



REGISTROS DE DESLIZAMENTOS A PARTIR DE NOTÍCIAS NA WEB: ESTUDO DE CASO DO MUNICÍPIO DE PETRÓPOLIS

Clara Moreira Cardoso ¹
Gean Paulo Michel ²
Artur Nonato Vieira Cereto ³
Franciele Zanandrea ⁴
Danúbia Teixeira ⁵

RESUMO

Nos últimos anos, as atividades antrópicas e as mudanças climáticas contribuíram para um aumento na frequência e magnitude de desastres. Entre eles, destacam-se os deslizamentos de terra, que, por conta também de uma ocupação desordenada de áreas de risco em diversas regiões do Brasil, tornaram-se cada vez mais recorrentes e destrutivos. É imprescindível, portanto, aprimorar a tecnologia voltada à gestão de risco de deslizamentos, que depende diretamente de inventários, catálogos e bancos de dados de deslizamentos prévios. Seguindo esta linha, este trabalho objetiva verificar a aplicabilidade de metodologia baseada na recuperação de dados a partir de notícias de jornais, disponibilizadas online, com foco no município de Petrópolis/RJ, reconhecido pela grande suscetibilidade a deslizamentos. Para tanto, o método escolhido faz uso do Projeto GDELT, que procura monitorar notícias do mundo todo, codificando fatores como locais, datas, fontes e temas citados, disponibilizando-os gratuitamente. Com isso, foi possível coletar notícias brasileiras que mencionam “deslizamento” e “Petrópolis” entre os anos de 2015 e 2024, organizando-as em um banco de dados. Esse resultado foi então comparado com o banco de dados elaborado pela Defesa Civil de Petrópolis, relacionando os dados anuais, mensais, semanais e diários de cada fonte. Observou-se, respectivamente, uma correlação excelente ($R^2 = 0,97$); boa ($R^2 = 0,57$) e baixa ($R^2 = 0,28$ nos semanais e $R^2 = 0,21$ nos diários). Com isso, considera-se que os resultados foram promissores, principalmente em relação a períodos de tempo mais abrangentes. Contudo, vale ressaltar que os registros da Defesa Civil são referentes aos acionamentos por parte da população, não representando com exatidão o momento, localização ou número de ocorrências, o que pode prejudicar a precisão da comparação. Por fim, conclui-se que a elaboração de bancos de dados pode enriquecer com uma estratégia que englobe diferentes métodos, sendo o estudo de notícias um importante componente.

Palavras-chave: Deslizamento, Banco de Dados, GDELT, Petrópolis, Notícias.

¹ Graduando do Curso de Engenharia de Recursos Hídricos e do Meio Ambiente da Universidade Federal Fluminense - UFF, claramoreiracardoso@id.uff.br;

² Doutor pelo Curso de Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul - UFRGS, geanpmichel@gmail.com;

³ Mestrando do Curso de Engenharia de Biossistemas da Universidade Federal Fluminense - UFF, arturcereto@id.uff.br;

⁴ Doutor pelo Curso de Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul - UFRGS, francielez@id.uff.br;

⁵ Doutoranda pelo Curso de Engenharia Civil da Universidade Federal do Rio de Janeiro - COPPE/UFRJ, danubia.teixeira@ifal.edu.br;

INTRODUÇÃO

Observou-se, nos últimos anos, um aumento generalizado na frequência e gravidade de desastres geomorfológicos, impulsionado principalmente por atividades antrópicas, que afetam cada vez mais o funcionamento dos processos geomorfológicos (CENDRERO *et al.*, 2020). Além disso, outro importante fator influenciador são as mudanças climáticas, que, segundo o IPCC (2019), podem exacerbar os processos de degradação do solo, por meio de fenômenos como: aumentos na intensidade das chuvas, inundações, vento, aumento do nível do mar e ação de ondas, entre outros.

Além da maior frequência e magnitude de desastres, consequentes das ações antrópicas e das mudanças climáticas, é importante destacar o aumento da vulnerabilidade por parte da população. Como ressaltado por OZTURK *et al.* (2022), a aceleração da urbanização, especialmente em países de baixa e média-baixa renda em regiões tropicais, resultou em um aumento de habitações posicionadas em áreas de alto risco, elevando as chances de serem vítimas de eventos como deslizamentos.

