

## CORRELAÇÃO DE NÍVEIS D'ÁGUA EM BACIA URBANA COM REGISTROS DE INUNDAÇÃO NO MUNICÍPIO DE CAMPINAS/SP: PREPARAÇÃO DOS DADOS PARA APLICAÇÃO DE MACHINE LEARNING

Vinicius Araujo <sup>1</sup>; Paulo Henrique Barreto Ribeiro <sup>2</sup>; Ana Elisa Silva de Abreu <sup>3</sup>; Paula Dornhofer Paro Costa <sup>4</sup>; Filipe Antonio Marques Falcetta <sup>5</sup>; Omar Yazbek Bitar <sup>6</sup>,

**Resumo** – Diversas medidas de gerenciamento de inundações vêm sendo tomadas para mitigar os danos decorrentes destes eventos no mundo todo, dentre elas a geração de modelos hidrológicos baseados em dados para subsidiar sistemas de alerta antecipados. Esse artigo apresenta o processo de mineração de dados com aplicação de filtros em dados de níveis d'água a fim de selecionar aqueles que representem eventos de inundação para futura aplicação de técnicas de previsão de nível em *Machine Learning*. Para tal, dados sub-horários de nível d'água do ribeirão Proença, situado em Campinas, SP, referentes ao período de novembro de 2014 a junho de 2019 foram utilizados para compor a série histórica de dados do presente trabalho. Rotinas em *Python* foram criadas para selecionar, pré-processar, transformar e minerar tais dados. Com base na distribuição de frequência dos valores de nível d'água da série histórica, foram testados três limiares, que selecionaram 10, 6 e 3% do total de dias com dados disponíveis. A assertividade dos limiares na separação de datas com ocorrência de inundação foi avaliada com base em relatos de inundações em sites de notícias e cadastro da Defesa Civil. Com base nos resultados obtidos, concluiu-se que o limiar que separa 3% dos dias possui a melhor assertividade dentre os limiares testados, sendo assim escolhido para separação dos dados a serem utilizados no desenvolvimento do modelo em *Machine Learning* para previsão de inundações na bacia.

**Abstract** – Several flood management measures have been undertaken worldwide to mitigate the damage resulting from these events, including the generation of data-based hydrological models to support early warning systems. This article presents the data mining process with the application of filters on water level data in order to select data that represent flood events for future application of level prediction techniques in Machine Learning. To this end, sub-hourly water level data from the Proença river, located in Campinas, SP, for the period from November 2014 to June 2019 were used to form the historical series of data for this work. Python routines were created to select, pre-process, transform and mine such data. Based on the frequency distribution of the water level values of the historical series, three thresholds were tested, which selected 10, 6 and 3% of the total days with available data. The assertiveness of the thresholds in separating dates with occurrence of flooding was evaluated based on reports of floods on news sites and Civil Defense records. The conclusion is that the thresholds that separates 3% of the days has the best assertiveness among the tested thresholds, being thus chosen for the separation of the data to be used in the development of the model in Machine Learning for forecasting floods in the basin.

**Palavras-Chave** – Alerta antecipado; Modelagem baseada em dados; Inundações urbanas; Risco geológico

<sup>1</sup> Geól., MSc, Universidade Estadual de Campinas (UNICAMP), (16) 99397-8467, [vinicius.ari@hotmail.com](mailto:vinicius.ari@hotmail.com)

<sup>2</sup> Geól., BSc, Universidade Estadual de Campinas (UNICAMP), (19) 99195-655, [paulohbribeiro@outlook.com](mailto:paulohbribeiro@outlook.com)

<sup>3</sup> Geól., PhD, Universidade Estadual de Campinas (UNICAMP), (19) 3521-4575, [anaelisa@ige.unicamp.br](mailto:anaelisa@ige.unicamp.br)

<sup>4</sup> Eng., PhD, Universidade Estadual de Campinas (UNICAMP), (19) 98801-3938, [paulad@unicamp.br](mailto:paulad@unicamp.br)

<sup>5</sup> Eng., PhD, Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT), (11) 3767-4096, [falcetta@ipt.br](mailto:falcetta@ipt.br)

<sup>6</sup> Geól., PhD, Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT), [omar@ipt.br](mailto:omar@ipt.br)

## 1. INTRODUÇÃO

É notória a intensificação de ocorrências de desastres naturais nos últimos anos, principalmente os eventos associados a enchentes e inundações. 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 de eventos obtidos entre o período de 2000-2019.

Desta forma, é evidente a necessidade de medidas para gerenciar tais fenômenos. No contexto do gerenciamento de inundações, uma das principais tarefas é a realização de uma acurada modelagem hidrológica, associada a um bom levantamento geológico/geomorfológico, buscando assim caracterizar e compreender o comportamento do rio e sua respectiva bacia hidrográfica, e utilizar esta ferramenta para fornecer dados de entrada para diversas aplicações, dentre elas a criação de um sistema de alerta antecipado de inundações.

