1º HIDROGEOMINAS
SIMPÓSIO MINEIRO DE HIDROGEOMORFOLOGIA E RECURSOS HÍDRICOS
07 – 10 de setembro de 2026
Departamento de Geografia – UFMG
Belo Horizonte - MG

CARACTERIZAÇÃO DOS DESLIZAMENTOS NO SUDESTE DO BRASIL VIA APRENDIZADO DE MÁQUINA

Vittor Hugo Santos Courbassier [1]
Michelle Simões Reboita [2]

RESUMO

Os deslizamentos de terra figuram entre os desastres naturais mais letais do Brasil, com a Região Sudeste concentrando o maior número de ocorrências e fatalidades do país. Este trabalho investiga a dinâmica espaço-temporal entre precipitação e deslizamentos no Sudeste brasileiro entre 1981 e 2025, a partir de um *pipeline* de ciência de dados aplicado a 1082 registros de deslizamentos (movimento de massa) do Sistema Integrado de Informações sobre Desastres. As séries temporais de precipitação foram extraídas do produto *CHIRPS* v2.0 via *Google Earth Engine* pela metodologia *Ponto-a-Pixel*, que associa cada evento registrado ao *pixel CHIRPS* de localização mais próxima. Com base no limiar de evento extremo de precipitação obtido para todo o Sudeste de 47,2 mm, as análises revelaram que a maioria dos casos de deslizamento corresponde a rupturas por saturação no solo antecedente e não por chuva intensa pontual. Algoritmos de aprendizado de máquina não supervisionado identificaram três tipologias de deslizamento e nove *hotspots* críticos, nos quais a letalidade é 12,4 vezes maior que em eventos isolados. O principal resultado do estudo é que, nos eventos de maior impacto, a população exposta é 4,7 vezes maior do que em eventos comuns, enquanto a chuva acumulada é apenas 1,2 vezes superior. Os resultados indicam que o fator de maior peso na ocorrência de eventos letais não é a intensidade pluviométrica, mas a concentração de população em áreas de encosta.

Palavras-chave: Precipitação; Movimentos de massa; CHIRPS; Desastres naturais; Inteligência artificial.

INTRODUÇÃO

Os deslizamentos estão entre os desastres naturais mais recorrentes e letais do Brasil, e a Região Sudeste consolida-se como o epicentro dessas tragédias. Essa concentração é condicionada pela geomorfologia abrupta da Serra do Mar e da Serra da Mantiqueira, que, associada a décadas de urbanização desordenada sobre encostas, transformou o perigo geológico em risco social estrutural (FERNANDES et al., 2004; RAMOS et al., 2022).

A precipitação é o principal agente deflagrador. O colapso ocorre quando a infiltração eleva a poropressão além do limite de resistência ao cisalhamento do solo, sendo o ponto crítico a partir do

[1] Graduando em Ciências Atmosféricas, Universidade Federal de Itajubá (UNIFEI), Instituto de Recursos Naturais (IRN), Itajubá, MG, Brasil. vittorcourbassier@gmail.com;
[2] Professora Doutora, Universidade Federal de Itajubá (UNIFEI), Instituto de Recursos Naturais (IRN), Itajubá, MG, Brasil. reboita@unifei.edu.br.

qual a encosta perde estabilidade e sofre ruptura (FERNANDES et al., 2004). Esse processo manifesta-se tanto por gatilhos imediatos de chuva intensa quanto pela saturação progressiva promovida pela persistência da Zona de Convergência do Atlântico Sul (ZCAS) ao longo de dias durante a estação chuvosa no Sudeste do país. Esse segundo mecanismo é particularmente crítico porque induz rupturas generalizadas mesmo sob intensidades pluviométricas não excepcionais (ESCOBAR, 2019).

Nesse contexto, o objetivo deste trabalho é investigar a relação espaço-temporal entre precipitação e deslizamentos na Região Sudeste do Brasil entre 1981 e 2025, identificando limiares pluviométricos críticos, classificando mecanismos de gatilho predominantes e quantificando o papel da exposição demográfica na letalidade dos eventos. A principal motivação do estudo é fornecer subsídios quantitativos para o aprimoramento de sistemas de alerta precoce e para a elaboração de políticas públicas de gestão de riscos geológicos mais eficazes na região.