A cidade de Petrópolis (RJ), área de estudo do presente projeto, é situada na região serrana e é caracterizada pela presença de encostas íngremes, rochas falhadas e fraturadas e clima tropical com pluviosidade significativa ao longo do ano (GUERRA *et al.*, 2007). Além disso, sua história é marcada por uma urbanização acelerada e desordenada, onde a alta taxa de crescimento populacional resultou em uma ocupação substancial de áreas de risco elevado (ASSIS DIAS *et al.*, 2018). Como consequência das condições geológicas, climáticas e sociais listadas, Petrópolis é um dos municípios mais vulneráveis a deslizamentos no Brasil.

Por isso, é imprescindível o aprofundamento de pesquisas voltadas à prevenção, mitigação e redução de riscos de desastres. Como destacado por TAYLOR *et al.* (2015), a tomada de decisões na gestão de riscos depende diretamente da qualidade das avaliações de riscos. Portanto, os bancos de dados de deslizamentos são de extrema importância, especialmente para documentação da extensão dos fenômenos de deslizamento; avaliação da suscetibilidade, do perigo e do risco de deslizamentos de terra; investigação da distribuição, tipos e padrões de deslizamentos em relação a características geomorfológicas; e estudo da evolução de paisagens dominadas por processos de movimentos de massa (GUZZETTI *et al.*, 2012).

Apesar de sua imprescindibilidade, a elaboração de bancos de dados de deslizamentos, no geral, ainda é marcada por inúmeras lacunas, onde uma grande



quantidade de bancos de dados de deslizamentos nacionais ao redor do mundo é marcada por falta de informações, de padronização e de disponibilização digital dos dados (VAN DEN EECKHAUT e HERVÁS, 2012).

Com isso, destaca-se a importância de pesquisas voltadas ao desenvolvimento de metodologias de elaboração de bancos de dados de deslizamentos. BATTISTINI *et al.* (2017) destacam três abordagens importantes: mapeamento de deslizamentos históricos a partir de sensoriamento remoto; uso de imagens aéreas e de satélites em conjunto com levantamentos de campo; compilação de informações coletadas a partir de notícias, artigos publicados, fotografias e outros. O presente trabalho tem como objetivo principal explorar a terceira abordagem, com foco no uso de notícias.

O emprego de notícias para elaboração de bancos de dados de deslizamentos não é uma estratégia recente, sendo sua primeira aplicação publicada na literatura a de GUZZETTI *et al.* em 1994. Com o tempo, inúmeras estratégias já foram exploradas por pesquisadores diversos, incluindo o estudo da possibilidade de coleta automática de notícias (BATTISTINI *et al.*, 2017). Seguindo esta linha, o presente trabalho também visou executar uma coleta de notícias automatizada, porém utilizando a ferramenta GDELT (*Global Database of Events, Language and Tone*), um projeto que se apresenta como um banco de dados global de monitoramento de notícias do mundo inteiro, com acesso gratuito.

A partir disso, foi possível coletar 1118 notícias, lançadas dentro do período de 2015 a 2024, que mencionam o tema de deslizamentos e a cidade de Petrópolis. A fim de verificar a completude e confiabilidade dessa metodologia, os registros coletados foram comparados ao banco de dados de deslizamentos elaborado e disponibilizado pela Defesa Civil de Petrópolis. A comparação foi feita de maneira visual, a partir da plotagem de gráficos, assim como de maneira quantitativa, a partir do cálculo de coeficientes de determinação.

Para averiguar a precisão temporal do banco de dados desenvolvido, a comparação foi feita considerando quatro diferentes escalas de dados: anuais, mensais, semanais e diários. Os resultados mostraram que a correlação entre os dados anuais foi excelente ($R^2 = 0,97$); entre os dados mensais foi boa ($R^2 = 0,57$); e entre os dados semanais, assim como entre os dados diários, foi fraca ($R^2 = 0,28$ e $R^2 = 0,21$, respectivamente). Com isso, concluiu-se que o banco de dados baseado em notícias demonstrou um resultado promissor, especialmente para os dados anuais e mensais.



