

## XXVI SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

### ACURÁCIA TEMPORAL EM BANCOS DE DADOS DE DESLIZAMENTOS: POTENCIAL DO USO DE NOTÍCIAS ONLINE

Clara Moreira Cardoso<sup>1</sup>; Gean Paulo Michel<sup>1,2</sup>; Artur Nonato Vieira Cereto<sup>1</sup>; Danúbia Teixeira<sup>3</sup>; Franciele Zanandrea<sup>1</sup>

**Palavras-Chave** – Desastres, Movimentos de Massa, GDELT.

#### INTRODUÇÃO

A tomada de decisões relacionadas à gestão de riscos depende diretamente de bancos de dados históricos, que podem ser elaborados a partir de uma variedade de abordagens, como pelo emprego de notícias e informações na internet (Battistini *et al.*, 2017). Assim, o objetivo deste trabalho foi avaliar a confiabilidade de um banco de dados construído exclusivamente a partir de notícias coletadas automaticamente.

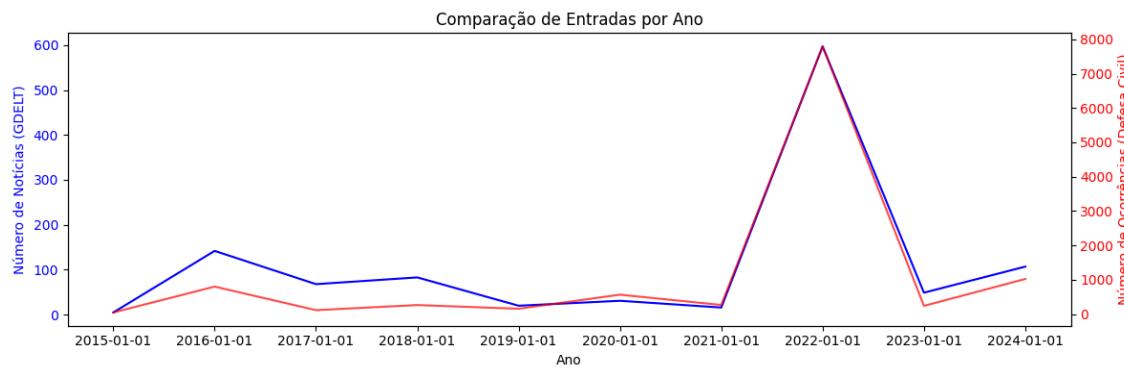
#### MATERIAIS E MÉTODOS

Petrópolis (RJ), área de estudo desta pesquisa, é vulnerável a deslizamentos devido à intensa incidência de chuvas e à ocupação desordenada em áreas de risco (Alcântara *et al.*, 2023). Assim, foram coletadas notícias referentes a deslizamentos em Petrópolis a partir do banco de dados global GDELT e comparadas com os registros da Defesa Civil local. A análise foi feita a partir de gráficos e dos coeficientes de determinação ( $R^2$ ) e de correlação de Spearman ( $p$ ), considerando escalas anual, mensal, semanal e diária. Buscando maior precisão temporal, foram identificadas nas notícias palavras-chave relacionadas a períodos do dia (“manhã”, “tarde”, “noite” e “madrugada”). Para os 10 maiores eventos, ocorridos de 2015 a 2024, analisaram-se casos em que os horários mencionados nas notícias divergiam dos registros da Defesa Civil.

#### RESULTADOS E DISCUSSÃO

As Figuras 1 e 2 apresentam a comparação de distribuições anuais e diárias obtidas através de um banco de dados de 1118 notícias de deslizamentos em Petrópolis.

Figura 1 - Comparação de entradas anuais entre dados coletados por notícias e pela Defesa civil.