MATERIAIS E MÉTODOS

O estudo fundamenta-se em um *pipeline* de ciência de dados estruturado sobre 1082 registros de deslizamentos de terra do S2iD (https://s2id.mi.gov.br/paginas/index.xhtml), cobrindo o período de 1981 a 2025 na Região Sudeste do Brasil. A partir desse banco de dados, foram extraídas informações como data da ocorrência do evento, coordenadas geográficas (latitude e longitude), município e Unidade da Federação (UF). Essa base do S2iD também forneceu os dados de impacto humano e demográficos, que incluem a quantidade de óbitos, desabrigados, desalojados e o tamanho populacional dos municípios afetados. Os dados de deslizamento foram auditados seguindo protocolos de consistência geográfica aplicados em bancos nacionais de desastres (ALVALÁ et al., 2019).

Os dados de precipitação foram obtidos do *CHIRPS* v2.0, produto híbrido que combina dados de satélite no canal infravermelho termal com calibração por estações meteorológicas *in situ* (FUNK et al., 2015). Esse conjunto de dados possui resolução horizontal de 0,05° (~5,5 km) e frequência diária, o que é adequado para estudos de regiões com forte gradiente orográfico e baixa densidade instrumental. Os dados do CHIRPS são adquiridos via *API* do *Google Earth Engine* seguindo a abordagem Ponto-a-Pixel, que consiste em extrair o valor do pixel *CHIRPS* coincidente com a localização mais próxima do evento registrado no S2iD. Essa metodologia armazena a chuva de cada evento em uma janela de 11 dias, com o dia do deslizamento (D_0) e os 10 dias antecedentes (D_{-1} *a* D_{-10}), semelhante a análise de ESCOBAR (2019).

Para definir o limiar de evento extremo de precipitação, reuniram-se todos os valores diários de chuva (excluindo valores iguais a zero) das janelas de 11 dias do evento (D_0 *a* D_{-10}) dos 1082 registros, formando uma distribuição única de dados. Sobre essa distribuição, aplicou-se o percentil 90 (P_{90}), obtendo-se o limiar crítico de 47,2 mm/dia. Definiu-se também o limiar de evento extremo de precipitação acumulada pelo somatório de D_0 *a* D_{-10} correspondente ao P_{90} dessa distribuição, resultando em 226,2 mm. Essa abordagem de análise de precipitação antecedente é semelhante à utilizada por Escobar (2019) para eventos de ZCAS.

A clusterização *K-Means* foi parametrizada para a extração de três grupos ($k = 3$), validado pelo critério de cotovelo. Para mitigar a disparidade de escalas entre as variáveis pluviométricas (mm), demográficas (habitantes) e de impacto (óbitos), aplicou-se uma transformação logarítmica do tipo $ln(1 + x)$ sobre os atributos, seguido de uma padronização estatística pelo método *StandardScaler*, reescalando todas as distribuições para média zero e variância unitária. O algoritmo *DBSCAN* com métrica *Haversine* foi parametrizado com raio de vizinhança (ϵ) de 30 km e calibrado pela curva de k-distância de densidade mínima de 15 pontos para mapear *hotspots* de alta concentração espacial de deslizamentos (REICHSTEIN et al., 2019; ALCÂNTARA et al., 2025). As relações entre as variáveis meteorológicas e demográficas foram exploradas por meio de uma rede de correlação de *Spearman*, na qual os nós representam as variáveis e as arestas os coeficientes de correlação significativos. A validação estatística da influência da exposição demográfica na letalidade dos desastres foi realizada através do teste não-paramétrico de *Mann-Whitney U* com nível de significância de 5% ($\alpha = 0,05$).

Uma vez realizada a clusterização, os resultados são organizados em tipologias de risco (*K-Means*) e em aglomerados espaciais críticos (*DBSCAN*), permitindo a caracterização regional dos padrões de deslizamento e a identificação de hotspots prioritários para gestão de riscos.

RESULTADOS E DISCUSSÃO

A Figura 1 apresenta a distribuição espacial dos 1082 registros de deslizamentos na Região Sudeste entre 1981 e 2025. A maior ocorrência desses episódios é encontrada nos eixos metropolitanos e nas escarpas da Serra do Mar e da Serra da Mantiqueira. Também destaca-se que São Paulo apresenta a maior média anual de ocorrências, seguido por Minas Gerais.