Assim, a metodologia se apresenta como uma importante ferramenta para complementação de dados de deslizamentos.

MATERIAIS E MÉTODOS

O GDELT, segundo seu próprio *website*, é um projeto, criado por Kalev Leetaru e a Universidade de Georgetown e apoiado pelo *Google Jigsaw*, que procura monitorar notícias transmitidas, impressas e da web do mundo todo, com atualização contínua a cada 15 minutos. Além disso, o projeto identifica e categoriza, a partir de técnicas de *machine learning* e análise de texto, elementos diversos mencionados nas reportagens, como pessoas, locais, datas, temas, fontes, organizações, emoções, contagens, citações, imagens e eventos, criando uma plataforma aberta e gratuita para computação.

Com isso, o presente trabalho visa coletar notícias que tenham relação com o tema de deslizamentos, a fim de verificar seu potencial para enriquecimento de bancos de dados de deslizamentos. A vantagem deste método é a possibilidade de acessar uma quantidade considerável de notícias podendo filtrá-las de acordo com o tema, localização e data, viabilizando uma busca específica e refinada.

O GDELT armazena e disponibiliza as notícias coletadas através de bancos de dados variados, dos quais o *Global Knowledge Graph* (GKG) foi considerado mais apropriado para o objetivo do presente artigo, por indicar pessoas, organizações, localizações, contagens, temas, fontes e eventos mencionados nas notícias.

A extração de dados neste arquivo foi feita através do *Google Cloud*, plataforma que serve como ponte para inúmeros bancos de dados. Utilizando o *Python*, foi consultado o *BigQuery*, um armazém de dados que fornece acesso ilimitado aos bancos de dados do GDELT. Foram aplicados filtros para selecionar somente notícias que mencionam a localização “Petrópolis”; envolvem o tema de “deslizamento” e foram lançadas no período de tempo entre 2015 e 2024. Além disso, foi aplicada uma condição onde as notícias devem citar o nome “Petrópolis” no mínimo mais de uma vez, em uma tentativa de evitar a inclusão de notícias que anunciam deslizamentos em outros locais, mas referenciam o nome “Petrópolis” em passagem.

É importante ressaltar que o Projeto GDELT traduz para o inglês todas as notícias monitoradas pelo projeto, a fim de facilitar o processamento e a pesquisa de notícias do mundo inteiro. Por conta disso, foram empregadas palavras-chave em inglês, sendo elas “*landslide*” e “*mudslide*”. Esses termos foram escolhidos a partir do trabalho



de TAYLOR et al., 2015, que apresenta as palavras mais comumente usadas na língua inglesa para este conceito.

Com isso, foram coletadas 3212 notícias do mundo inteiro. Para refinar este resultado e reduzir a quantidade de notícias que anunciam o mesmo evento, optou-se por manter somente aquelas cuja fonte é brasileira, considerando pouco provável que um deslizamento ocorrido em Petrópolis fosse anunciado por uma fonte estrangeira e não por uma fonte nacional. Assim, o banco de dados final contém 1118 notícias.

RESULTADOS E DISCUSSÃO

A fim de verificar a confiabilidade dos resultados, foi empregado como elemento comparativo o banco de dados elaborado pela Defesa Civil de Petrópolis, que registra todas as vezes em que foi acionada, incluindo o motivo, a hora e a localização do pedido de socorro. Vale ressaltar, no entanto, que as informações registradas são referentes aos acionamentos por parte da população, não representando com exatidão o momento, localização ou número de ocorrências.

A comparação entre os bancos de dados foi feita, em primeiro lugar, de maneira visual, a partir de gráficos representando os dados coletados tanto a partir do GDELT quanto pela Defesa Civil de Petrópolis. A análise foi feita de quatro maneiras: empregando-se dados anuais (Figura 1), dados mensais (Figura 2), dados semanais (Figura 3) e dados diários (Figura 4).