De modo geral, a metodologia atual empregada para modelagem hidrológica se subdivide em: modelos baseados em processos físicos e modelos orientados a dados (Adnan *et al.* 2019; Mehdizadeh *et al.* 2019). Enquanto o primeiro necessita de diversos parâmetros físicos da bacia e um robusto banco de dados hidroclimáticos (Ochoa-Tocachi *et al.* 2016; Adnan *et al.* 2019), cenário pouco comum para as bacias brasileiras, o segundo é baseado em modelos matemáticos construídos apenas a partir de observações de séries temporais (Mehdizadeh *et al.* 2019).

Nos últimos anos, com a sensível redução do custo das tecnologias de processamento computacional e o acesso ao grande volume de dados de sensores, os modelos orientados a dados, principalmente os gerados a partir de *Machine Learning*, tornaram-se comuns e eficientemente aplicáveis (Alizamir *et al.* 2018; Fotovatikhah *et al.* 2018; Yaseen *et al.* 2018; Yaseen *et al.* 2019). Segundo Yaseen *et al.* (2019), os algoritmos de *Machine Learning* possuem a capacidade de identificar as inter-relações complexas e intrínsecas entre os parâmetros que controlam, por exemplo, a transformação de chuva em vazão, não necessitando assim entender completamente os mecanismos de interação.

Tratando especificamente dos modelos orientados a dados, são cada vez mais numerosos os trabalhos que utilizam o nível d'água dos rios como variável de monitoramento (Gautam e Dulal, 2013; Adhikari *et al.* 2014, Smith *et al.* 2017). Deve-se destacar que a característica comum destes modelos é a definição de níveis d'água onde os eventos de inundação/enchentes estão próximos de ocorrerem, permitindo assim a estruturação de um sistema de alerta antecipado a inundações.

Dentro deste contexto, o presente trabalho tem como objetivo avaliar a eficácia de três limiares aplicados para filtragem dos dados de nível d'água para seleção de eventos de inundações em uma bacia urbana localizada no município de Campinas, São Paulo. Os dados selecionados pelo melhor limiar de filtragem serão futuramente utilizados como entrada para aplicações de algoritmos de *Machine Learning* para previsão de inundações, o que será objeto de trabalhos futuros.