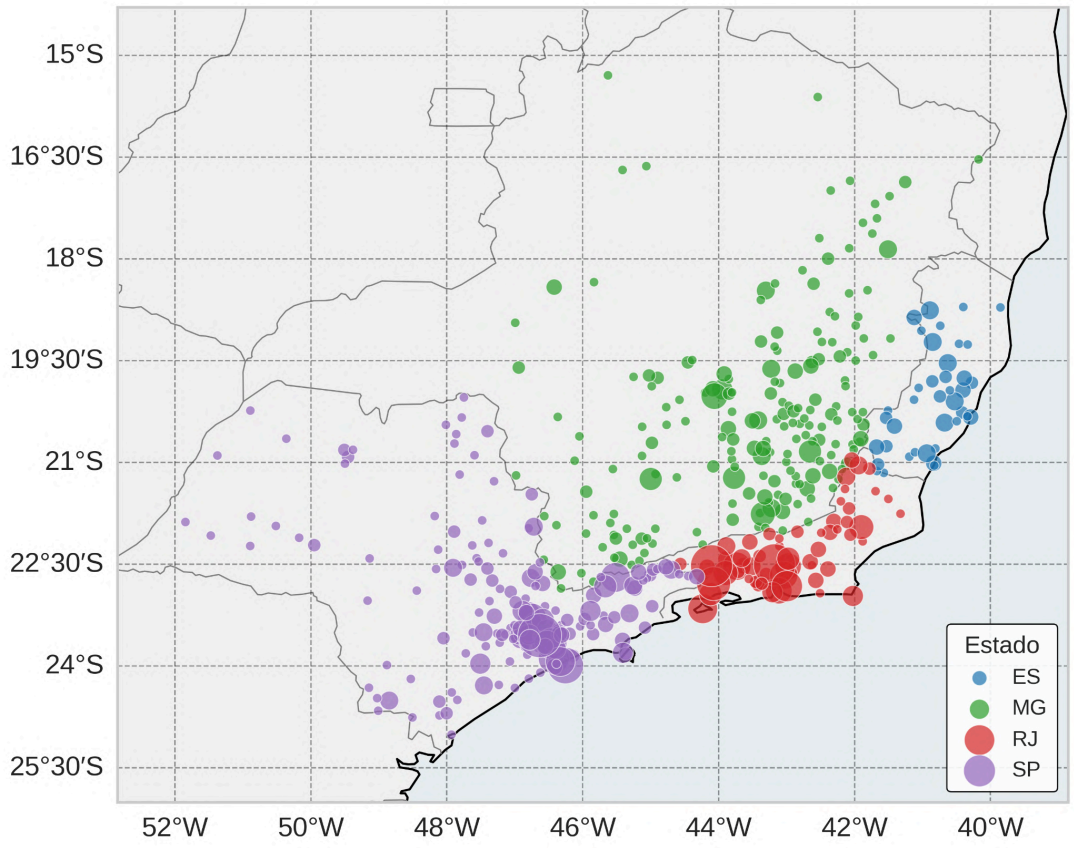


Figura 1 — Distribuição espacial de deslizamentos na Região Sudeste (1981–2025). O diâmetro dos círculos é proporcional à frequência de eventos por município; as cores distinguem as Unidades da Federação: São Paulo (roxo), Minas Gerais (verde), Rio de Janeiro (vermelho) e Espírito Santo (azul).

A análise das séries temporais entre 2000 e 2025 mostra que meses com maior número de deslizamentos (Figura 2a) são também os com maior número de óbitos (Figura 2b), registrando um coeficiente de correlação de *Spearman* de $\rho = 0,581$ ($p < 0,001$). Há uma indicação de que os acumulados de precipitação de D_{-1} a D_{-10} são maiores (Figura 2c), com correlação direta entre a precipitação acumulada de D_{-1} a D_{-10} e a frequência de desastres, com valores de $\rho = 0,942$ ($p < 0,001$). Apenas 3,2% dos meses analisados concentram 21,6% de toda a mortalidade histórica, totalizando 473 vítimas fatais. Esse comportamento ocorreu em janeiro de 2007 e 2010, que registraram 40 eventos cada, com picos de 121 e 61 óbitos respectivamente, evidenciando que o risco está concentrado em episódios de magnitude extrema combinada a vulnerabilidades locais (FERNANDES et al., 2004).