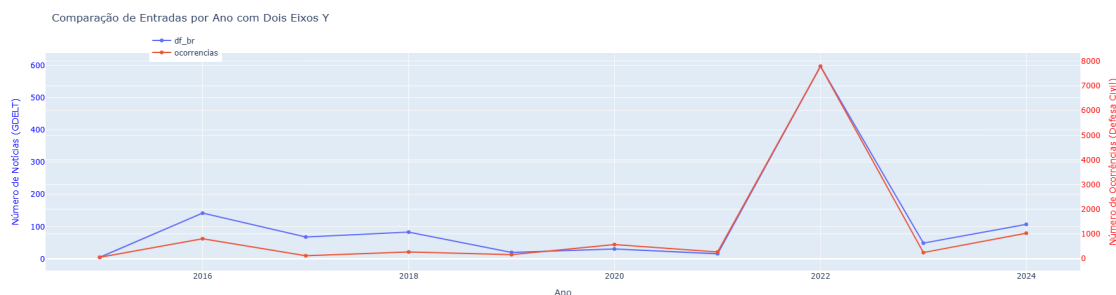


Figura 1 - Gráfico de comparação dos dados anuais coletados através do GDELT (azul) com os registrados pela Defesa Civil de Petrópolis (vermelho).

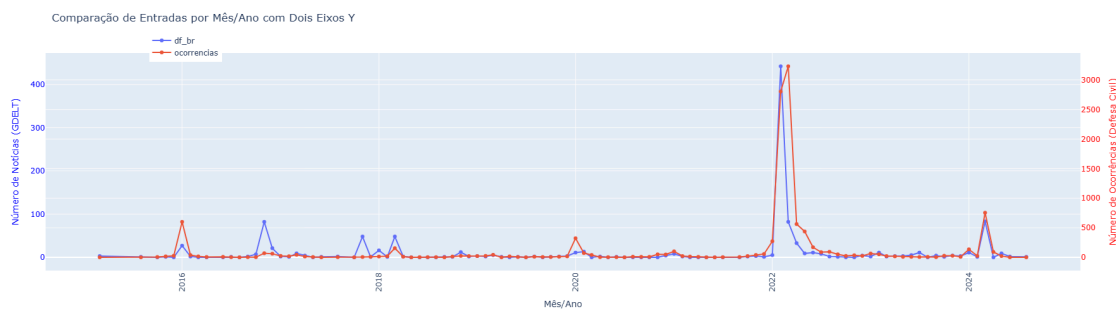


Figura 2 - Gráfico de comparação dos dados mensais coletados através do GDELT (azul) com os registrados pela Defesa Civil de Petrópolis (vermelho).

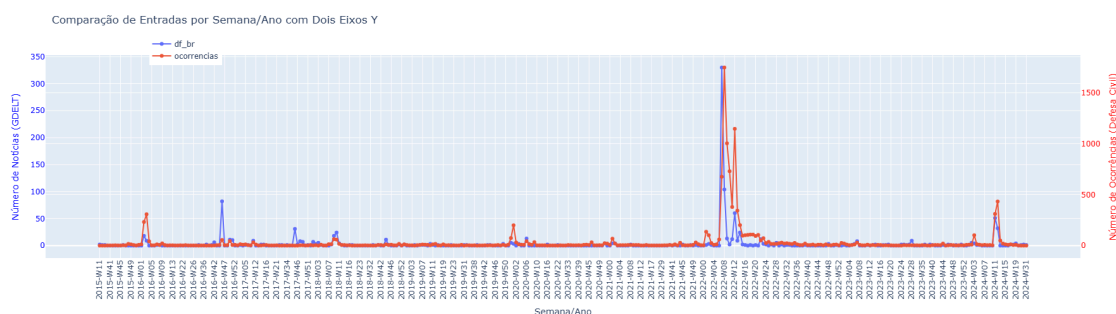


Figura 3 - Gráfico de comparação dos dados semanais coletados através do GDELT (azul) com os registrados pela Defesa Civil de Petrópolis (vermelho).

