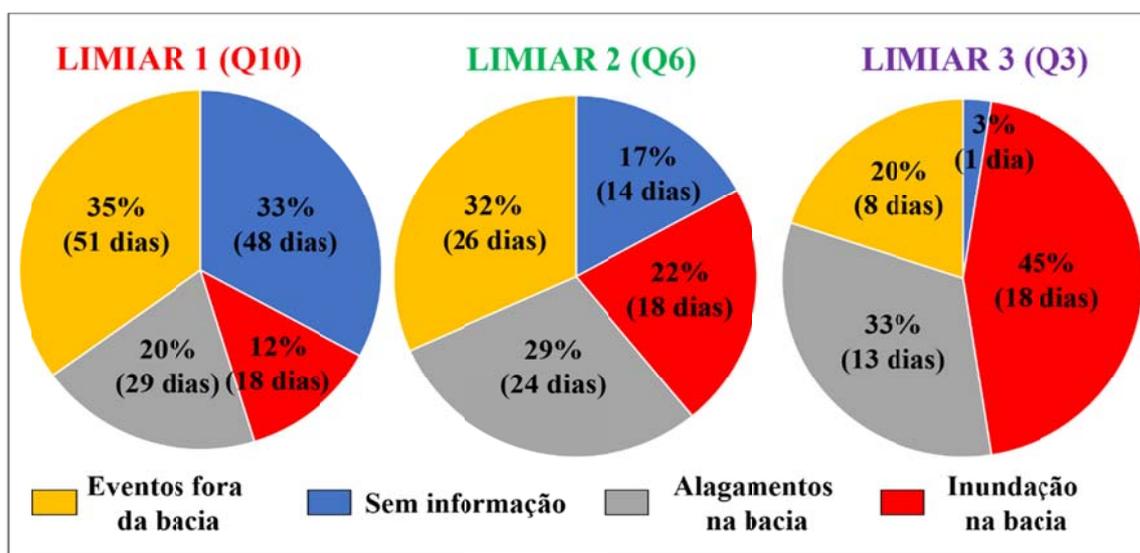
	Número de dias
Limiar 1 (Q10)	145 dias
Limiar 2 (Q6)	82 dias
Limiar 3 (Q3)	40 dias

Avaliando a efetividade dos limiares empregados na seleção de possíveis eventos de inundação, cada um dos 145 dias selecionados pelo limiar mais abrangente foi classificado conforme a presença ou não de informações sobre a ocorrência de inundações ou fenômenos correlatos registrados em notícias de jornal e no banco de dados da Defesa Civil de Campinas, tanto na bacia do ribeirão Proença, quanto no município. A Figura 6.4 exibe o total de dias e suas respectivas porcentagens em cada uma das classes utilizadas.

Baseado na Figura 6.4, pode-se destacar que o número de dias sem registro de ocorrências de inundações e fenômenos correlatos diminuiu à medida que o nível d'água estabelecido no limiar aumentou, corroborando assim com o fato de que maiores valores de nível d'água tendem a estar associados a fenômenos mais intensos. Nota-se também que a

medida que aumentamos o limiar, o número de eventos fora da bacia e sua respectiva percentagem diminui.

Figura 6.4: Distribuição do número de dias e suas respectivas percentagens em cada classe para cada um dos limiares aplicados.



Podemos destacar que a assertividade do Limiar 1 (1,58 metros) na seleção de eventos de inundação na bacia do ribeirão Proença é de apenas 12%. Uma vez que tal percentagem é relativamente baixa, podemos constatar que o Limiar 1 inclui dias em que ocorreram eventos de inundação e dias em que não houve eventos.

Já o Limiar 2 (1,96 metros), por sua vez, apresentou uma assertividade de 22%, 10% a mais que a observada pelo Limiar 1. Deve-se destacar que o número de registros de eventos de inundações, somado ao número de registros de alagamentos, totaliza mais de 50% dos dias avaliados, embora a assertividade continue baixa.

Por fim, o Limiar 3 (2,30 metros), limiar menos abrangente, apresentou uma assertividade de 45%, 32% a mais que a observada pelo Limiar 1. De longe, tal limiar é o que possui maior percentagem de seleção de eventos na bacia (78%), sendo 33% dos registros relacionados a alagamentos e 45% relacionados a inundações. Ademais, nota-se também que a percentagem de dias sem informações é quase nula, denotando assim uma íntima correlação com eventos mais severos. Tal característica é de se esperar, uma vez que ao aumentar o limiar de nível d'água, restringimos os eventos menos recorrentes na bacia e com maiores probabilidades de estarem associados a inundações e alagamentos.

Com base no exposto, o Limiar 3 (2,30 metros) é o limiar que possui maior potencial para filtragem de eventos correlacionados a inundações e/ou alagamentos na bacia, uma vez que 78% dos dias selecionados apresentaram ocorrência de eventos na bacia. Dito isso, o mesmo poderá ser utilizado como um nível de atenção para potenciais eventos de inundação e/ou alagamentos na bacia do ribeirão Proença.

Por outro lado, o Limiar 1 (1,58 metros) é o limiar menos abrangente e consequentemente será o que mais beneficiará o processo de aprendizado de máquina, uma vez que é o que seleciona a maior quantidade de dados. Cabe destacar que, embora os modelos gerados a partir de algoritmos de *Machine Learning* não necessitem de diversos parâmetros, eles apresentam melhores desempenhos quando gerados a partir de maiores quantidades de dados. A Tabela 6.4 exibe os 145 dias selecionados por este limiar, com destaque para os dias em que há registros de inundação na bacia do ribeirão Proença (vermelho), registros de alagamentos (cinza), registros fora da bacia (laranja) e sem informações (azul).

6.2 Estimativa do valor do *lag-time*

Partindo da premissa que os valores de *lag-time* variam conforme a precipitação ocorrida na bacia, foram calculados valores de *lag-time* para cada um dos 145 eventos avaliados. A partir dos resultados, verificou-se que os valores de *lag-time* obtidos variaram de 0 minutos, denotando ausência de tempo de retardo entre os sinais, até 110 minutos.

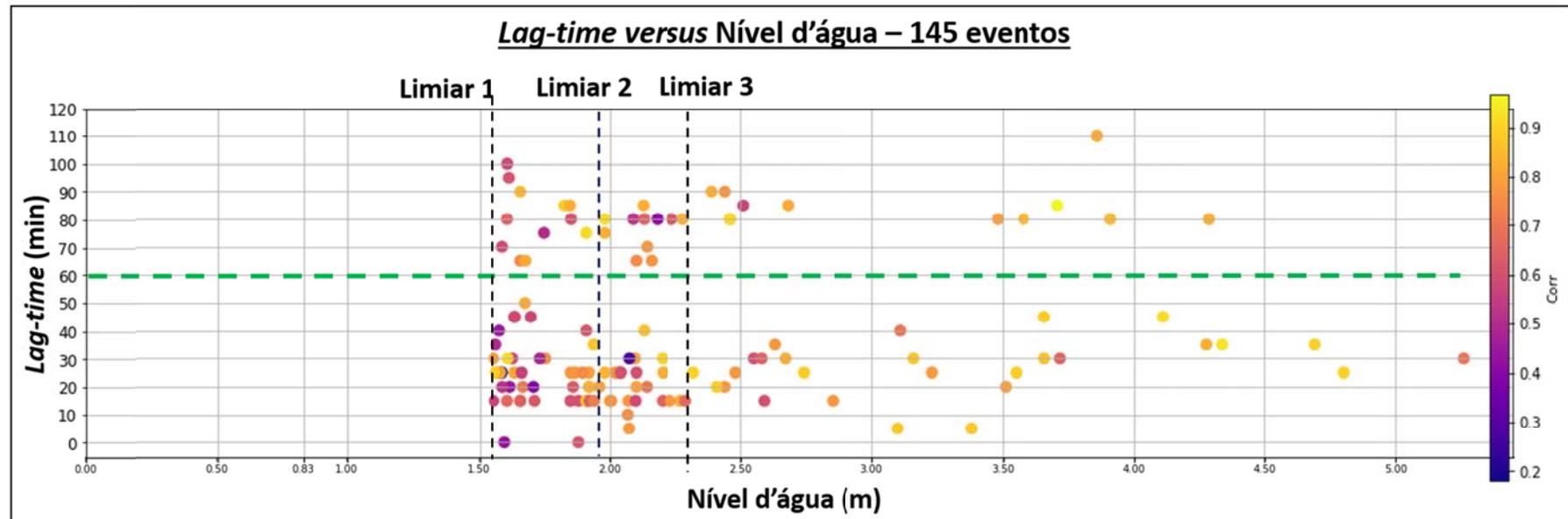
No que tange à correlação de Pearson, os valores observados variaram de 0,18 até 0,97. A Figura 6.5 exibe os resultados supracitados, relacionando o nível d'água máximo de cada um dos 145 eventos com o seu melhor ajuste de *lag-time*. A escala de cores nesta figura indica o valor da melhor correlação cruzada em cada evento.

Baseado na Figura 6.5, nota-se a presença de dois grupos distintos de *lag-time*, estes delimitados pela linha verde correspondente ao *lag-time* de 60 minutos. Tais grupos podem ser representados pelos valores de 30 e 70 minutos, com valores de correlação similares conforme ilustrado no histograma da Figura 6.5. Cabe ressaltar que a presença dos dois grupos supracitados concretiza a premissa assumida que os valores de *lag-time* variam conforme a precipitação ocorrida.

Tabela 6.4: Data dos 145 dias selecionados pelo Limiar 1. As cores representam a classificação dada para cada um dos eventos ocorridos, sendo eles: inundação na bacia (vermelho - sublinhado), alagamentos na bacia (cinza - itálico), registros fora da bacia (laranja - negrito) e sem informações (azul).