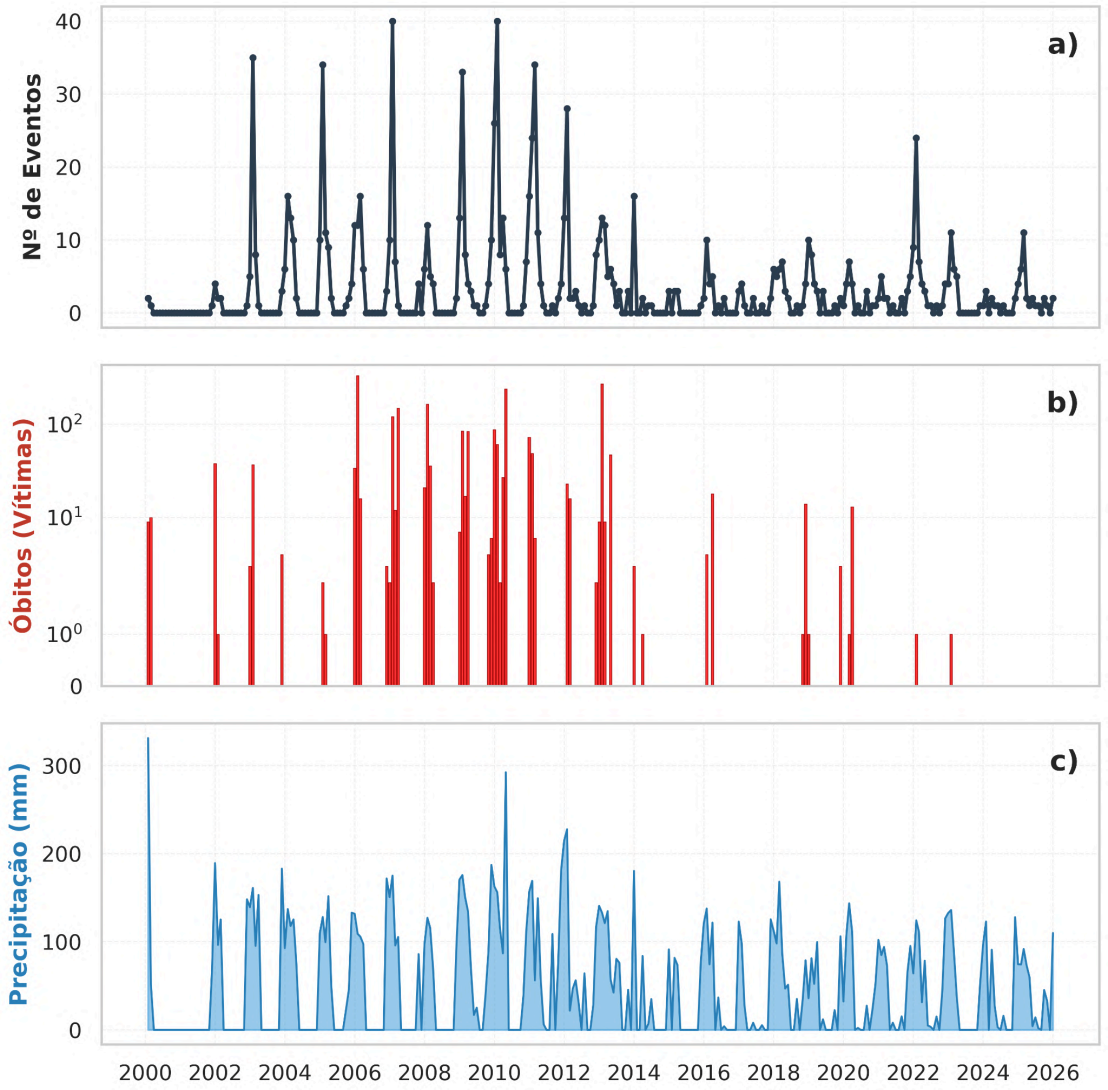


Figura 2 — Características dos desastres no Sudeste (2000–2025). O painel (A) apresenta a frequência mensal de movimentos de massa; o painel (B) detalha a letalidade absoluta mensal em escala logarítmica; o painel (C) exibe o acumulado de precipitação de D_{-1} *a* D_{-10}.