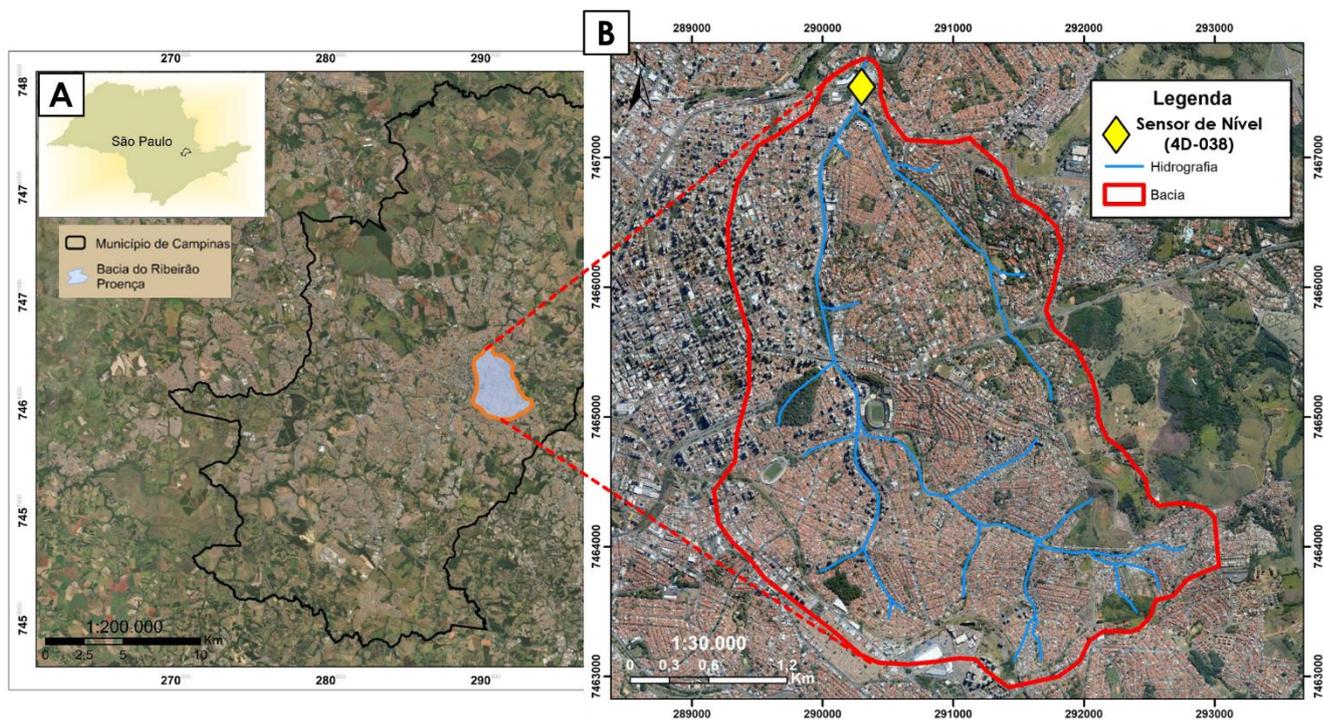
## 2. METODOS E TÉCNICAS

### 2.1. Área de Estudo

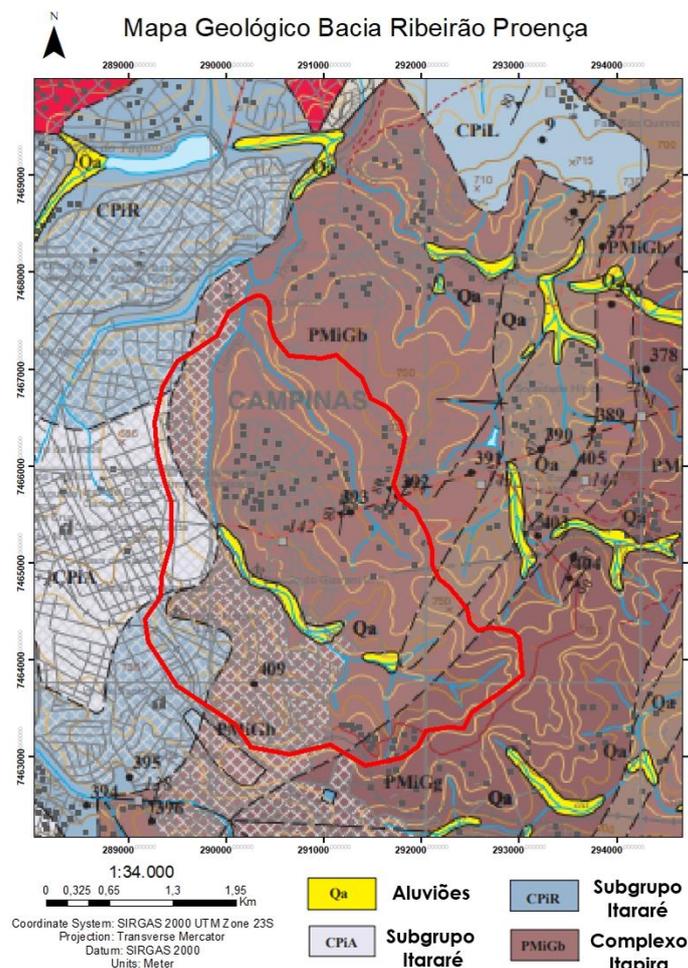
A bacia hidrográfica do Ribeirão Proença, área de estudo, está localizada na região central do município de Campinas, estado de São Paulo, contendo uma área de aproximadamente 12 km<sup>2</sup> e delimitada pelas coordenadas aproximadas: 7.462.900 a 7.467.700 mN e 289.100 a 293.000 mE (UTM, Zona 23) (Figura 1). De modo geral, possui uma densa ocupação urbana, intensa supressão das áreas verdes ciliares e diversos registros de inundações e enchentes.

Com relação a geologia, a área de estudo está inserida majoritariamente nas rochas gnáissicas bandadas do Complexo Itapira e em menores porções ocorrem rochas areníticas e ritmitos do Subgrupo Itararé e aluviões próximos aos canais de drenagem. A Figura 2 exibe o mapa geológico da bacia do ribeirão Proença. Deve-se destacar que grande parte das rochas

supracitadas não afloram, uma vez que a bacia hidrográfica em estudo se encontra em uma das áreas mais urbanizadas e antropizadas do município.