24/11/2014	<i>26/11/2014</i>	22/12/2014	23/12/2014	<i>28/12/2014</i>	29/12/2014	<i>06/01/2015</i>
<i>11/01/2015</i>	<i>14/01/2015</i>	<i>24/01/2015</i>	<i>26/01/2015</i>	05/02/2015	09/02/2015	<u>10/02/2015</u>
<i>15/02/2015</i>	17/02/2015	<i>19/02/2015</i>	<i>27/02/2015</i>	<i>28/02/2015</i>	<u>06/03/2015</u>	07/03/2015
<i>15/03/2015</i>	<u>16/03/2015</u>	<i>18/03/2015</i>	<i>21/03/2015</i>	<i>16/04/2015</i>	21/04/2015	10/05/2015
<i>20/05/2015</i>	25/07/2015	08/09/2015	<i>11/09/2015</i>	<i>28/09/2015</i>	<i>26/10/2015</i>	<i>31/10/2015</i>
19/11/2015	<i>20/11/2015</i>	<i>28/11/2015</i>	<u>29/11/2015</u>	<i>06/12/2015</i>	<i>07/12/2015</i>	13/12/2015
<i>19/12/2015</i>	<u>25/12/2015</u>	<i>27/12/2015</i>	28/12/2015	<u>31/12/2015</u>	<i>01/01/2016</i>	<i>12/01/2016</i>
13/01/2016	<i>14/01/2016</i>	27/01/2016	<i>02/02/2016</i>	06/02/2016	<u>07/02/2016</u>	15/02/2016
<i>19/02/2016</i>	22/02/2016	24/02/2016	26/02/2016	<i>27/02/2016</i>	<i>28/02/2016</i>	<i>02/03/2016</i>
<i>10/03/2016</i>	11/03/2016	<u>23/03/2016</u>	<i>13/10/2016</i>	<i>14/10/2016</i>	<i>25/10/2016</i>	09/11/2016
11/11/2016	<i>28/11/2016</i>	<i>12/12/2016</i>	<i>18/12/2016</i>	<i>25/12/2016</i>	<i>27/12/2016</i>	<i>28/12/2016</i>
09/01/2017	<i>10/01/2017</i>	<i>13/01/2017</i>	<u>16/01/2017</u>	<i>17/01/2017</i>	19/01/2017	26/01/2017
<u>30/01/2017</u>	31/01/2017	<i>01/02/2017</i>	03/02/2017	04/02/2017	26/02/2017	<i>03/03/2017</i>
05/03/2017	<i>07/03/2017</i>	<i>17/03/2017</i>	<i>18/03/2017</i>	19/03/2017	<i>06/04/2017</i>	26/04/2017
<i>04/05/2017</i>	<i>05/05/2017</i>	<i>18/05/2017</i>	19/05/2017	21/05/2017	<i>05/06/2017</i>	30/09/2017
26/10/2017	<i>04/11/2017</i>	<i>10/11/2017</i>	<u>18/11/2017</u>	<i>27/11/2017</i>	01/12/2017	06/12/2017
07/12/2017	<u>25/12/2017</u>	30/12/2017	<i>31/12/2017</i>	<u>03/01/2018</u>	11/01/2018	20/01/2018
06/02/2018	<i>10/02/2018</i>	<i>18/02/2018</i>	<i>26/02/2018</i>	<i>02/03/2018</i>	<i>15/03/2018</i>	<i>21/03/2018</i>
<u>01/04/2018</u>	<i>03/04/2018</i>	<i>19/11/2018</i>	<i>23/11/2018</i>	24/11/2018	<u>25/11/2018</u>	<u>01/12/2018</u>
<i>18/12/2018</i>	<i>02/01/2019</i>	03/01/2019	<u>04/01/2019</u>	<i>08/01/2019</i>	<u>24/01/2019</u>	<i>26/01/2019</i>
<i>29/01/2019</i>	<i>06/02/2019</i>	<i>16/02/2019</i>	<i>13/04/2019</i>	<i>29/05/2019</i>		

Figura 6.5: *Lag-time versus* nível d'água para cada um dos 145 eventos analisados. Cores mais quentes representam correlações cruzadas maiores, enquanto cores mais frias representam correlações cruzadas menores. A linha em verde segmenta dois grupos de valores de *lag-time*: maiores e menores que 60 minutos.



Nota-se também que, independentemente do grupo, para maiores valores de níveis d'água, maiores são os valores da correlação cruzada, sendo raros os eventos onde a correlação obtida foi menor que 0,5.

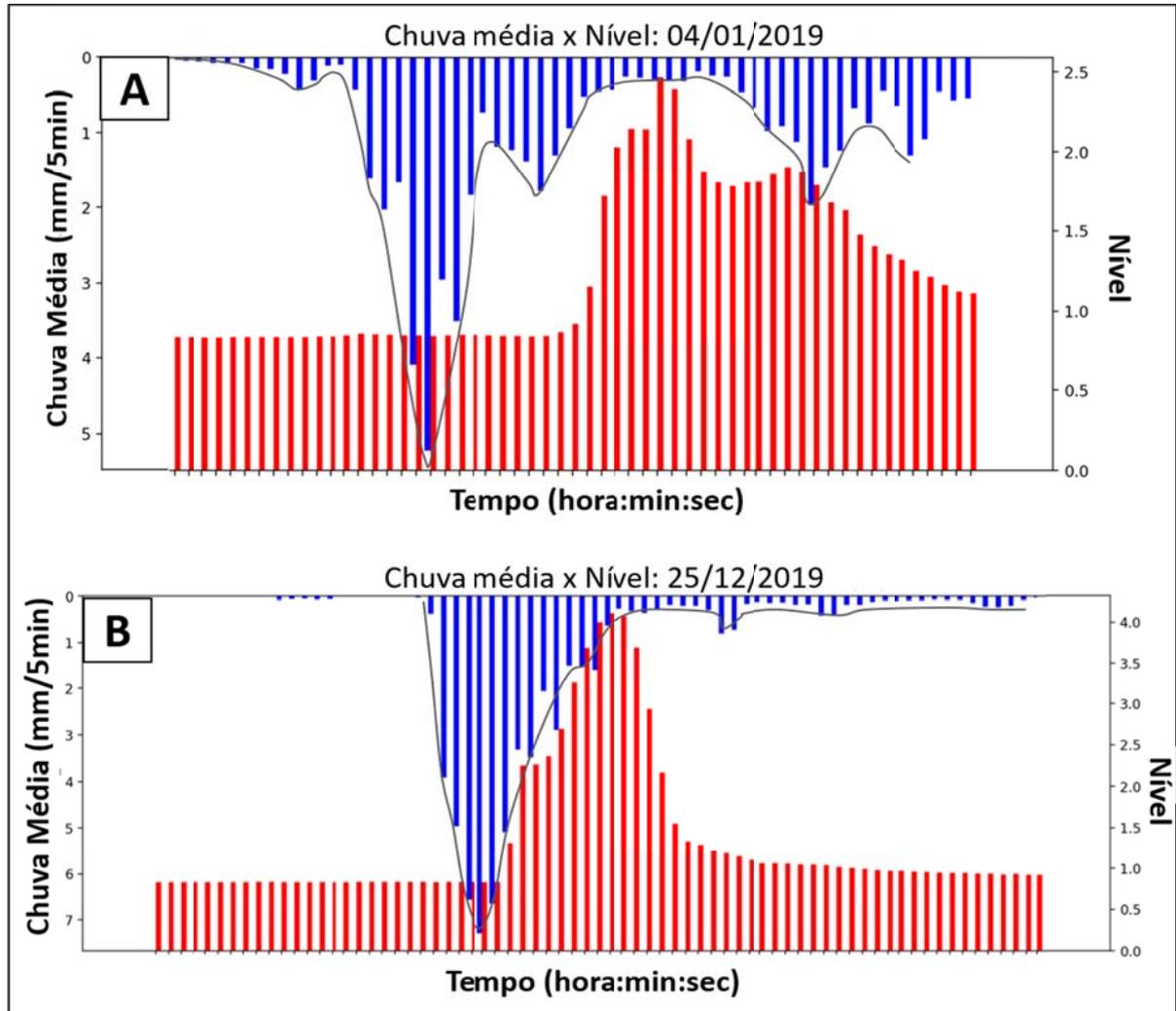
Por outro lado, para níveis d'água menores a correlação obtida variou amplamente, porém apresentou uma predominância de valores baixos. Isto sugere que a precipitação possui papel importante para a previsão dos níveis d'água mais elevados do ribeirão Proença no ponto instrumentado, principalmente para níveis superiores ao Limiar Q3, o que corrobora a sua utilização para a construção de um sistema de alerta antecipado de inundação para a bacia.

Buscando destacar as possíveis explicações para a ocorrência dos dois grupos de *lag-time* observados, as precipitações nos 47 dias em que há registros de inundação ou alagamentos na bacia do ribeirão Proença foram classificadas em 2 tipos: eventos de menor intensidade pluviométrica e eventos de maior intensidade pluviométrica.

Tais classificações foram baseadas na conformidade do evento de precipitação, ou seja, eventos de menor intensidade pluviométrica foram aqueles que apresentaram precipitações ao longo de um tempo considerável, enquanto eventos de maior intensidade pluviométrica foram aquelas que apresentaram precipitações restritas e em pequenos intervalos de tempo.

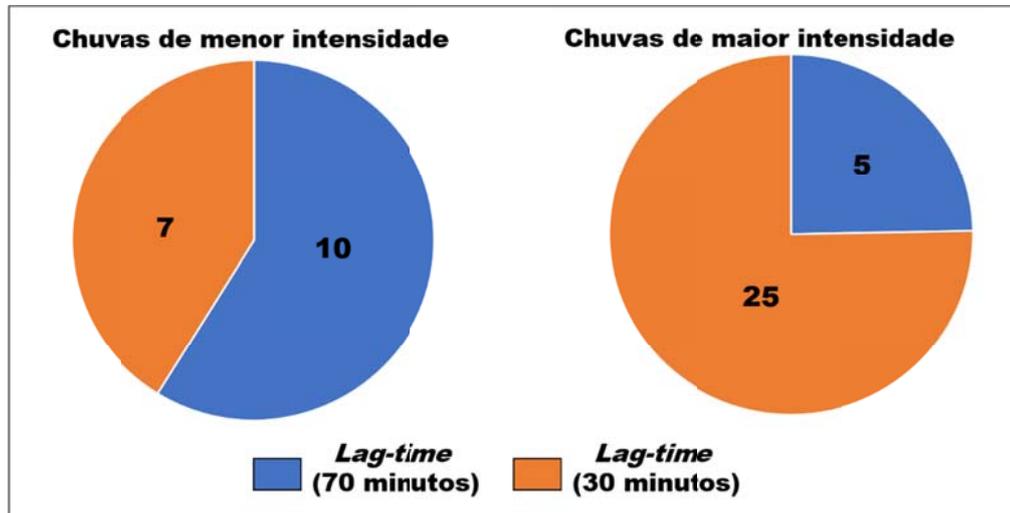
A Figura 6.6 exhibe exemplos para ambos os tipos de eventos de precipitação adotados neste estudo, sendo denominados de eventos de menor intensidade pluviométrica as precipitações similares àquela da Figura 6.6A e eventos de maior intensidade pluviométrica aqueles semelhantes à precipitação exibida na Figura 6.6B.

Figura 6.6: Precipitação média (azul) *versus* nível d'água registrado (vermelho). (A) ilustra as precipitações classificadas como eventos de menor intensidade pluviométrica; (B) ilustra as precipitações classificadas como eventos de maior intensidade pluviométrica.