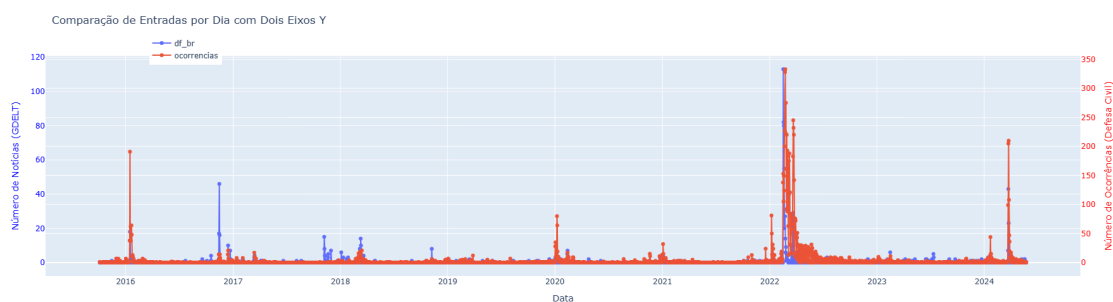


Figura 4 - Gráfico de comparação dos dados diários coletados através do GDELT (azul) com os registrados pela Defesa Civil de Petrópolis (vermelho).

Além disso, para verificar a correlação entre os bancos de dados, foram elaborados gráficos de dispersão, contendo suas respectivas linhas de tendência, para cada tipo de dados: anuais (Figura 5), mensais (Figura 6), semanais (Figura 7) e diários (Figura 8). Vale ressaltar que os gráficos foram montados aplicando uma escala logarítmica em ambos os eixos, a fim de diminuir a dispersão dos dados.

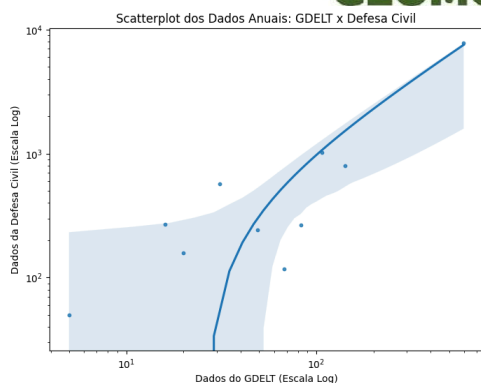


Figura 5 - Gráfico de dispersão comparando os dados anuais da Defesa Civil com as notícias anuais coletadas a partir do GDELT.

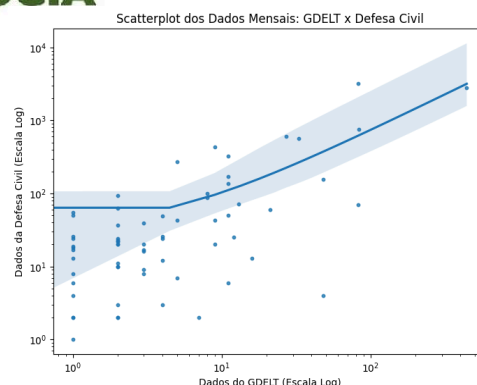


Figura 6 - Gráfico de dispersão comparando os dados mensais da Defesa Civil com as notícias mensais coletadas a partir do GDELT.

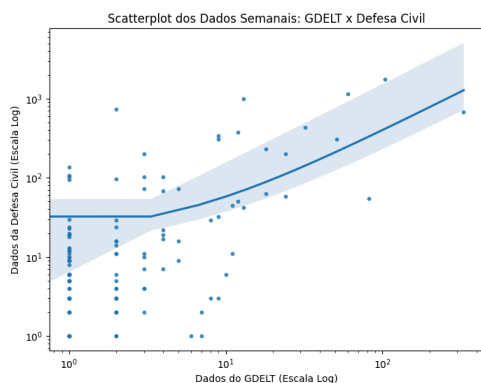


Figura 7 - Gráfico de dispersão comparando os dados semanais da Defesa Civil com as notícias semanais coletadas a partir do GDELT.