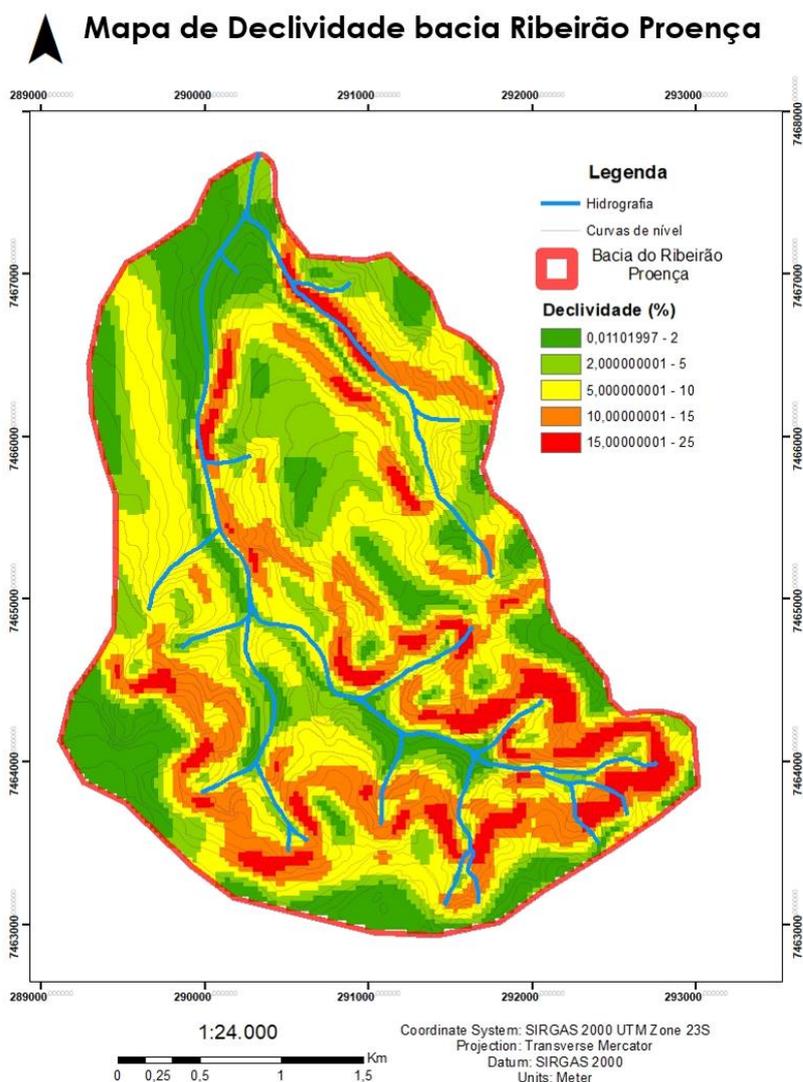


**Figura 1.** (A) Localização da bacia hidrográfica do ribeirão Proença (área de estudo) no município de Campinas (SP); (B) zoom na bacia hidrográfica.



**Figura 2.** Mapa geológico da bacia do ribeirão Proença (retirado e modificado de: Instituto Geológico, 2009).

A Figura 3 ilustra o mapa de declividade da área de estudo. Os subgrupos de declividade foram subdivididos em cinco grupos, conforme a tendência de ocorrência de inundações e enchentes proposta por Prina & Tretin (2014).



**Figura 3.** Mapa de declividade da bacia do ribeirão Proença.

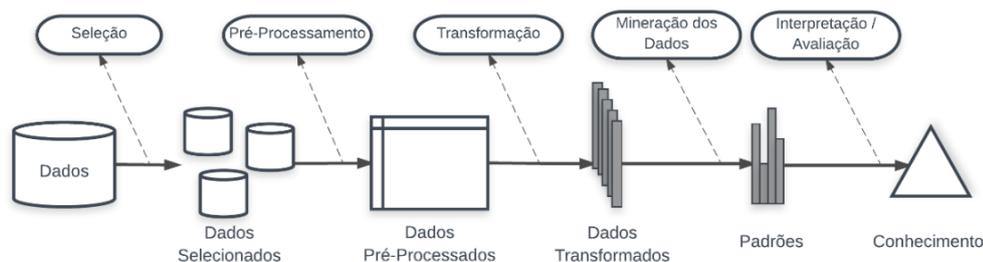
Segundo os autores, declividades entre 0 a 2% possuem uma propensão muito alta a ocorrência de inundações, seguida de alta para 2 a 5%, média 5 a 10%, baixa 10 a 15% e por fim muito baixa acima de 15%.

## 2.2. Banco de Dados de nível d'água no ribeirão Proença

Foram utilizados dados sub-horários coletados por um sensor de nível d'água (4D-038) localizado na foz do ribeirão Proença (Figura 1), extraídos do banco de dados hidrológicos do Departamento de Águas e Energia Elétrica (DAEE) os quais compreendem o período de novembro de 2014 até junho de 2019.

## 2.3. Análise Exploratória dos Dados

De posse dos dados, foi realizado o processo de seleção, pré-processamento, transformação e mineração dos dados, conforme proposto por Fayaad *et al.* (1996), apresentado na Figura 4. Algoritmos em linguagem *Python* foram desenvolvidos para cada uma destas etapas da pesquisa.



**Figura 4.** Fluxograma representando os principais componentes presentes no processamento da informação (adaptado de Fayaad *et al.* 1996).

Na etapa de seleção, foram selecionados os dados que compreendem o período de estudo (novembro de 2014 até junho de 2019), no qual há o menor número de falhas nos dados, enquanto que na etapa de pré-processamento, códigos em linguagem *Python* foram criados para unificar os níveis registrados no referido período e transformá-los em análises diárias, ao selecionar o maior valor observado no respectivo dia.

Após pré-processamento e transformação, iniciou-se a etapa de mineração dos dados. Nesta etapa, foram desenvolvidas rotinas em linguagem *Python* para avaliar os principais parâmetros estatísticos dos dados, a saber: número de dados (dias avaliados), média, desvio padrão, valor máximo e mínimo e quartis (25%, 50%, 75%, 90% e 95%).

Além disso, foi gerado o histograma contendo os níveis d'água diários máximos agrupados em doze intervalos obtidos através da fórmula de Sturges (Eq. 1), a fim de definir visualmente limiares de filtragem com o objetivo de selecionar os dias no qual há maior probabilidade de terem ocorrido eventos de inundação e verificar a hipótese de que esses estão associados aos dias em que o sensor de nível registrou valores incomuns e pouco recorrentes. A fórmula de Sturges é dada pela seguinte equação:

$$k = \log_2(N) + 1 \quad (\text{Eq.1})$$

Onde: N = número de eventos; k = número de intervalos.