Já a Figura 6.7 exibe a quantidade de eventos classificados como chuvas concentradas e contínuas além do grupo de *lag-time* associado a cada uma. Com base nesta figura, temos que para os 47 eventos de inundação e alagamentos registrados na bacia, 30 deles estão associados a chuvas concentradas e 17 associados a chuvas contínuas. Além disso, nota-se que dentre as chuvas concentradas, predominam-se os valores de *lag-time* de 30 minutos, enquanto para as chuvas contínuas, apesar dos números de eventos com valores de *lag-time* de 30 e 70 serem relativamente próximos (7 e 10 eventos, respectivamente), há um predomínio de valores de *lag-time* de 70 minutos.

Figura 6.7: Números de eventos classificados como eventos de menor ou de maior intensidade pluviométrica e seus respectivos *lag-time*.



6.3 Estimativa do tempo de concentração da bacia

A Tabela 6.5 exibe o tempo de concentração da bacia do ribeirão Proença obtido a partir das equações mencionadas no capítulo de métodos e os tempos de concentração obtidos a partir dos *lag-times* médios descritos no item anterior.

Tabela 6.5: Tempo de concentração da bacia do Ribeirão Proença calculada para diferentes equações.

Equação	Tempo de Concentração (minutos)
Kirpich modificada	108 minutos
Kirpich (California Culverts Practice)	72 minutos
Tsuchiya	60 minutos
Adaptação Chow <i>et al.</i> (1988)	65 minutos
<i>Lag-time</i> (30 minutos)	50 minutos
<i>Lag-time</i> (70 minutos)	116 minutos

Com base na Tabela 6.5, temos que o tempo de concentração da bacia obtido a partir das equações varia entre 60 e 110 minutos, aproximadamente. Verifica-se que os tempos de concentração obtidos para os diferentes *lag-times* médios se aproxima dos limites superior e inferior dos valores obtidos a partir das equações. Enquanto tempo de concentração

para o *lag-time* de 30 minutos apresentou resultado próximo aos obtidos pelas equações de Tsuchiya e a adaptação de Chow *et al.* (1988), o tempo de concentração para o *lag-time* de 70 minutos apresentou resultado próximo ao obtido pela equação de *Kirpich* modificada.

6.4 Base de dados unificada (BDU)

A base de dados unificada foi gerada a partir da integração de dados de nível d'água oriundos do sensor 4D-038 e dados de radar meteorológico. Os dados de nível d'água para o período de estudo, organizados de 5 em 5 minutos, como descrito anteriormente, totalizavam 50.543 dados. Os dados oriundos do radar meteorológico, também organizados de 5 em 5 minutos, mas apenas para as 3 horas que antecedem e que sucedem o valor máximo de nível d'água para os dias em que se registrou elevação maior que o Limiar Q3, totalizavam 12.265 dados.

A base de dados que une os dados comuns a estes dois bancos de dados tem 12.265 linhas e é composta pelas variáveis: data, horário, a precipitação de cada uma das dez quadriculas do radar meteorológico, o nível d'água, a temperatura da água e a precipitação média. A rotina em Python que realiza esta integração está disponível em <https://github.com/datasci4water>.

Assim sendo, a base de dados unificado (BDU) organizada nesta pesquisa tem 12.265 linhas e está disponibilizado no Repositório de Dados Institucionais da Unicamp (REDU) como Araujo *et al.* (2022) (<https://doi.org/10.25824/redu/J2NLAY>). Essa base de dados é composta por 15 colunas, sendo 2 delas referente a data e hora e as demais colunas as variáveis de interesse para a modelagem. A Tabela 6.6 exibe o nome das variáveis que compõe a BDU.

Tabela 6.6: Estrutura da base de dados unificada.

Colunas	Descrição
Date	Data do evento (dia, mês e ano)
Time_match	Hora, minuto e segundo do registro
79809	Precipitação na quadrícula 79809
78367	Precipitação na quadrícula 78367
78368	Precipitação na quadrícula 78368
78847	Precipitação na quadrícula 78847
78848	Precipitação na quadrícula 78848
79327	Precipitação na quadrícula 79327
79328	Precipitação na quadrícula 79328
79329	Precipitação na quadrícula 79329
79807	Precipitação na quadrícula 79807
79808	Precipitação na quadrícula 79808
Mean_rain	Precipitação média
Level_corr	Nível d'água
Temperature	Temperatura

Dentre as colunas supracitadas, as colunas Date, Time_match, Level_corr e Temperature foram obtidas por meio do sensor 4D-038, enquanto as colunas com números foram obtidas a partir de dados do radar meteorológico.

Por fim, para a modelagem foram adicionadas ao BDU mais três novas variáveis a saber: taxas de variação do nível do rio, da intensidade da precipitação e nível d'água 30 minutos no futuro, está última sendo a variável que desejamos prever. Para o cálculo destas

novas variáveis e pelo fato da série histórica de dados e precipitação não ser contínua, foi necessário suprimir algumas linhas de dados do BDU. Ao final, para previsão dos níveis d'água foi utilizado um *dataset* contendo um total de 9503 dados e 18 variáveis, onde 15 variáveis foram utilizadas como dados de entrada do modelo, uma como dado de saída e outras duas apenas para avaliação temporal.

6.5 Modelagem XGBoost

A Figura 6.8 exibe a subdivisão da série temporal dos dados em três subconjuntos de dados: treinamento (cinza), validação (roxo) e teste (vermelho). A proporção de dados utilizada foi de 80%, 10% e 10%, respectivamente. As setas em amarelo indicam os eventos de inundação na bacia no período estudado. Nota-se que o número de setas amarelas é de 17, enquanto o número de registros de inundações na bacia totaliza 18 (Tabela 6.4). Isso se deve ao fato da ausência de imagens de radar correspondente ao evento do dia 24/01/2019.

A partir da Figura 6.8, nota-se que os dados de treinamento correspondem ao período de 24/11/2014 até 06/02/2018, enquanto os dados de validação correspondem ao período de 10/02/2018 até 24/11/2018 e os dados de teste correspondem ao período de 25/11/2018 até 29/05/2019. O número total de eventos de inundações foi de 17 e em cada subgrupo mencionado foi de 13, 1 e 3, respectivamente.

Deve-se destacar que dos 17 eventos de inundações registrados, 16 ultrapassaram 3 metros de nível d'água. Por outro lado, 6 eventos ultrapassaram tal cota, porém não correspondem a eventos de inundação (3 deles estão associados a alagamentos na bacia, 2 a registros de eventos fora da bacia e um único evento sem informações). Cabe ressaltar que a última inundação não ultrapassou a cota de 3 metros.

Desta forma, a presente pesquisa considerou que a cota de 3 metros de nível d'água pode ser utilizada, com considerável confiança, como cota de alerta para a ocorrência de inundações na bacia, enquanto o Limiar Q3, que corresponde ao nível de 2,3 metros, será utilizado como cota de atenção para a ocorrência de inundações na bacia. Tais cotas supracitadas serão utilizadas para avaliar a capacidade do modelo gerado para prever inundações e a possível ocorrência de falsos alertas (alertas vazios).

Após a segregação dos dados supracitados, os dados de treinamento foram utilizados para definição dos hiperparâmetros e treinamento do modelo denominado de XGBoost 15VAR. A Tabela 6.7 exibe os resultados de avaliação do modelo XGBoost 15VAR.

Tabela 6.7: Resultados de avaliação do modelo XGBoost 15VAR para os dados de treinamento e validação.

Métrica	Resultado
RMSE (m)	0,268
R²	0,626
R² ajustado	0,582
NSE	0,626

Com base na Tabela 6.7, temos que o modelo XGBoost 15VAR possui um coeficiente de determinação (R^2) igual a 0,626, ou seja, consegue explicar cerca de 62,6% da variância dos dados.

Por outro lado, apresenta um R^2 ajustado igual a 0,582, fruto da penalização pela quantidade de variáveis presentes no modelo. No que tange a raiz quadrática média dos erros entre os valores observados e os preditos, o modelo apresentou o valor de 0,268 metros. A partir do resultado de NSE e com base na Tabela 5.2, temos que o modelo XGBoost 15VAR possui um desempenho satisfatório.

Após a avaliação do modelo XGBoost 15VAR, novos modelos foram gerados a partir da adição de novas variáveis, conforme exposto na Tabela 5.4.

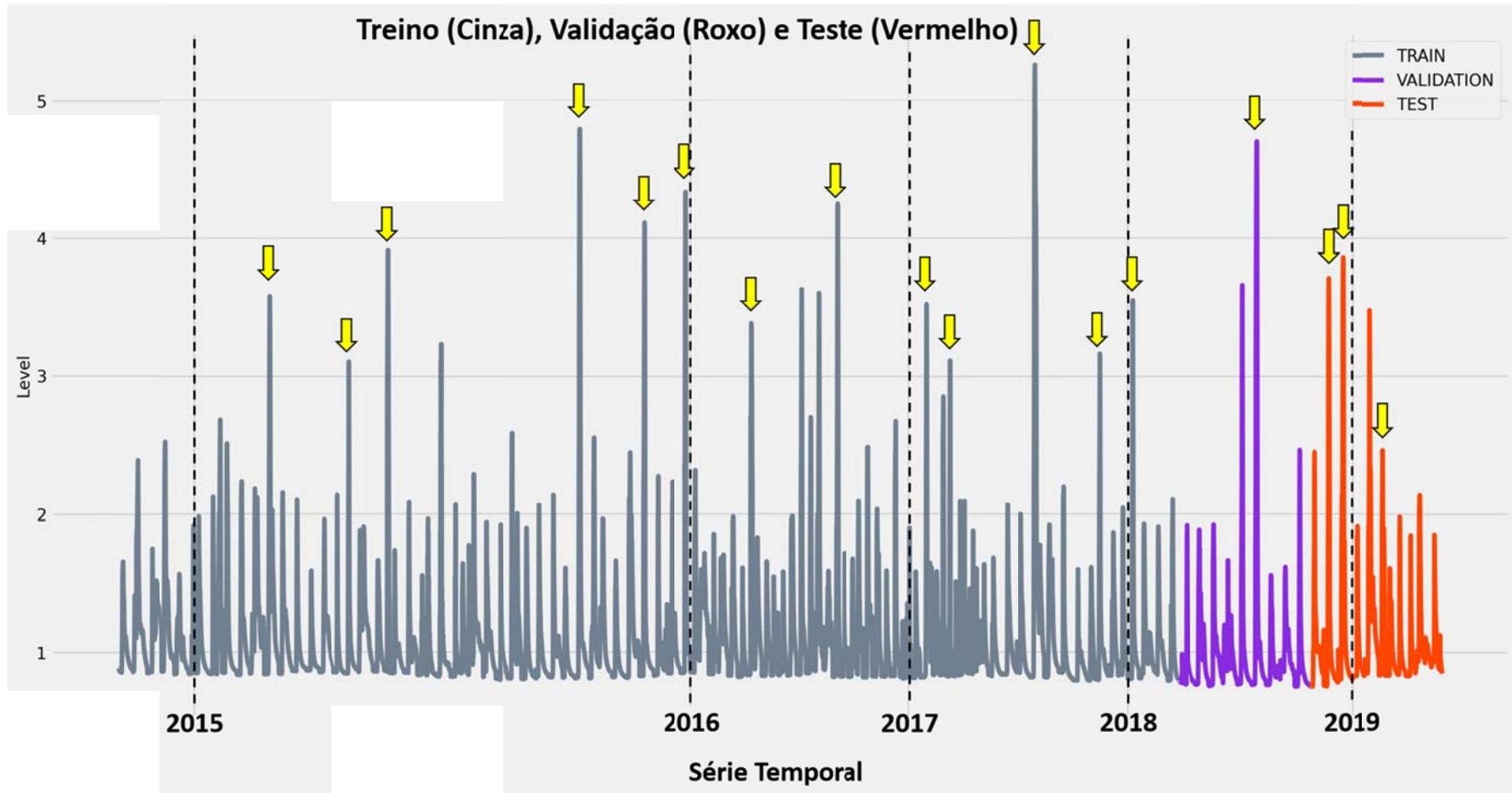
A Tabela 6.8 exhibe os resultados da avaliação dos modelos gerados a partir da adição de novas variáveis.