A Figura 3 apresenta a sazonalidade e a tendência interanual da frequência de deslizamentos. A distribuição climatológica (Figura 3a,c) mostra o pico no trimestre de verão (71% dos casos), com janeiro respondendo a cerca de 35% das ocorrências nessa estação. Isso pode ser reflexo direto da atuação da ZCAS sobre a saturação do solo (ESCOBAR, 2019). A série histórica de frequência anual de deslizamentos apresenta tendência negativa com variação inferior a um evento por ano (Figura 3b). Essa tendência não é estatisticamente significativa ao nível de 5%.

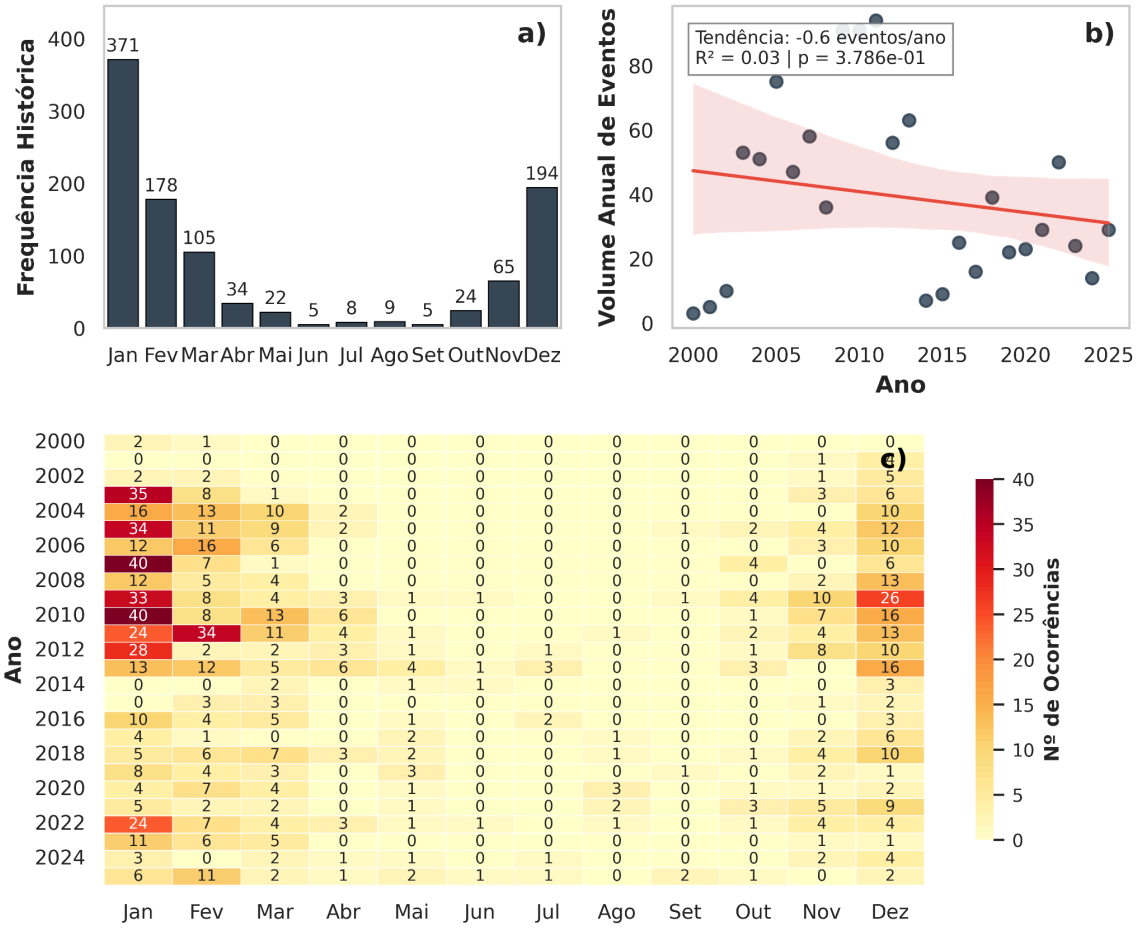


Figura 3 — Sazonalidade e assinatura interanual no Sudeste (2000–2025). (A) Distribuição da frequência histórica por mês; (B) Regressão linear de tendência interanual, sem tendência estatisticamente significante ($p = 0,38$); (C) *Heatmap* de densidade temporal com o volume mensal de ocorrências por ano.