#### 2.4. Correlação com relatos de inundações na bacia

As datas que resultaram da aplicação dos filtros na etapa anterior foram pesquisadas em notícias de jornais e no bancos de dados da Defesa Civil, a fim de buscar informações que corroboram a ocorrência dos eventos de inundação na bacia e vizinhanças. Tal procedimento, além de definir certamente os dias nos quais houve inundações registradas, auxiliou na avaliação da qualidade do limiar de filtragem aplicado frente à seleção dos possíveis eventos.

Nesta etapa, foram definidas três categorias para classificação dos dados de nível filtrados: 1) Datas onde há relatos de inundações/alagamentos na bacia do ribeirão Proença; 2) Datas onde há relatos de eventos, porém fora da bacia; 3) Datas sem relatos de quaisquer eventos.

Por fim, selecionou-se o melhor limiar entre os três empregados e organizou-se um banco de dados composto apenas pelas datas em que há relatos de inundações e fenômenos correlatos na bacia. Tais datas serão integradas com dados de precipitação interpretados a partir de radar meteorológico e formarão o banco de dados de entrada para processamento e treinamento de algoritmos de *Machine Learning* para previsão de inundações na área de estudo.

### 3. RESULTADOS E DISCUSSÃO

A Tabela 1 exibe os principais parâmetros estatísticos observados nos dados de nível após pré-processamento e transformação. Deve-se destacar que o registro corresponde ao nível máximo registrado em cada dia pelo sensor. Verifica-se que o número de dias avaliados é de 1378 dias,

embora o número de dias presentes no período de estudo seja de 1702. Desta forma, temos que a porcentagem de falhas no sensor é de 19%.

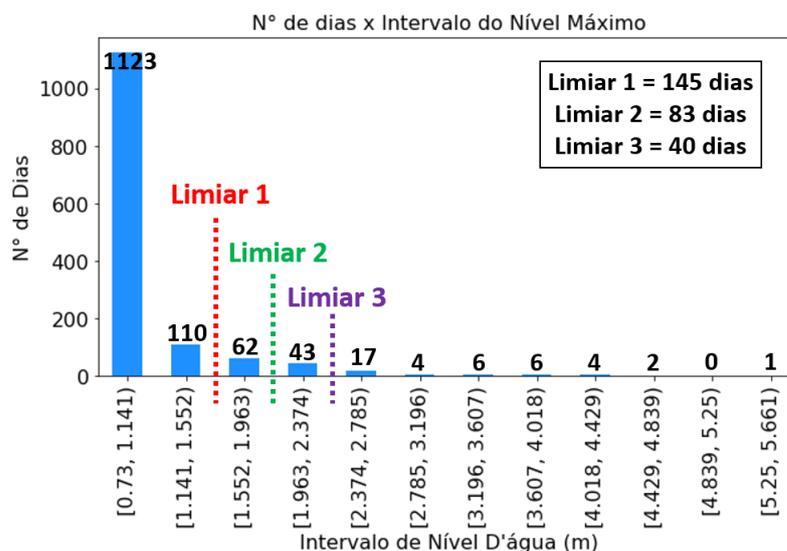
No que tange às demais estatísticas, a média dos níveis d'água diários máximos é de 1,040 metros, com desvio padrão de 0,511 metros e valores variando entre 0,737 a 5,258 metros. Ademais, a distribuição dos valores ocorre de forma heterogênea, uma vez que 90% dos dados são menores ou iguais a 1,587 metros. Tal característica era esperada, uma vez que a ocorrência de eventos onde os níveis d'água atingem valores incomuns são relativamente raros, até mesmo para bacias urbanas.

Buscando selecionar apenas os dias no quais os níveis registrados apresentaram valores incomuns e que estão possivelmente correlacionados a eventos de inundação, três limiares para filtragem de dados foram aplicados.

A Figura 5 exibe a posição dos limiares aplicados. Nesta figura, o número de intervalos utilizados (k) foi de 12, enquanto que o valor de separação empregado foi de 0,41 metros, calculados conforme a equação 1.