Tabela 6.8: Resultado de avaliação dos modelos gerados a partir da adição de novas variáveis defasadas no tempo. Como exemplo, o modelo XGBoost 15VAR + N5 representa o modelo padrão mais a adição da variável nível d'água (N) 5 minutos no passado.

	Modelos de Teste		
Métrica	XGBoost 15VAR + N5	XGBoost 15VAR + N10	XGBoost 15VAR + N5-10
RMSE (m)	0,299	0,395	0,392
R²	0,627	0,623	0,627
R² ajustado	0,562	0,378	0,387
NSE	0,627	0,623	0,627
Métrica	XGBoost 15VAR + PM5	XGBoost 15VAR + PM10	XGBoost 15VAR + PM5-10
RMSE (m)	0,293	0,382	0,378
R²	0,623	0,620	0,621
R² ajustado	0,578	0,419	0,431
NSE	0,623	0,620	0,621

Com base nos resultados presentes na Tabela 6.8, nota-se que em alguns casos houve um aumento sutil do R² e NSE, o que poderia indicar uma possível melhoria do modelo. Contudo, de modo geral, temos que o valor de RMSE aumentou ao adicionarmos as novas variáveis, indicando assim uma queda no desempenho dos novos modelos. Além disso, nota-se que o R² ajustado caiu significativamente, indicando uma penalização do resultado fruto da adição de variáveis pouco significativas para o modelo.

Figura 6.8: Segmentação do banco de dados em dados de treino (Cinza), Validação (Roxo) e Teste (Vermelho). As setas em amarelo indicam os dias em que há registros de inundação na bacia do ribeirão Proença.



Desta forma, podemos constatar que a adição de novas variáveis prejudicou o modelo previamente gerado (XGBoost 15VAR). Dito isso, o modelo final da presente pesquisa, modelo que será avaliado pelos dados de teste, será o modelo denominado de XGBoost 15VAR.

6.6 Avaliação do Modelo Final (XGBoost 15VAR)

A Tabela 6.9 exibe as estatísticas avaliativas do modelo final (XGBoost 15VAR) calculadas para os dados de teste do modelo. Com base na Tabela 6.9, nota-se que modelo gerado apresentou um coeficiente de determinação (R^2) igual a 0,613, denotando assim uma cobertura de 61,3% da variância dos dados de teste. No que tange a raiz quadrática média dos erros (RMSE) entre os valores observados e os preditos, o modelo apresentou o valor de 0,284 metros, um valor que pode ser considerado baixo.

Tabela 6.9: Avaliação do modelo final XGBoost 15VAR aplicado aos dados de teste.

Métrica	Resultado
RMSE (m)	0,284
R^2	0,613
R^2 ajustado	0,540
NSE	0,614
MAPE (%)	9,78

Ademais, nota-se que o modelo gerado realiza previsões que diferem, em média, cerca de 9,78% dos valores reais (MAPE). Por fim, com base no NSE observado tal modelo pode ser classificado como satisfatório. Deve-se destacar que os valores dos hiperparâmetros obtidos que apresentaram os melhores resultados foram: *colsample_bytree* = 0.7, *learning_rate* = 0.01, *max_depth* = 7, *n_estimators* = 500, *nthread* = 24, *objective* = 'reg:squarederror' e *subsample* = 0.75.

A Figura 6.9 exibe as previsões realizada pelo modelo para os dados de treino, validação e teste. Note que para os dados de treino (azul) e validação (verde) o modelo gerado apresentou ótimas previsões, enquanto para os dados de teste o modelo apresentou previsões razoáveis.

Cabe ressaltar que a previsão dos níveis d'água realizada pelo modelo para os dados de treinamento e validação apresentou apenas um único caso de falso alerta (estrela amarela), onde o nível previsto pelo modelo ultrapassou o Limiar Q3 (nível de atenção), enquanto os dados reais (em preto) denotam um nível d'água abaixo de tal limiar. Além disso, é possível notar que de modo geral o modelo subestima os níveis d'água, principalmente em momentos em que o nível d'água atinge valores elevados.

A Figura 6.10 exibe a previsão detalhada dos níveis d'água (em rosa) para os dados de teste. Embora os níveis previstos pelo modelo apresentem uma defasagem de tempo quando comparados com os dados reais e subestimem os picos de nível d'água, nota-se que a previsão realizada apresenta uma razoável conformidade com os dados reais (em preto), corroborando com a classificação satisfatória atribuída ao modelo devido ao valor do NSE.

No que tange aos 3 eventos de inundações pertencentes aos dados de teste, em 2 o modelo foi capaz de prever níveis acima do Limiar Q3, remetendo assim à detecção antecipada da ocorrência de níveis acima do nível de atenção, embora em nenhum dos eventos o modelo tenha sido capaz de prever níveis acima de 3 metros.

Figura 6.9: Predições do nível d'água 30 minutos no futuro realizadas pelo modelo XGBoost 15VAR para os dados de treinamento (azul), validação (verde) e teste (rosa). As linhas em preto representam os níveis reais registrados, enquanto as setas amarelas destacam os eventos de inundações na bacia.

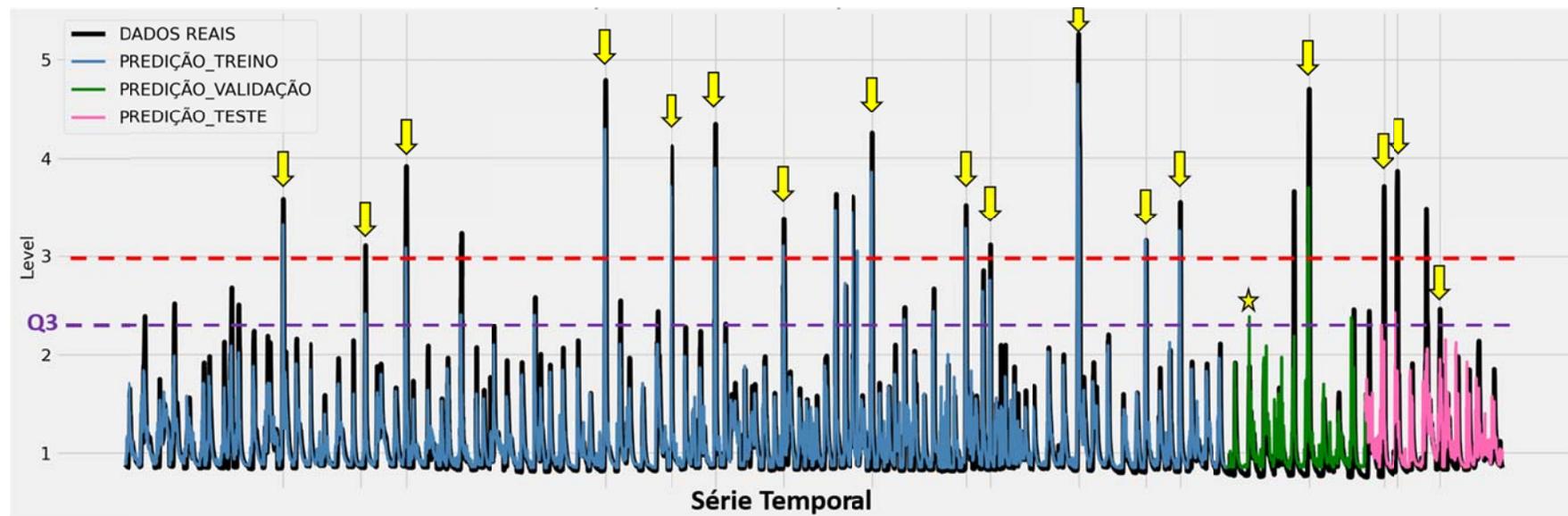
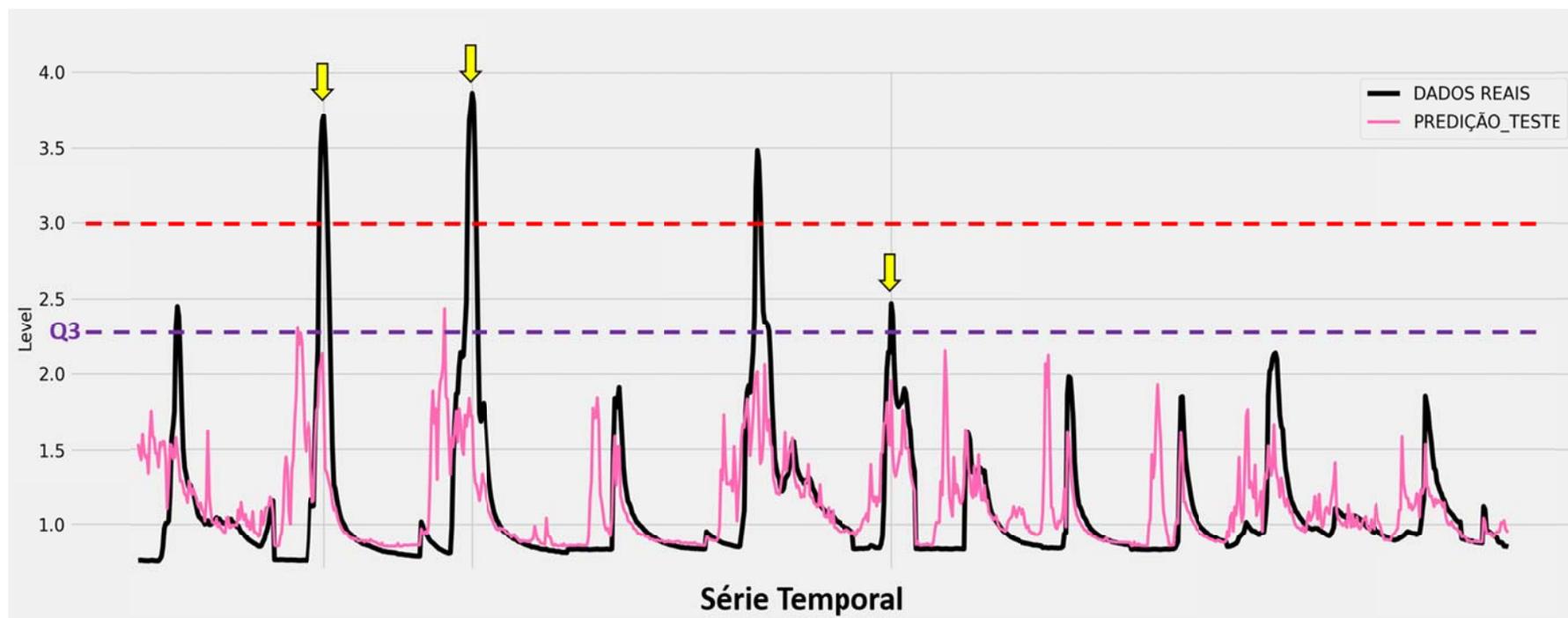
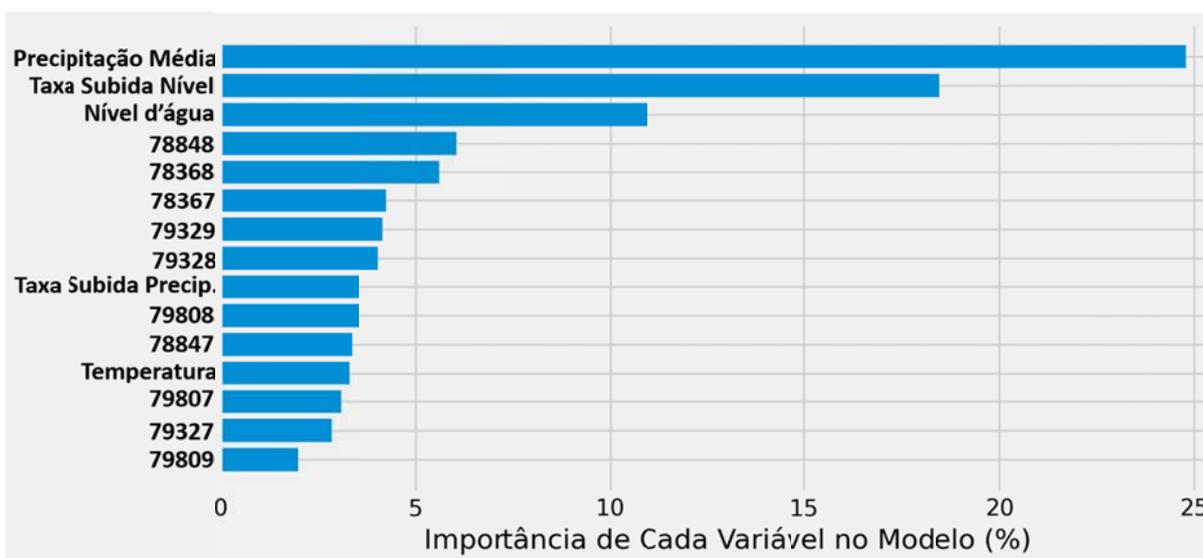