A Figura 4 apresenta as tipologias de deslizamento identificadas neste estudo através da clusterização K-Means. O algoritmo foi aplicado sobre as variáveis normalizadas de precipitação em D_0, D_{-1} *a* D_{-10}, população exposta e letalidade, extraindo três tipologias distintas. A Tipologia 1, denominada Vulnerabilidade Urbana, caracteriza-se por baixos acumulados pluviométricos de D_0 *a* D_{-10}, evidenciando que a exposição demográfica supera a forçante meteorológica como fator de risco. A Tipologia 2, Saturação Moderada, é a mais frequente com 480 eventos, combinando chuva diária moderada em D_0 com maiores valores de precipitação de D_{-1} *a* D_{-10}. A Tipologia 3, Saturação Crítica, caracteriza-se por altos acumulados de D_{-1} *a* D_{-10} com dependência nula da

chuva em D_0, dinâmica típica de regiões interioranas como o sul de Minas Gerais (ESCOBAR, 2019).

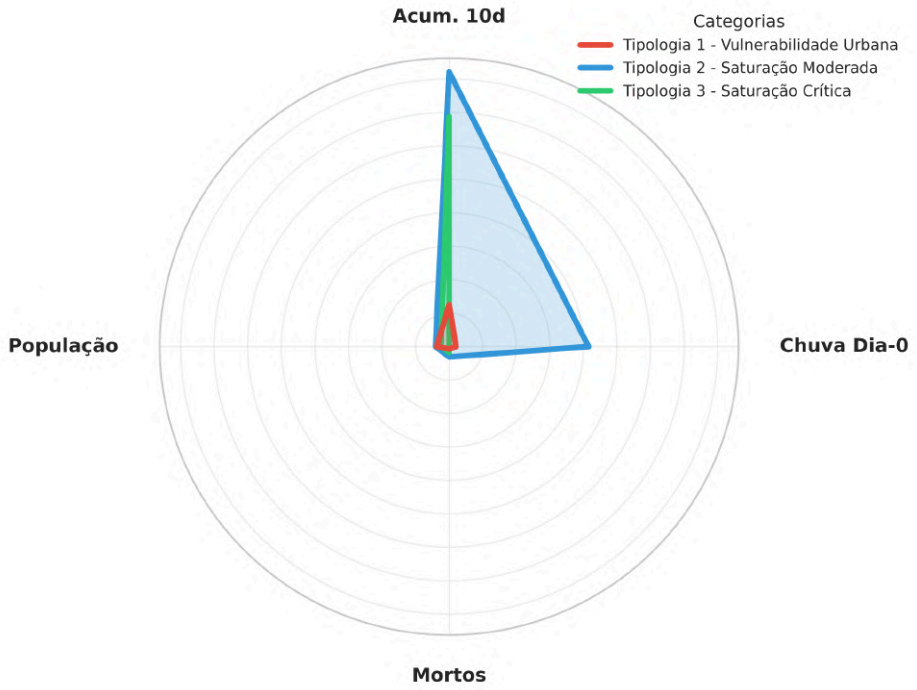


Figura 4 — Assinatura morfológica das tipologias de deslizamento via *K-Means*. O diagrama de radar exibe os pesos das variáveis normalizadas para a Tipologia 1 (vermelho), Tipologia 2 (azul) e Tipologia 3 (verde). Os eixos descrevem a precipitação em D_0, o acumulado antecedente em D_{-1} *a* D_{-10}, a população exposta e a letalidade.

A Figura 5 apresenta os hotspots críticos identificados pelo algoritmo *DBSCAN*. A análise identificou 9 aglomerados espaciais de deslizamentos que concentram 57,6% dos eventos (592 registros) e 90,1% de toda a letalidade histórica, revelando um fator de agravo de 12,4 vezes na taxa de mortalidade em relação a ocorrências isoladas. Esses hotspots correlacionam-se com aglomerados urbanos em encostas de alta declividade (FERNANDES et al., 2004; RAMOS et al., 2022). Os principais hotspots localizam-se nas seguintes regiões: (1) Região Serrana do Rio de Janeiro, com destaque para Petrópolis e Volta Redonda (443 óbitos); (2) Região Metropolitana de São Paulo, incluindo São Paulo, Guarujá, Mauá e Cubatão (1297 óbitos); (3) Serra da Mantiqueira em São Paulo, abrangendo Campos do Jordão e Santo Antônio do Pinhal (192 óbitos); (4) Zona da Mata Mineira Norte, com Natividade e Muriaé; (5) Região Metropolitana de Belo Horizonte, incluindo Ibirité e Contagem; (6) Zona da Mata Mineira Central, com Juiz de Fora como núcleo principal; (7) Região Serrana Fluminense Interior, abrangendo Santa Maria Madalena e Cantagalo;