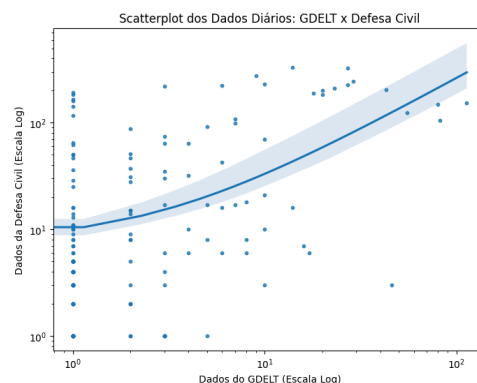


Figura 8 - Gráfico de dispersão comparando os dados diários da Defesa Civil com as notícias diárias coletadas a partir do GDELT.

Analisando os gráficos, é possível notar uma similaridade no padrão de variação de número de registros entre os dois bancos de dados, especialmente nos gráficos referentes a dados anuais e mensais. Visualmente, observa-se uma tendência de maior semelhança entre os bancos de dados para maiores abrangências temporais. Por outro lado, quanto maior a precisão temporal, menor a correlação entre os bancos de dados.

Essa suposição foi confirmada pela comparação quantitativa, realizada a partir da análise linear dos resultados, calculando os coeficientes de determinação (R^2) para cada categoria. Os resultados apontaram uma correlação excelente entre os dados anuais



($R^2 = 0,97$); boa entre os dados mensais ($R^2 = 0,57$); e precária entre os dados semanais ($R^2 = 0,28$) e entre os dados diários ($R^2 = 0,21$).

Observando os gráficos de dispersão, nota-se que cada valor no eixo das notícias está associado a uma grande variedade de valores de registros pela Defesa civil, e vice-versa. Com isso, é impossibilitada a análise generalizada do comportamento entre os dados. Tal afirmação fica mais evidente ao notar que existem dias sem notícias, mas com registros da Defesa Civil, assim como existem dias com lançamento de notícias, mas sem registros por parte da Defesa Civil.

CONSIDERAÇÕES FINAIS

Visto que os resultados obtidos empregando-se dados anuais e dados mensais foram particularmente próximos dos padrões dos registros de ocorrência feitos pela Defesa Civil de Petrópolis, é possível concluir que os resultados foram promissores. Mais especificamente, destaca-se os dados mensais, que podem ser úteis para perceber padrões, identificando o mês ou a época do ano mais propícios a terem deslizamentos.

Os dados semanais e diários, por sua vez, se distanciaram mais dos registros da Defesa Civil, o que pode ser consequência de uma série de fatores. Um exemplo é o fato do banco de dados montado a partir de notícias coletadas pelo GDELT não registrar, necessariamente, a data específica do deslizamento, e sim a data em que a notícia foi publicada. Outra inconsistência observada nos dados, tanto nas notícias quanto nos registros da Defesa Civil, é a variância de registros de acordo com os dias da semana. Em ambos os casos, o dia da semana que menos houve registros foi domingo, representando uma falha sistemática de registros. Isso pode ser um sinal de que nos fins de semana costuma haver uma queda de registro na Defesa Civil, seja por conta de uma quantidade menor de indivíduos acionando a Defesa Civil ou seja por uma diminuição de funcionários ou de eficiência de trabalho nos fins de semana.

Outra importante ressalva é que notícias podem anunciar mais de um deslizamento de uma só vez, portanto uma entrada pode representar múltiplos deslizamentos. Ao mesmo tempo, por outro lado, a Defesa Civil pode fazer mais de um registro referente ao mesmo deslizamento, visto que ela pode ser acionada múltiplas vezes para atender ao mesmo incidente. Além disso, um fator influente são deslizamentos que não atingem, direta ou indiretamente, a população, visto que estes geralmente não são noticiados, mas podem resultar em acionamento da Defesa Civil.



Uma proposta para pesquisas futuras é a investigação de soluções automáticas para identificar se as notícias separadas realmente anunciam uma nova ocorrência. Foi observado que uma parcela das notícias coletadas referem-se a deslizamentos que ocorreram previamente, e que foram tão marcantes que são mencionados novamente por conta de suas repercussões e danos causados. Além disso, foi observado que em épocas de eleição há um aumento no número de notícias que anunciam planos de obras voltadas à gestão de riscos de deslizamentos. Esse tipo de notícia pode ser lançada em qualquer momento do ano, trazendo ruídos dentre os registros de deslizamentos.