Figura 6.10: Dados reais (em preto) *versus* dados preditos pelo modelo XGBoost 15VAR (em rosa). As setas em amarelo indicam os eventos de inundação na bacia, enquanto as setas em roxo e vermelho indicam os limiares de atenção e alerta, respectivamente.



A Figura 6.11 denota a importância de cada uma das 15 variáveis utilizadas para o modelo final gerado. Note que a precipitação média da bacia possui, por si só, uma contribuição de aproximadamente 25% para o modelo, seguida pela taxa de subida do nível d'água e pelo valor do nível d'água.

Figura 6.11: Importância em porcentagem de cada uma das 15 variáveis utilizadas para o modelo.



Dentre as 10 quadrículas de radar meteorológico utilizadas como variáveis (Figura 5.2), temos que as quadrículas mais próximas da foz do ribeirão Proença (78848, 78368 e 78367), por sua vez mais próximas do sensor, apresentaram maiores pesos para o modelo, indicando assim uma maior importância para a previsão dos dados futuros. Por outro lado, as quadrículas que apresentaram menores pesos para o modelo foram as que possuem maiores distâncias da foz do ribeirão, dentre elas: 79807, 79327 e 79809.

Nota-se também que as quadrículas localizadas na porção central da bacia, quadrículas terminadas em 8, possuem maior importância quando comparadas com as quadrículas localizadas nas laterais, sugerindo assim um reflexo do tipo de ocupação urbana na bacia.

De modo geral, podemos destacar as quadrículas analisadas individualmente apresentaram valores de importância para o modelo menores que 5%. Apenas as quadrículas 78848 e 78368, localizadas próximas a foz do ribeirão Proença e que abrangem o córrego Serafim, afluente do ribeirão Proença, exibiram valores maiores que 5%.

Por fim, tanto a variável taxa de subida da precipitação média quanto a temperatura da água apresentaram importâncias para o modelo menores que 5%.

6.7 Modelagem para diferentes tempos futuros

A Tabela 6.10 compara o desempenho do modelo XGBoost 15VAR na previsão de níveis d'água 30 minutos no futuro com o desempenho de novos modelos gerados para diferentes tempos de previsões futuras.

A partir da Tabela 6.10, fica nítido que à medida que diminuimos o tempo de previsão futura, diminuimos o valor de RMSE, enquanto aumentamos o valor de R^2 e NSE, indicando assim um melhor desempenho do modelo. Por outro lado, à medida que aumentamos o tempo de previsão futura, diminuimos o desempenho do modelo e consequentemente a sua capacidade de realizar previsões.

No que tange à importância das variáveis para os modelos gerados, note que para previsões temporais mais próximas, tais como 10 e 15 minutos, o modelo atribui importâncias mais elevadas as variáveis relacionadas ao nível d'água observado (nível e taxa de subida do nível d'água), em detrimento das variáveis associadas a precipitação, tais como a precipitação média.

Em contrapartida, para as previsões temporais mais distantes, tais como 30 e 40 minutos, as variáveis que recebem maiores importâncias são as associadas a precipitação na bacia, como exemplo a variável precipitação média.

Além disso, são atribuídos pesos cada vez menores as variáveis associadas ao nível d'água. Deve-se destacar que, para tempos superiores a 40 minutos o desempenho do modelo deixa de ser satisfatório.

Tabela 6.10: Desempenho de diferentes modelos para a previsão de distintos níveis d'água futuros.

Modelo 15VAR para previsão de nível d'água futuro	R²	RMSE (m)	NSE	Importância das Variáveis		
				Nível	Taxa Subida Nível	Precipitação Média
10 minutos	0,883	0,150	0,883	53%	14%	5%
15 minutos	0,844	0,173	0,844	43%	16%	6%
20 minutos	0,747	0,219	0,748	28%	20%	17%
25 minutos	0,681	0,256	0,683	23%	19%	17%
30 minutos	0,613	0,284	0,614	25%	18%	11%
40 minutos	0,541	0,315	0,543	24%	12%	6%

7. CONCLUSÃO

A partir da análise dos dados de nível d'água coletados pelo sensor 4D-038 na foz do ribeirão Proença, área de estudo deste trabalho, para o período de estudo estabelecido, referente a novembro de 2014 até junho de 2019, três limiares de nível d'água no ribeirão Proença foram avaliados e comparados com registros de ocorrências de inundações, a fim de verificar suas respectivas correlações. Com base nesta análise, considerou-se que o nível d'água de 2,30 metros (limiar correspondente à permanência de 3% do tempo) poderia ser utilizado como um nível de atenção para a ocorrência de eventos de inundação na bacia do ribeirão Proença.

Considerando-se apenas os 22 eventos em que o nível d'água de 3,00 metros foi ultrapassado, verificou-se que 16 destes estavam relacionados a eventos de inundação na bacia, 3 associados a alagamentos na bacia, 2 a registros fora da bacia e um único evento sem informações. Desta forma, a presente pesquisa considerou que o registro de 3,00 metros de nível d'água poderia ser utilizado como nível de alerta para a ocorrência de inundações na bacia.

Nesta pesquisa foi criado um banco de dados reais, a partir da integração de dados de diferentes fontes, compondo um banco de dados unificado multimodal. Este banco de dados é composto por 15 variáveis, a saber: nível d'água medido no sensor 4D-038, temperatura da água medida no sensor 4D-038, precipitação em cada quadrícula obtida a partir de dados de radar meteorológico (sendo 10 quadrículas no total), precipitação média, além da data e da hora de aquisição dos dados. Este banco de dados está disponibilizado no Repositório de Dados de Pesquisa da Unicamp e foi utilizado para o desenvolvimento de um modelo baseado em dados para previsão de níveis no Ribeirão Proença.

Ao avaliar as estatísticas do modelo gerado para previsão de níveis d'água 30 minutos à frente, constatou-se que ele apresentou uma abrangência de 61,3% da variância total dos dados de teste, enquanto o valor de NSE foi de 0,614, denotando assim um modelo satisfatório. Das 15 variáveis empregadas, as variáveis precipitação média na bacia, taxa de subida do nível d'água e valor do nível d'água foram as que apresentaram maiores importâncias para o modelo. Constatou-se que o emprego de novas variáveis defasadas no tempo, além das 15 variáveis iniciais, prejudicou o desempenho do modelo. Logo, o modelo escolhido para subsequente previsão de níveis d'água foi o modelo inicial, denominado de XGBoost 15VAR.

Na avaliação do modelo XGBoost 15VAR com os dados de teste, notou-se que os níveis previstos, além de apresentarem uma defasagem no tempo, subestimaram os picos de níveis d'água, principalmente nos eventos onde o nível d'água atingiu valores extremos.

Por outro lado, dos 3 eventos de inundações pertencentes aos dados de teste, em 2 o modelo foi capaz de prever níveis acima do nível de atenção (Limiar 3), remetendo assim a uma assertividade de 66%.

Ao realizar a modelagem para diferentes tempos futuros, notou-se que à medida que diminuía o tempo futuro no qual o nível d'água era previsto, havia melhora no desempenho do modelo e ele atribuía importâncias mais elevadas para variáveis relacionadas ao nível d'água, diminuindo assim a importância da precipitação.