(8) Região Serrana do Espírito Santo, com Santa Leopoldina e Fundão; e (9) Zona da Mata Mineira Sul, incluindo Ervália e Araponga.

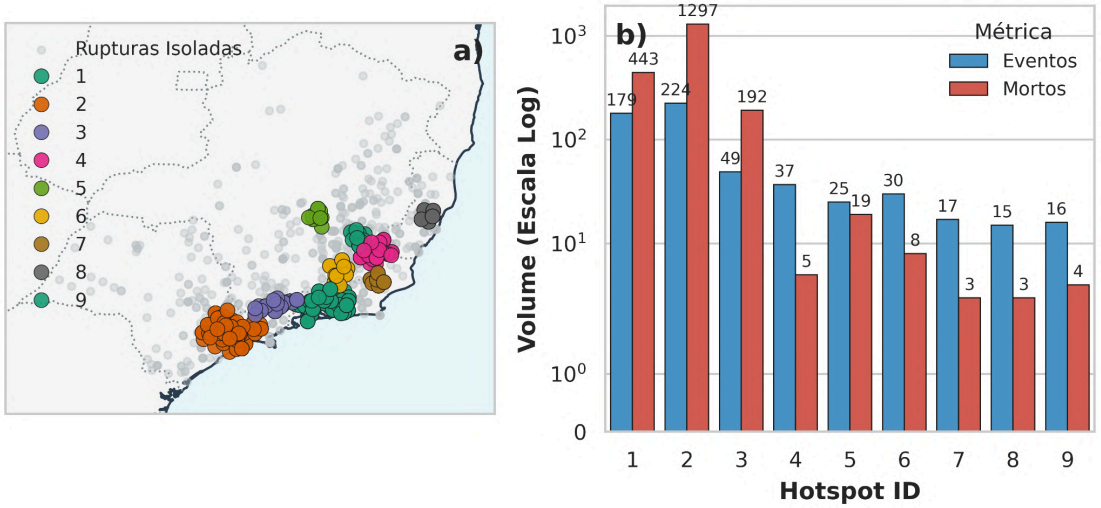


Figura 5 — Detecção de *hotspots* críticos e perfil de letalidade via *DBSCAN*. No mapa (A), os círculos coloridos representam os 9 aglomerados identificados (raio = 30 km) e os pontos cinza indicam eventos isolados. Os hotspots localizam-se em: (1) Região Serrana do Rio de Janeiro; (2) Região Metropolitana de São Paulo; (3) Serra da Mantiqueira; (4) Zona da Mata Mineira Norte; (5) Região Metropolitana de Belo Horizonte; (6) Zona da Mata Mineira Central; (7) Região Serrana Fluminense Interior; (8) Região Serrana do Espírito Santo; e (9) Zona da Mata Mineira Sul. O gráfico de barras (B) detalha, em escala logarítmica, o volume de eventos (azul) e o total de óbitos (vermelho) por ID de *hotspot*.

A rede de covariância de *Spearman* apresentou correlações moderadas para os dias antecedentes ao evento, corroborando o acumulado de D_{-1} *a* D_{-10} como principal preditor meteorológico para a ocorrência de deslizamentos na região. A associação positiva entre letalidade e densidade populacional indica que a exposição demográfica é o principal condicionante da severidade dos impactos, sendo o acumulado de D_{-1} *a* D_{-10} a variável com maior grau de conectividade na rede e maior relevância preditiva para a ocorrência dos eventos (RAMOS et al., 2022; ALVALÁ et al., 2019).