**Tabela 1.** Principais parâmetros estatísticos dos dados de nível d'água referentes ao período de estudo avaliado.

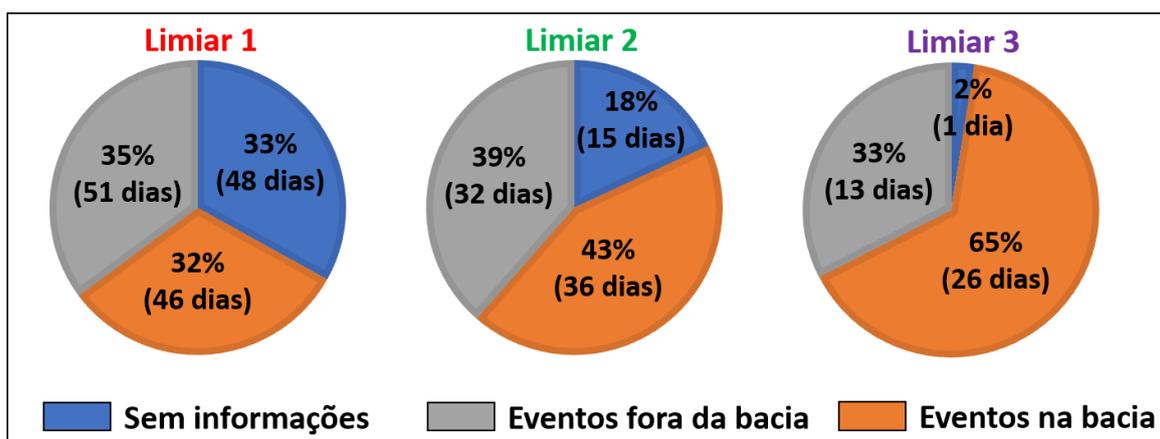
Parâmetro	Resultado
Nº de dias	1378
Média	1,040 m
Desvio Padrão	0,511 m
Valor Mínimo	0,737 m
Valor Máximo	5,258 m
25%	0,826 m
50%	0,840 m
75%	0,982 m
90%	1,587 m
95%	2,025 m



**Figura 5.** Histograma da distribuição dos níveis diários máximos em intervalos de nível d'água. As linhas pontilhadas destacam a posição de aplicação de cada limiar de filtragem. À direita da figura há o número de dias selecionados por cada limiar.

Com base na Figura 5, nota-se que o Limiar 1 (em vermelho) foi o mais abrangente (145 dias, ou 10% do total de dias analisados no período), uma vez que selecionou níveis diários máximos maiores que 1,552 metros. Por outro lado, o Limiar 3 (em roxo) foi o mais seletivo (40 dias, ou 3% do total de dias analisados no período), selecionando apenas níveis maiores que 2,374 metros. Por fim, o limiar intermediário em verde (Limiar 2), selecionou valores maiores que 1,963 metros, totalizando assim 82 dias. Deve-se destacar que os limiares de filtragem 1, 2 e 3 selecionaram os 10%, 6% e 3%, respectivamente, maiores valores observados, os quais equivaleriam aos Q10, Q6 e Q3.

Buscando avaliar a efetividade dos limiares na seleção dos possíveis eventos de inundação, cada um dos 145 dias selecionados foi classificado conforme a presença ou não de informações sobre a ocorrência dos eventos de inundação ou fenômenos correlatos registrados tanto na bacia do Ribeirão Proença, quanto no município de Campinas. A Figura 6 apresenta a distribuição dos dias em cada uma das classes pré-definidas e suas respectivas porcentagens.



**Figura 6.** Distribuição do número de dias e suas respectivas porcentagens em cada classe para cada um dos limiares de filtragem aplicados.

Com base na Figura 6, pode-se destacar que o número de dias sem registros de ocorrências de inundações e fenômenos correlatos diminuiu a medida que o nível d'água estabelecido no limiar aumentou, corroborando assim com o fato de que maiores valores tendem a estar associados a fenômenos mais intensos. Tal fato é nítido também nos eventos que ocorrem na bacia, local de instalação do sensor, ou seja, a porcentagem de ocorrência de eventos aumenta à medida que aumenta o nível filtrado. No que tange aos eventos fora da bacia, percebe-se que a sua porcentagem mantém-se constante, embora o número de dias diminua.

Tratando especificamente do Limiar 1 (1,552 metros), foram identificados 46 dias no qual há registros de ocorrências de eventos na bacia, embora tais eventos somem apenas 32% do total de dias filtrados. Por outro lado, 33% dos dias não apresentaram relatos de ocorrências e 35% apresentaram ocorrências fora da bacia. Desta forma, podemos destacar que a assertividade do Limiar 1 é de apenas 32%. Ademais, uma vez que a porcentagem nas três categorias foi similar, podemos constatar que o limiar 1 sequer cumpriu seu papel na filtragem de informações.