Por outro lado, para previsões de níveis d'água em tempos mais distantes, tais como 30 e 40 minutos, os modelos gerados apresentaram piores desempenhos. Porém, diferentemente dos modelos gerados para previsões temporais mais próximas, tais modelos atribuíram importâncias mais elevadas para as variáveis associadas à precipitação em detrimento das variáveis associadas a nível d'água. Para tempos superiores a 40 minutos, o desempenho do modelo foi insatisfatório (NSE inferior a 0,5), evidenciando assim a limitação do modelo para previsão de níveis d'água para tempos muito distantes.

Analisando-se o *lag-time* entre a precipitação e a subida do nível d'água para os 145 eventos delimitados pelo Limiar 3 (nível de atenção), constatou-se a presença de dois grupos distintos de valores de *lag-time*, sendo estes: 30 e 70 minutos. Deve-se destacar que independente dos grupos, para maiores valores de níveis d'água, maiores foram os valores de correlação obtidos, sugerindo assim a existência de uma elevada importância da precipitação na previsão de níveis d'água mais elevados. Este resultado, se comparado ao desempenho do modelo para previsões de mais curto e longo prazo, sugere que há grande variabilidade intrínseca ao problema, que ainda não está sendo completamente capturada pelo modelo baseado em dados gerado na presente pesquisa.

Além disso, ao avaliar o tipo de chuva predominante em cada grupo de *lag-time* supracitado, notou-se que os eventos pertencentes ao grupo de 30 minutos estão associados a precipitações de maior intensidade, enquanto os eventos pertencentes ao grupo de 70 minutos estão associados a precipitações de menor intensidade. Este resultado sugere que o padrão de precipitação controla os tempos de concentração na bacia do ribeirão Proença. Estudar esta correlação também é uma alternativa para melhor compreensão dos fenômenos na bacia e melhor desenvolvimento de modelos, sejam eles físicos ou baseados em dados.

Como trabalhos futuros, sugere-se ainda a aquisição de um maior número de dados de nível d'água e precipitação a fim de diminuir o efeito de subestimação dos valores de nível d'água máximo no modelo gerado, além de melhorar o desempenho do modelo para tempos futuros próximos e distantes. Outra proposta de trabalho futuro, atrelada à aquisição de um maior número de dados, é a utilização de algoritmos de *Machine Learning* baseados em séries temporais, como é o caso do modelo denominado de *Long Short-Term Model* (LSTM).

8. REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMI, S.F.; NOGUEIRA, F.P.; MORAES, J.F.L.; COSTA, D.C.; PEREZ FILHO, A.; CARVALHO, D.RE.; PRADO, M.S.B. & VUKOMANOVIC, C.R. Cartografia, sensoriamento remoto e sistemas de informações geográficas. In: TORRES, R.B.; COSTA, M.C.; NOGUEIRA, F.P. & PEREZ FILHO, A. (coord.). Recuperação ambiental, participação e poder público: uma experiência em Campinas. 2006. Relatório de Pesquisa. Disponível em: <<http://www.iac.sp.gov.br/projetoanhumas/index.htm>>. Acesso em: 10 de janeiro de 2022.
- ADHIKARI, T., RAKHAL, B., MAHARJAN, L. D., & KUSHI, P. S. Determination of flood warning and danger of Mohana and Macheli river. *Final Report*, Mercy Corps Nepal, p. 157, 2014.
- ADNAN, R. M., LIANG, Z., HEDDAM, S., ZOUNEMAT-KERMANI, M., KISI, O., LI, B. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *Journal of Hydrology*, 586:124371, 2019.
- ALFIERI, L., SALAMON, P., PAPPENBERGER, F., WETTERHALL, F., & THIELEN, J. Operational early warning systems for water-related hazards in Europe. *Environmental Science & Policy*, 21, p. 35-49. 2012.
- ALFIERI, L., BUREK, P., DUTRA, E., KRZEMINSKI, B., MURARO, D., THIELEN, J., & PAPPENBERGER, F. GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, 17(3), 1161-1175, 2013.
- ALIZAMIR, M., KISI, O., ZOUNEMAT-KERMANI, M. Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. *Hydrological Sciences Journal*. v. 63, n. 1, p. 63–73, 2018.
- AMARAL, R.; RIBEIRO, R.R. Inundação e Enchentes. In: TOMINAGA, L.K.; SANTORO, J.; AMARAL, R. 2009. Desastres Naturais: conhecer para prevenir. São Paulo, Instituto Geológico, p. 39-52. 2009.
- APEL, H.; MARTINEZ TREPAT, O., HUNG, N.N., CHINH, D.T., MERZ, B. & DUNG, N.V. Combined fluvial and pluvial urban flood hazard analysis: concept development and application to Can Tho city, Mekong Delta, Vietnam. *Nat. Hazards Earth Syst Sci* 2016, 16, p. 941–961. 2016.
- ARAUJO, V.; ABREU, A.E.S.; COSTA, P.D.P.; FALCETTA, F.A.M.; PEREZ, G.A.C.; Bitar, O.Y. Multimodal hydrological dataset for flash flood modeling in an urban watershed. Repositório de Dados de Pesquisa da Unicamp. 2022. <https://doi.org/10.25824/redu/J2NLAY>.

- ASWAD, F. M., KAREEM, A. N., KHUDHUR, A. M., KHALAF, B. A., & MOSTAFA, S. A. Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction. *Journal of Intelligent Systems*, v. 31(1), p. 1-14. 2022.
- BARBAROTTO JÚNIOR, J. L.; ZUFFO, A. C. Mapeamento sistemático da literatura sobre a influência de fatores desencadeadores para ocorrência de inundações bruscas. *XXII Simpósio Brasileiro de Recursos Hídricos (SBRH)*, Vol. 1, pp.1-8, Florianópolis, SC, Brasil, 2017.
- BARCELLOS, P. da C. L.; SILVA, F. P. da.; VISSIRINI, C. A. M.; TERRA, J. M.; DUTRA, M. R. F.; AMARAL, I. C. F. Diagnóstico Meteorológico dos Desastres Naturais Ocorridos nos Últimos 20 Anos na Cidade de Duque de Caxias. *Revista Brasileira de Meteorologia*, v. 31, n. 3, p. 319-329, 2016.
- BORGA, M.; ANAGNOSTOU, E.N.; BLÖSCHL, G.; CREUTIN, J.D. Flash flood forecasting, warning and risk management: the HYDRATE project. *Environmental Science & Policy*, v. 14, n. 7, p. 834-844. 2011.
- BREINL, K. Pluvial and fluvial flooding: integration in probabilistic flood hazard assessment using a coupled rainfall-discharge generator. In: EGU General Assembly Conference Abstracts. p. 1990. 2012.
- BRIGUENTI, Ederson Costa. O uso de geoindicadores na avaliação da qualidade ambiental da Bacia do Ribeirão Anhumas, Campinas/SP. 2005. 179p. Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Geociências, Campinas, SP.
- BROWN, S., BUDIMIR, M., LOVELL, E., WIKINSON, E., & MEECHAIYA, C. The governance of Nepal's flood early warning system: opportunities under federalism. 2019.
- BRUBACHER, J. P.; OLIVEIRA, G.G.; GUASSELLI, L.A. Suscetibilidade de enchentes a partir da análise das variáveis morfométricas na bacia hidrográfica do rio dos Sinos/RS. In: XV SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO - SBSR, Curitiba, Anais. Curitiba: INPE, 2011. p.1279-1286. 2011.
- CANHOLI, A. P. Drenagem Urbana e Controle de Enchentes. Oficina de Textos. São Paulo, 2005.
- CARVALHO, C. S.; MACEDO, E. S; OGURA, A.T; (Orgs.). Mapeamento de Riscos em encostas e margens de rios. Ministério das Cidades / IPT – Instituto de Pesquisas Tecnológicas. Brasília, 176 p. 2007.
- CASTELLANO, M. S. Extremos de chuva na Região Metropolitana de Campinas (SP): impactos, análise socioeconômica e políticas públicas. 2016. (290 p.). Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Geociências, Campinas, SP.

- CASTRO, A. L. C. Manual de Desastres Naturais. v. 1. Brasília (DF): Ministério da Integração Nacional, 2003.
- CHOUDHURY, N. Y.; POUL, A.; POUL, B.K. Impact of costal embankment on the flash flood in Bangladesh: a case study. *Applied Geography* v. 24, pp. 241-258. 2004.
- CHOW, V. T.; MAIDMENT, D. R. & MAYS, L. W. Applied Hydrology. McgrawHill Book Company, New York, 572p., 1988.
- COELHO, R. M.; GOMES, D. C. H.; ROSSI, M.; MENK, J. R. F.; ADAMI, S. F.; SOUZA, L. A. de; GUADALUPE. G. W. Levantamento pedológico semidetalhado da bacia do ribeirão das anhumas. In: TORRES, R. B.; COSTA, M. C.; NOGUEIRA, F. P. & PEREZ FILHO, A. (coord.). Recuperação ambiental, participação e poder público: uma experiência em Campinas. 2006. Relatório de Pesquisa. Disponível em: <<http://www.iac.sp.gov.br/projetoanhumas/index.htm>>. Acesso em: 10 de janeiro de 2022.
- COSTA, C. M., da SILVA, Q. D., dos SANTOS COSTA, I. R., Barros, D. V., & TEIXEIRA, C. Suscetibilidade a inundações a partir das variáveis morfométricas na região hidrográfica Eliezer Silva--alto curso do rio Paciência. *Entorno Geográfico*, v. 13, p. 68-81. 2017.
- CRED Crunch 58 - Disaster Year in Review (2019). Brussels: CRED, 2020. Disponível em: <<https://www.cred.be/publications>>. Acesso em 04 de janeiro de 2022.
- CRED Crunch 62 - 2020 Annual Report. Brussels: CRED, 2021. Disponível em: <<https://www.cred.be/publications>>. Acesso em 04 de janeiro de 2022.
- CREUTIN, J. D.; BORGA, M. “Radar hydrology modifies the monitoring of flash-flood hazard.” *Hydrological Processes*, v. 17, p. 1453-1456, 2003.
- CHRISTOFOLETTI, A. Geomorfologia Fluvial. 1ed. São Paulo: Edgard Blucher. 1981.
- DAWSON, R. J., BALL, T., WERRITTY, J., WERRITTY, A., HALL, J. W., & ROCHE, N. Assessing the effectiveness of non-structural flood management measures in the Thames Estuary under conditions of socio-economic and environmental change. *Global Environmental Change*, v. 21, n.2, p. 628-646. 2011.
- DNIT - Departamento Nacional de Infra_Estrutura de Transportes, 2005. Manual de Hidrologia Básica para Estruturas de Drenagem. Publicação IPR-715. 2ª Edição. Rio de Janeiro, 2005.
- FAISAL, I. M., KABIR, M. R., NISHAT, A. Non-structural flood mitigation measures for Dhaka City. *Urban Water* v1, p.145–153. 1999.
- FEDERAL EMERGENCY MANAGEMENT AGENCY (FEMA). Design guidelines for flood damage reduction, 1981.