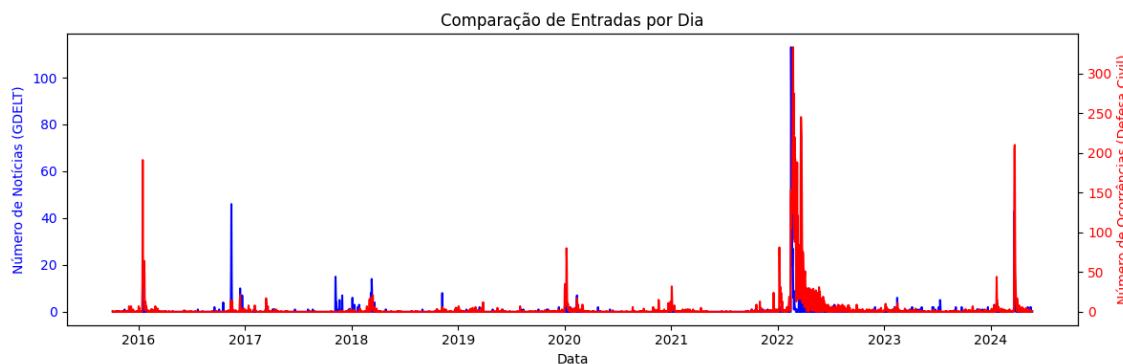


<sup>1</sup>) Universidade Federal Fluminense - UFF, [claromoreiracardoso@id.uff.br](mailto:claromoreiracardoso@id.uff.br), [geanpmichel@gmail.com](mailto:geanpmichel@gmail.com), [arturcereto@id.uff.br](mailto:arturcereto@id.uff.br), [francielez@id.uff.br](mailto:francielez@id.uff.br)

<sup>2</sup>) Instituto de Pesquisas Hidráulicas - IPH, UFRGS, [geanpmichel@gmail.com](mailto:geanpmichel@gmail.com)

<sup>3</sup>) Universidade Federal do Rio de Janeiro - COPPE/UFRJ, [danubia.teixeira@ifal.edu.br](mailto:danubia.teixeira@ifal.edu.br)

Figura 2 - Comparação de entradas diárias entre dados coletados por notícias e os coletados pela Defesa Civil.



Observa-se uma grande correlação entre os dados anuais, enquanto nota-se um maior desencontro entre os dados diários. Essa interpretação foi confirmada pelos métodos estatísticos de regressão linear ( $R^2 = 0,97$  para os dados anuais e  $R^2 = 0,21$  para os dados diários) e de relações não-paramétricas (coeficiente de Spearman igual a 0,71 para os dados anuais e igual a 0,25 para os dados diários). Ou seja, quanto maior a abrangência temporal, maior a correlação entre os dados, enquanto abrangências temporais menores resultam em correlações mais fracas. Um dos fatores que pode influenciar esse resultado é a imprecisão nas datas registradas em ambos os bancos de dados analisados. Na identificação de menções a períodos do dia, verificou-se que, dentre as datas dos 10 maiores eventos de deslizamentos, 4 apresentaram divergência entre os dados da Defesa Civil e os dados coletados nas notícias (16/01/2016, 03/01/2021, 22/03/2022 e 23/03/2024). Após a análise, somente em duas datas foi possível identificar o momento real de deflagração (16/01/2016 e 03/01/2021), e as palavras-chave mencionadas corroboraram com os momentos. Nas demais datas, as notícias referiam-se a consequências de chuvas que ocorreram em dias anteriores.

## CONCLUSÃO

Apesar de suas limitações, o método se apresenta como uma solução viável, ágil e de baixo custo para a construção automatizada de bancos de dados de deslizamentos, especialmente para análises com foco em escalas temporais de maior abrangência. A comparação mostrou que o banco de dados da Defesa Civil tende a ser mais abrangente e menos específico. O emprego de notícias, por outro lado, resulta em um maior destaque a eventos de impacto significativo à população.

## AGRADECIMENTOS

Os autores agradecem a bolsa de estudos concedida para a primeira autora fornecida pelo Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico no projeto Desenvolvimento do Sistema de Alerta Antecipado para Deslizamentos – SALAD (processo CNPq 446162/2023-0). Agradecimentos são estendidos à Secretaria Municipal de Proteção e Defesa Civil (SEMPDEC) de Petrópolis pelo fornecimento de dados.

## REFERÊNCIAS

- ALCÂNTARA, E.; MARENKO, J. A.; MANTOVANI, J.; LONDE, L. R.; LAU YU SAN, R.; PARK, E.; LIN, Y. N.; WANG, J.; MENDES, T.; CUNHA, A. P.; PAMPUCH, L.; SELUCHI, M.; SIMÕES, S.; CUARTAS, L. A.; GONCALVES, D.; MASSI, K.; ALVALÁ, R.; MORAES, O.; SOUZA FILHO, C. S.; MENDES, R.; NOBRE, C. (2023). “*Deadly disasters in southeastern South America: flash floods and landslides of February 2022 in Petrópolis, Rio de Janeiro*”. Nat. Hazards Earth Syst. Sci., 23, pp. 1157–1175.

- BATTISTINI, A.; ROSI, A.; SEGONI, S.; LAGOMARSINO, D.; CATANI, F.; CASAGLI, N. (2017). “Validation of landslide hazard models using a semantic engine on online news”. Applied Geography 82, pp. 59 - 65.