Outra ação que pode ser desenvolvida no futuro é aprofundar a precisão horária. Seria interessante, por exemplo, explorar a possibilidade de identificar, dentro do corpo das notícias, indicações da hora em que houve a ocorrência, seja a partir de horários mencionados no texto ou então a partir de palavras-chave, como “manhã”, “tarde”, “noite” ou “madrugada”, frequentemente empregadas no corpo de notícias.

Por conta dos obstáculos destacados, é importante ressaltar que, para a elaboração de inventários completos, detalhados e robustos, é de grande importância o emprego também de outras fontes e métodos. Algumas possibilidades seriam a inclusão de uma investigação de redes sociais, que têm potencial para trazer uma precisão temporal mais precisa; de informações obtidas a partir de trabalhos de campo, imagens de satélite e sensoriamento remoto, que podem auxiliar a precisão espacial; e de estudos de outros tipos de registros escritos, como artigos ou jornais físicos, a fim de possibilitar a classificação dos tipos de deslizamentos.

Além disso, é de grande importância que haja uma atualização na forma em que se registra os deslizamentos por parte de entidades governamentais, para que se tenha uma atualização constante e precisa do momento de deflagração do deslizamento. Por fim, como ressaltado por VAN DEN EECKHAUT e HERVÁS (2012), é essencial que todos os dados coletados sejam concentrados em softwares estruturados para coletar e armazenar informações, disponibilizando-as para o público.

Apesar das falhas e incertezas listadas na presente seção, de forma geral, esta metodologia se apresenta como uma opção rápida, barata e viável para identificar deslizamentos, visto que o emprego de notícias coletadas pelo GDELT possui uma boa precisão mensal. Com isso, destaca-se o potencial deste método para aprimorar os bancos de dados existentes, assim como para auxiliar outras localidades, especialmente aquelas com menor preparo para gestão de riscos.



AGRADECIMENTOS

Os autores agradecem a bolsa de estudos concedida para o primeiro autor deste trabalho 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 dos registros de ocorrências relacionadas a deslizamentos de terra.

REFERÊNCIAS

- ASSIS DIAS, Mariane Carvalho De *et al.* Estimation of exposed population to landslides and floods risk areas in Brazil, on an intra-urban scale. **International Journal of Disaster Risk Reduction**, v. 31, p. 449–459, out. 2018.
- BATTISTINI, Alessandro *et al.* Validation of landslide hazard models using a semantic engine on online news. **Applied Geography**, v. 82, p. 59–65, maio 2017.
- CENDRERO, Antonio *et al.* Anthropocene Geomorphic Change. Climate or Human Activities? **Earth's Future**, v. 8, n. 5, maio 2020.
- GUERRA, Antonio. Catastrophic events in Petrópolis city (Rio de Janeiro state), between 1940 and 1990. **GeoJournal**, v. 37, n. 3, p. 349–354, nov. 1995.
- GUZZETTI, Fausto *et al.* Landslide inventory maps: New tools for an old problem. **Earth-Science Reviews**, v. 112, n. 1–2, p. 42–66, abr. 2012.
- GUZZETTI, Fausto; CARDINALI, Mauro; REICHENBACH, Paola. The AVI project: A bibliographical and archive inventory of landslides and floods in Italy. **Environmental Management**, v. 18, n. 4, p. 623–633, jul. 1994.
- OZTURK, Ugur *et al.* How climate change and unplanned urban sprawl bring more landslides. **Nature**, v. 608, n. 7922, p. 262–265, 11 ago. 2022.
- TAYLOR, Faith E. *et al.* Enriching Great Britain's National Landslide Database by searching newspaper archives. **Geomorphology**, v. 249, p. 52–68, nov. 2015.
- VAN DEN EECKHAUT, Miet; HERVÁS, Javier. State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk. **Geomorphology**, v. 139–140, p. 545–558, fev. 2012.
- SHUKLA, P.R. *et al.* IPCC, 2019: Summary for Policymakers. In: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. <https://doi.org/10.1017/9781009157988.001>