- FERREIRA, R. S.; LIMA, G. S.; AMORIM, R. R. Áreas suscetíveis a inundações e alagamentos na bacia hidrográfica do Córrego Proença, Campinas-SP. *In: Simpósio Brasileiro de Geografia Física Aplicada*, XVIII, 2019, Fortaleza - CE. pag. 1 – 11. 2019.
- FOTOVATIKHAH, F., HERRERA, M., SHAMSHIRBAND, S., ARDABILI, S.F., PIRAN, J. Mechanics Survey of computational intelligence as basis to big flood management: challenges, research directions and future work. *Engineering Applications of Computational Fluid Mechanics*, v. 12, n. 1, p. 411-437, 2018.
- FURQUIM, G. A. Uma abordagem tolerante a falhas para a previsão de desastres naturais baseada em IoT e aprendizado de máquina. Tese de Doutorado. Universidade de São Paulo (USP), São Paulo, p. 111, 2017.
- GAUTAM, D. K., & DULAL, K. Determination of threshold runoff for flood early warning in Nepalese Rivers. *IDRiM Journal*, v. 3, n.1, p. 126-136. 2013.
- GEORGAKAKOS, K. P. "On the design of natural, real-time warning systems with capability for site-specific, flash-flood forecast". *Bulletin American Meteorological Society*, v 67, n.10, pp. 1233-1239. 1986.
- Glen, S. "Mean Absolute Percentage Error (MAPE)". *Statistics How To. Elementary Statistics for the rest of us!*. 2022. Disponível em: <<https://www.statisticshowto.com/mean-absolute-percentage-error-mape/>>. Acesso em: 23 de agosto de 2022.
- GOERL, R.F.; KOBIYAMA, M. Considerações sobre as Inundações no Brasil. XVI SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 2005, João Pessoa. Anais... Porto Alegre, ABRH, 2005.
- HAN, L.; XU, Y.; PAN, G.; DENG X.; HU, C.; XU, H.; SHI, H. Changing properties of precipitation extremes in the urban areas, Yangtze River Delta, China, during 1957–2013. *Natural Hazard*, v.79, p. 437–454, 2015.
- HILL, C.; VERJEE, F.; BARRETT, C. Flash flood early warning system reference guide. University Corporation for Atmospheric Research: Boulder, CO, USA, 2010.
- IGNÁCIO, C. F. Considerações sobre a drenagem urbana: o caso das inundações em Campinas (2016-2018). 2020. 114 p. Dissertação (Programa de Pós-Graduação em Urbanismo) - Pontifícia Universidade Católica de Campinas, Campinas-SP.
- INSTITUTO GEOLÓGICO. Mapa geológico do município de Campinas e mapas de pontos de descrição geológica e de pontos de descrição geomorfológica. Instituto Geológico, Secretaria do Meio Ambiente, Governo de São Paulo, 2009.

- JAVELLE, P., SAINT-MARTIN, C., GARANDEAU, L., & JANET, B. Flash flood warnings: Recent achievements in France with the national Vigicrues Flash system. United Nations Office for Disaster Risk Reduction, Contributing Paper to the Global Assessment Report on Disaster Risk Reduction (*GAR 2019*), v. 60, p. 20. 2019.
- JACOMAZZI, M. A. Avaliação de cenários para elaboração de plano diretor de macrodrenagem urbana: Estudo de caso: bacia do Ribeirão das Anhumas em Campinas SP. 2015. Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Campinas, SP, 263 p.
- KOBIYAMA, M.; MENDONÇA, M.; MORENO, D. A.; MARCELINO, I. P. V. O.; MARCELINO, E. V.; GONÇALVES, E. F.; BRAZETTI, L. L. P.; GOERL, R. F.; MOLLERI, G.; RUDORFF, F. Prevenção de desastres naturais: conceitos básicos. Curitiba: Organic Trading, p. 109. 2006.
- KRON, W. “Keynote lecture: Flood risk = hazard x exposure x vulnerability”. Proceedings of Second International Symposium of Flood Defense, Beijing, p. 82-97. 2002.
- KRON, W.; CHENG, X. China Floods 2016: “Like a hole in the sky”. *Water Solutions*, v. 3, n. 2017, p. 18-21, 2017.
- KUNDZEWICZ, Z. W., SU, B., WANG, Y., XIA, J., HUANG, J., & JIANG, T. Flood risk and its reduction in China. *Advances in Water Resources*, v. 130, p. 37-45. 2019.
- LI, D.; FANG, Z. N.; BEDIENT, P. B. Flood early warning systems under changing climate and extreme events. In: *Climate Change and Extreme Events*. Elsevier, p. 83-103, 2021.
- LIMA, G. R. T; SCOFIELD, G. B. Feasibility study on operational use of neural networks in a flash flood early warning system. *Revista Brasileira de Recursos Hídricos*, v. 26, 2021.
- LIU, C. J.; GUO, L.; YE, L.; ZHANG, S. F.; ZHAO, Y. Z.; SONG, T. Y. A review of advances in China’s flash flood early-warning system. *Natural Hazards*, v. 92, p. 619-634, 2018.
- LY, S., CHARLES, C., DEGRÉ, A. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. *Rev. Biotechnol. Agron. Soc. Environ.* v. 17 (2013), p. 392. 2013.
- MA, M., WANG, H., YANG, Y., ZHAO, G., TANG, G., HONG, Z., ... & HONG, Y. Development of a new rainfall-triggering index of flash flood warning-case study in Yunnan province, China. *Journal of Flood Risk Management*, v. 14, n.1, p. e12676. 2021.
- MADDOX, I. Three Common Types of Flood Explained. Intermap, 2014. Disponível em: <<https://www.intermap.com/risks-of-hazard-blog/three-common-types-of-flood-explained>>. Acesso em: 30 de março de 2021.

MARCELINO, E. V.; NUNES, L. H.; KOBİYAMA, M. Banco de dados de desastres naturais: análise de dados globais e regionais. Uberlândia: Caminhos de Geografia. Outubro de 2006. Vol. 6, Nº 19. p 130 -149. 2006.

MARCHI, Lorenzo et al. Characterisation of selected extreme flash floods in Europe and implications for flood risk management. *Journal of Hydrology*, v. 394, n. 1-2, p. 118-133, 2010.

MEHDIZADEH, S., FATHIAN, F., SAFARI, M. J. S., ADAMOWSKI, J. F. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach. *Journal of Hydrology*, v. 579, p. 124225. 2019.

MONTE, B. E. O., COSTA, D. D., CHAVES, M. B., MAGALHÃES, L. D. O., & UVO, C. B. (2016). Modelagem hidrológica e hidráulica aplicada ao mapeamento de áreas inundáveis. *RBRH*, 21, 152-167.

MORIASI, D. N., ARNOLD, J. G., VAN LIEW, M. W., BINGNER, R. L., HARMEL, R. D., & VEITH, T. L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900. 2007.

MÜLLER, A. S. Aplicabilidade de medidas não-estruturais no controle de cheias urbanas: estudo de caso Bacia Hidrográfica do Rio Cachoeira-Joinville-SC. Tese de Mestrado em Engenharia – Escola Politécnica, Universidade de São Paulo. São Paulo. p. 239. 2011.

MURALIDHAR K. S. V. Demystifying R-Squared and Adjusted R-Squared. Towards Data Science. 2021. Disponível em: <<https://towardsdatascience.com/demystifying-r-squared-and-adjusted-r-squared-52903c006a60>>. Acesso em: 23 de Agosto de 2022.

MUTHUSAMY, M., CASADO, M. R., BUTLER, D., & LEINSTER, P. Understanding the effects of Digital Elevation Model resolution in urban fluvial flood modelling. *Journal of Hydrology*, 596, 126088. 2021.

NOAA. Flash Flood Early Warning System Reference Guide; University Corporation for Atmospheric Research: Denver, CO, USA, 2010. Disponível em: <https://www.meted.ucar.edu/training_module.php?id=958>. Acesso em: 17 de novembro de 2021.

NOURANI, V., KISI, O., KOMASI, M. Two hybrid Artificial Intelligence approaches for modeling rainfall– runoff process. *Journal of Hydrology*. v. 402, n. 1-2, p. 41–59, 2011.

NWS/NOAA - NATIONAL WEATHER SERVICE/NATIONAL OCEANIC ATMOSPHERIC ADMINISTRATION. Glossary. 2004.

OCHOA-TOCACHI, B.F., BUYTAERT, W., DE BIÈVRE, B. Regionalization of land-use impacts on streamflow using a network of paired catchments. *Water Resour. Res.* v. 52, n. 9, p. 6710–6729, 2016.

OLIVEIRA, L. M. Guia de prevenção de acidentes geológicos urbanos. Curitiba: Mineropar, 1998.

OLIVEIRA, J. V. P.; LUCENA, D. B.; LIMA, P. R. C. Avaliação dos eventos extremos de chuva na região pluviometricamente homogênea do Alto Sertão da Paraíba entre 1994-2016. Universidade Federal de Santa Maria, Ci. e Nat., Santa Maria v.42, e102, 2020.

PLANO DIRETOR DO MUNICÍPIO DE CAMPINAS, 2018. Plano Diretor Estratégico – Prefeitura de Campinas – Lei Complementar n° 189 de 08 de janeiro de 2018. Mapa de Suscetibilidade a Inundações, Secretaria de Planejamento e Urbanismo, 2018. Disponível em: <https://planodiretor.campinas.sp.gov.br/timeline/timeline/59_mapas_finais_pd2018/anexoX_X_suscet_inundacoes.pdf>. Acesso em: 15 de janeiro de 2022.

PRACTICAL ACTION. Flood Early Warning Systems in Practice: experiences of Nepal, Practical Action, Kathmandu, Nepal, 2016.

PRINA, Z. B; TRETIN, R. Metodologia para Mapeamento de Áreas Suscetíveis à Inundação: Estudo de Caso para o Município de Jaguari – RS. Anais XXVI Congresso Brasileiro de Cartografia, Gramado, RS, Brasil, 03-07 de agosto de 2014, n.274. 2014.

RAGETTLI, S., CORTES, G., MCPHEE, J., PELLICCIOTTI, F. An evaluation of approaches for modeling hydrological processes in high-elevation, glacierized Andean watersheds. *Hydrological. Processes*, v. 28, n. 23, p. 5674–5695, 2014.

RAPANT, P.; KOLEJKA, J. Dynamic Pluvial Flash Flooding Hazard Forecast Using Weather Radar Data. *Remote Sensing*, v. 13, n. 15, p. 2943, 2021.