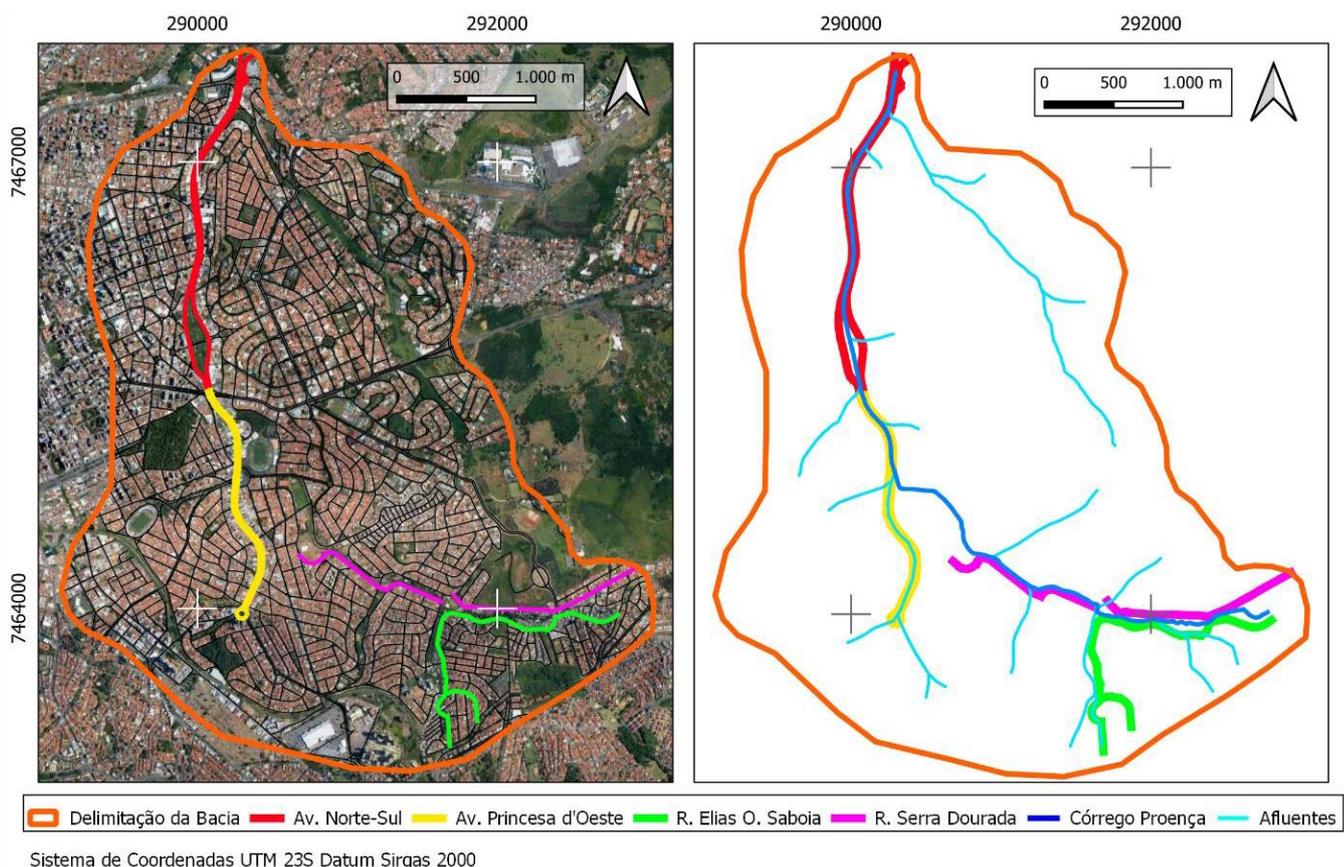
Já o Limiar 2 (1,963 metros), dos 83 dias selecionados, 36 apresentaram relatos de inundações, uma assertividade de 43%, aumento de 11% em relação ao Limiar 1. Com relação aos dias sem relatos, houve uma diminuição de 15% quando comparado com o Limiar 1, embora ainda assim haja muitos dias sem relatos (15 dias).

Por fim, o Limiar 3 (2,374 metros) selecionou apenas 40 dias e apresentou 26 dias contendo relatos de eventos na bacia, uma assertividade de 65%, aumento de 32% em relação ao Limiar 1. Além disso, nota-se que a porcentagem de dias sem relatos é de apenas 2%, evidenciando, assim, uma acurada seleção dos eventos de inundação na bacia e proximidades. Desta forma, o limiar 3 será adotado para seleção dos dados de entrada para o desenvolvimento do modelo baseado em dados para previsão de inundações na bacia nas etapas futuras da pesquisa.

Uma ressalva que se deve fazer em relação ao procedimento adotado para mineração dos dados de entrada para o *Machine Learning* é que, ao excluir datas do banco de dados, como acontece com as 62 datas situadas entre os limiares 1 e 2 ou com as 43 datas situadas entre os limiares 2 e 3 (vide Figura 3), estão sendo desconsiderados alguns dias nos quais há registros de inundações e fenômenos correlatos. Entretanto, ao avaliar os 20 dias de diferença onde há relatos de inundações (em laranja na Figura 6), constatou-se que grande parte das notícias ou registros da defesa civil relataram ocorrências isoladas de alagamentos em moradias, por exemplo, não correspondendo, portanto, a eventos em grandes escalas, tais como as inundações.

Por fim, ao analisar espacialmente os eventos relatados dentro da própria bacia de estudo, foi revelada uma distribuição desigual das informações presentes nos bancos de dados consultados para validação das ocorrências de inundações e correlação com a série histórica de níveis d'água.

De acordo com os dados veiculados nos sites de notícias, as avenidas Princesa d'Oeste e José de Souza Campos (Norte-Sul) são os locais da bacia onde a ocorrência de alagamentos e inundações são mais recorrentes. Por outro lado, as informações disponibilizadas pela Defesa Civil apontam que as ruas mais atingidas estão na região sudeste da bacia: ruas Serra Dourada e Elias de Oliveira Saboia. O mapa da Figura 7 traz a localização dessas vias mais atingidas e sugere que os eventos registrados em notícias de jornal concentram-se nas áreas de maior poder aquisitivo da bacia, enquanto os eventos registrados no banco de dados da Defesa Civil concentram-se nas áreas da população de menor poder aquisitivo, o que evidencia a importância de se utilizar dados de registro de ocorrências de inundações de fontes diversas para que haja representatividade.