RIBEIRO, P. H. B; ABREU, A. E. S. Mapeamento de áreas de inundação na bacia do Córrego Proença em Campinas, SP. In: CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UNICAMP, 22p., 2021, Campinas. Anais eletrônicos... Campinas: UNICAMP, 2021.

RIGHI, E. Metodologia para zoneamento de risco a inundações graduais. Tese de Doutorado em Geografia – Instituto de Geociências, Universidade Federal do Rio Grande do Sul. Porto Alegre. p. 176. 2016.

RITTER, J., BERENQUER, M., PARK, S., & SEMPERE-TORRES, D. (2021). Real-time assessment of flash flood impacts at pan-European scale: The ReAFFINE method. *Journal of Hydrology*, v. 603, p. 15, 127022, 2021.

ROCHA FILHO K. L. “Modelagem hidrológica da bacia do Rio Pirajuçara com TOPMODEL, telemetria e radar meteorológico.” 138 f. Dissertação (Mestrado em Meteorologia) – Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, 2010.

ROCHA FILHO, K. L.; CONDE, F.; ANDRIOLI, C. P.; BRITO, L. S. Previsão qualitativa para bacias hidrográficas com radar meteorológico. XIX SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, Maceió, AL, BRAZIL, ABRH, Associação Brasileira de Recursos Hídricos, 2011.

ROCHA FILHO, K. L.; CONDE, F.; ANDRIOLI, C. P. Correção em tempo real da precipitação estimada por um radar meteorológico com uma rede de superfície. XX SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, Bento Gonçalves, RS, BRAZIL, ABRH, Associação Brasileira de Recursos Hídricos, 2013.

ROCHA FILHO, K. L.; CONDE, F.; ANDRIOLI, C. P. Estimativas de Precipitação no Leste de São Paulo com Radar de Dupla Polarização. In: XXI Simposio Brasileiro de Recursos Hidricos. 2015.

RODRIGUEZ, J., VOS, F., BELOW, R., GUHA-SAPIR, D. Annual disaster statistical review 2008: The numbers and trends. Centre for Research on the Epidemiology of Disasters (CRED). 2009.

ROY, B., ISLAM, A. S., ISLAM, G. T., KHAN, M. J. U., BHATTACHARYA, B., ALI, M. H., HOSSAIN, S., SARKER, G. C.; PIEU, N. M. Frequency analysis of flash floods for establishing new danger levels for the rivers in the northeast haor region of Bangladesh. *Journal of Hydrologic Engineering*, 24(4), 05019004. 2019.

SAHARIA, M., KIRSTETTER, P. E., GOURLEY, J. J., HONG, Y., VERGARA, H. J. Mapping flash flood severity in the United States. *J. Hydrometeorol.* 18 (2), 397–411. 2015.

SÃO PAULO. MODELO DIGITAL DE ELEVAÇÃO (MDE) DO ESTADO DE SÃO PAULO OBTIDO A PARTIR DA BASE DO GISAT (CARTAS TOPOGRÁFICAS NA ESCALA 1:50.000). COORDENADORIA DE PLANEJAMENTO AMBIENTAL, SECRETARIA DO MEIO AMBIENTE DO ESTADO DE SÃO PAULO, 2013". 2013.

SANTOS, L. C. C. Estimativa de vazões máximas de projeto por métodos determinísticos e probabilísticos. Programa de Pós-Graduação em Engenharia Ambiental, Universidade Federal do Espírito Santo (UFES), Vitória (ES), p. 173, 2010.

SANTOS, M. A.; CARVALHO, S. M.; ANTONELI, V. Suscetibilidade de enchentes a partir da análise das variáveis morfométricas na bacia hidrográfica rio bonito em Irati-PR-brasil. **Revista Equador**, v. 5, n. 5, p. 152-167, 2016.

- SILVA, A. F. DA; TOLDO JUNIOR, E. E.; VAN ROOIJEN, A.; ABREU, C. F.; RODRIGUES FILHO, J. L.; ROCHA, R. DOS S.; AQUINO, R. DOS S. Inundação Costeira por Elevação do Nível do Mar em Imbé e Tramandaí - RS. *Revista Brasileira de Cartografia*, v. 72, n. 3, p. 541-557, 2020.
- 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.
- SOLOMATINE, D. P., XUE, Y. 2004. M5 model trees and neural networks: application flood forecasting in the upper reach the Huai River in China. *Journal of Hydrologic Engineering*, v. 9, n. 6, p. 491-501, 2004.
- SOLOMATINE, D. P. Data-driven modeling and computational intelligence methods in hydrology. In *Encyclopedia of Hydrological Sciences*, Anderson M (ed.). Wiley: New York, 2005.
- SOLOMATINE, D. P.; OSTFELD, A. Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, v. 10, n. 1, p. 3-22, 2008.
- SOMBRA, Y. B. da S; SILVA, T. P. Avaliação de suscetibilidade a inundações e a movimentos de massa no baixo curso da bacia de drenagem do Rio Muriaé (RJ) com base no geoprocessamento. *Revista da Pós-Graduação em Geografia da PUC-Rio*, v. 11, n. 21, p. 7-32, 2018.
- SUN, D.; ZHANG, D.; CHENG, X. Framework of national non-structural measures for flash flood disaster prevention in China. *Water*, v. 4, n. 1, p. 272-282, 2012.
- TANAKA, K., OMAR, H., TANAKA, S. Changes in extreme rainfall in arid and semi-arid region projected by super high resolution AGCM. In: *Fifth international conference on engineering geophysics (ICEG)*. Society of Exploration Geophysicists, Al Ain, UAE, pp 296–299. 2020.
- TANG, G., CLARK, M. P., PAPALEXIOU, S. M., MA, Z., & HONG, Y. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sensing of Environment*, p. 240, 111697. 2020.
- THIELEN-DEL POZO, J., THIEMIG, V., PAPPENBERGER, F., REVILLA-ROMERO, B., SALAMON, P., DE GROEVE, T., & HIRPA, F. The benefit of continental flood early warning systems to reduce the impact of flood disasters. *EUR Sci. Tech. Res. Rep.* 2015.
- TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. (orgs). *Desastres Naturais: conhecer para prevenir*. São Paulo, Instituto Geológico. 197 p. 2009.

- TOMINAGA, L.K. Inundações. In: TOMINAGA, L.K.; SANTORO, J.; AMARAL R. (orgs.). Desastres naturais: conhecer para prevenir. 3. ed. São Paulo: Instituto Geológico, p. 25-38. 2015.
- TORO, C.H.F., MEIRE, S.G., GALVEZ, J.F., FDEZ-RIVEROLA, F. A hybrid artificial intelligence model for river flow forecasting, *Applied Soft Computing*. v. 13, n. 8, p. 3449-3458, 2013.
- TRIGO, R. M., RAMOS, C., PEREIRA, S. S., RAMOS, A. M., ZÊRERE, J., LIBERATO, M.L.R. The deadliest storm of the 20th century striking Portugal: flood impacts and atmosphere circulation. *J. Hydrol.* 541, 597–610. 2016.
- TUCCI, C.E.M. Enchentes urbanas in: *Drenagem Urbana*, cap. 1 Editora da Universidade, ABRH, 1995.
- TUCCI, C.E.M., CLARKE, R.T. Impactos das mudanças da cobertura vegetal no escoamento. *Revista Brasileira de Recursos Hídricos*, v.2. Porto Alegre - RS. p135-152. 1997.
- TUCCI, C. E. M. *Drenagem Urbana e Controle de Inundações*. In: CAMPOS, Heraldo & CHASSOT, Attico (Org). *Ciências da Terra e meio ambiente*. São Leopoldo: Ed. Unisinos, 1999.
- TUCCI, C.E M.; BERTONI, J.C. (Orgs) *Inundações Urbanas na América do Sul* Ed. Brasileira de Recursos Hídricos, p. 471. 2003.
- UNISDR. *Hyogo Framework for Action 2005-2015: Building the Resilience of nations and Communities to Disasters*. Geneva: UNISDR. 2005.
- UNISDR. United Nations Office for Disaster Risk Reduction. *Developing Early Warning Systems: A Checklist*, Third International Conference on Early Warning, 27 – 29 March 2006, Bonn, Germany. 2006.
- UNISDR. United Nations Office for Disaster Risk Reduction. *Terminology on Disaster Risk*. Geneva: UNISDR. 2009.
- UNISDR. United Nations Office for Disaster Risk Reduction. *Proceedings of the third UN World Conference on Disaster Reduction*. Paper presented at the UN World Conference on Disaster Risk Reduction, 14-18 March 2015, Sendai - Japan. 2015.
- UNISDR. United Nations Office for Disaster Risk Reduction. *Terminology on disaster risk reduction*. Geneva: UNISDR. 2017.
- UNISDR. United Nations Office for Disaster Risk Reduction. *Annual Report 2017*. Geneva: UNISDR. 2018.

- UNISDR. “How to make cities more resilient a handbook for local government leaders”. 2016. Disponível em: <https://www.unisdr.org/files/26462_handbookfinalonlineversion.pdf>. Acesso em 05 de janeiro de 2022.
- VENKATESAN, E.; MAHINDRAKAR, A. B. Forecasting floods using extreme gradient boosting-a new approach. *International Journal of Civil Engineering and Technology*, v. 10, n. 2, p. 1336-1346, 2019.
- VILELA FILHO, L. R. Urbanização e fragilidade ambiental na Bacia do Corrego Proença, Município de Campinas (SP). 2006. 214p. Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Geociencias, Campinas, SP.
- VITOUSEK, S.; BARNARD, P. L.; FLETCHER, C. H.; FRAZER, N.; ERIKSON, L.; STORLAZZI, C. D. Doubling of coastal flooding frequency within decades due to sea - level rise. *Scientific Reports*, v. 7, p. 1399. 2017.
- WANG, W.P.H.A.J.M., VAN GELDER, J.K., VRIJLING, J. M. Forecasting daily streamflow using hybrid ANN models, *Journal of Hydrology*. v. 324, p. 383-399, 2006.
- YASEEN, Z.M., EL-SHAFIE, A., JAAFAR, O., AFAN, H.A., SAYL, K.N. Artificial intelligence-based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*. v. 530, p. 829–844, 2015.
- YASEEN, Z.M., ALLAWI, M.F., YOUSIF, A.A., JAAFAR, O., HAMZAH, F.M., EL-SHAFIE, A., 2018a. Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing and Applications*. v. 30, n. 5, p. 1479-1491, 2018.
- YASEEN, Z. M., EBTEHAJ, I., KIM, S., SANIKHANI, H., ASADI, H., GHAREB, M. I., ... & SHAHID, S. Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis. *Water*, 11(3), 502. 2019.