**Figura 7.** Localização das vias mais atingidas por eventos de inundações e fenômenos correlatos na bacia hidrográfica do ribeirão Proença, Campinas/SP.

#### 4. CONCLUSÕES

As análises realizadas com base nos dados de nível d'água coletados na foz do Ribeirão Proença, Campinas, permitiram conhecer diversas características sobre a variabilidade desse

parâmetro na área de estudo, embora haja a presença de aproximadamente 19% de falha nos dados no período de estudo selecionado.

Avaliando os níveis d'água máximos diários, constatou-se que a média é de 1,040 metros, sendo que em 90% dos dias analisados o nível d'água foi menor ou igual a 1,587 metros, denotando assim uma não homogeneidade nos dados. Tal característica era esperada, uma vez que ocorrência de eventos onde os níveis d'água atingem valores elevados são relativamente raros, até mesmo para bacias urbanas.

Buscando selecionar os eventos pouco recorrentes e eventos com maior potencial de estarem associados a inundações, três limiares de filtragem correspondentes aos valores de 1,552, 1,963 e 2,374 metros foram aplicados, valores maiores que 10, 6 e 3% de todos os dados de nível da série histórica, respectivamente. Da aplicação dos limiares 1, 2 e 3, apenas 145, 83 e 40 dias foram selecionados, respectivamente, para verificação da ocorrência de inundações e fenômenos correlatos. Comparando-se tais datas com os registros dos bancos de dados da Defesa Civil de Campinas e com notícias de jornal, o número de dias com inundações registradas foi de 46, 36 e 26 dias, respectivamente, correspondendo assim a uma assertividade de 32, 43 e 65%, respectivamente, para os limiares 1, 2 e 3. Ademais, embora haja uma diferença de 20 dias entre o número de inundações e eventos correlatos entre os limiares 1 e 3, os registros apontam que grande parte dos eventos corresponderam a alagamentos, fenômenos reduzidos espacialmente.

Desta forma, pode-se concluir que dentre os três limiares analisados, o Limiar 3, que seleciona apenas 3% do total de dias da série histórica analisada, apesar de possuir menor número de registros de inundações, foi que o que apresentou melhor assertividade e consequentemente pode ser o melhor limiar de filtragem empregado para classificação das datas com ocorrências de inundações na bacia do Ribeirão Proença na etapa de mineração dos dados para aplicação de Machine Learning.

## AGRADECIMENTOS

Os autores agradecem a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, processos: 2017/50343-2 e 2020/00058-2), o Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT) e a Universidade Estadual de Campinas (UNICAMP), pelo fomento à pesquisa e por todo suporte técnico/intelectual oferecido.

Agradecemos também a equipe técnica do Departamento de Águas e Energia Elétrica (DAEE) do Estado de São Paulo pela disponibilização dos dados de nível d'água.

## REFERÊNCIAS

- ADHIKARI, T., RAKHAL, B., MAHARJAN, L. D., & KUSHI, P. S. (2014). Determination of flood warning and danger of Mohana and Macheli river. *Final Report*, Mercy Corps Nepal, p. 157.
- ADNAN, R. M., LIANG, Z., HEDDAM, S., ZOUNEMAT-KERMANI, M., KISI, O., LI, B. (2019). 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.
- ALIZAMIR, M., KISI, O., ZOUNEMAT-KERMANI, M. (2018). Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. *Hydrological Sciences Journal*. v. 63, n. 1, p. 63–73.
- 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.
- FAYYAD, U., PIATETSKY-SHAPIRO, G., and SMYTH, P. (1996). Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>.

- FOTOVATIKHAH, F., HERRERA, M., SHAMSHIRBAND, S., ARDABILI, S.F., PIRAN, J. (2018). 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.
- GAUTAM, D. K., & DULAL, K. (2013). Determination of threshold runoff for flood early warning in Nepalese Rivers. *IDRiM Journal*, v. 3, n.1, p. 126-136.
- 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.
- MEHDIZADEH, S., FATHIAN, F., SAFARI, M. J. S., ADAMOWSKI, J. F. (2019). 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.
- OCHOA-TOCACHI, B.F., BUYTAERT, W., DE BIÈVRE, B. (2016). Regionalization of land-use impacts on streamflow using a network of paired catchments. *Water Resour. Res.* v. 52, n. 9, p. 6710–6729.
- 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.
- SMITH, P. J.; BROWN, S.; DUGAR, S. (2017). Community-based early warning systems for flood risk mitigation in Nepal. *Natural Hazards and Earth System Sciences*, v. 17, n. 3, p. 423-437.
- 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., BONAKDARI, H.; AL-ANSARI, N. & SHAHID, S. (2019). Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis. *Water*, v. 11, n. 3, p. 502.