A análise de *composites* da precipitação antecedente aos deslizamentos indica que eventos de alta letalidade acumulam 141,7 mm de D_{-1} *a* D_{-10}, com anomalia de 14,1% sobre a climatologia de verão. Entre D_0 e D_{-1}, os eventos de maior impacto registraram aumentos de até 115% na precipitação em relação à média histórica do período.

A Figura 6 apresenta a análise comparativa entre eventos comuns e extremos: nos eventos de maior impacto, o volume de precipitação é apenas 1,2 vezes maior que em eventos comuns, enquanto a população exposta é 4,7 vezes superior. Os resultados indicam que a combinação de chuva antecedente elevada e alta densidade populacional em encostas está fortemente associada à letalidade dos deslizamentos na Região Sudeste, resultado consistente com Alvalá et al. (2019).

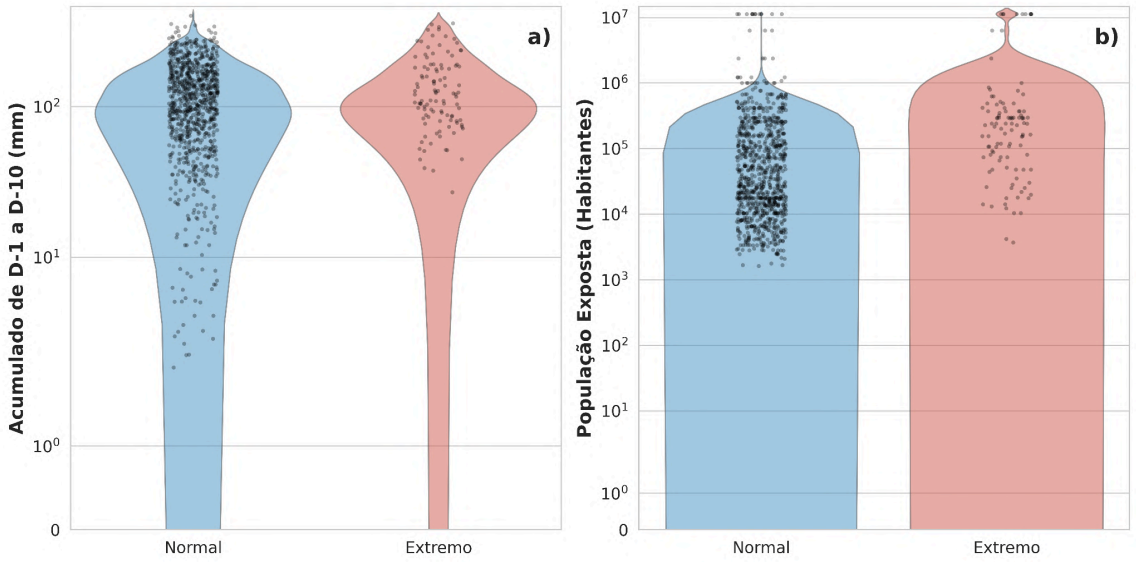


Figura 6 — Diagnóstico comparativo de extremos. Os gráficos de violino apresentam a distribuição em escala logarítmica para (A) o acumulado de precipitação de D_{-1} *a* D_{-10} e (B) a população exposta aos eventos, segmentados entre eventos normais e extremos. Os valores de significância estatística via teste de *Mann-Whitney U* registrou $p = 0,016$ para o acumulado pluviométrico e $p < 0,001$ para a população exposta, ambas para um nível de significância de 5%. Os pontos pretos representam a dispersão individual dos 1082 registros de deslizamentos.

CONCLUSÕES E CONSIDERAÇÕES FINAIS

Os resultados indicam que os deslizamentos no Sudeste do Brasil, em geral, decorrem de processos de saturação progressiva do solo, e não de gatilhos imediatos isolados. A assinatura temporal da precipitação antecedente aos deslizamentos sugere que monitorar o acumulado pluviométrico de D_{-1} *a* D_{-10} é mais eficaz para a previsão de deslizamentos do que o foco exclusivo em alertas de intensidade horária, processo condicionado, principalmente, pela persistência da ZCAS no trimestre de verão.

A abordagem por aprendizado de máquina permitiu identificar padrões espaciais e meteorológicos estruturados no risco geológico: o *K-Means* isolou três tipologias de ruptura com

dinâmicas meteorológicas e sociais distintas, enquanto o *DBSCAN* indicou que deslizamentos em *hotspots* críticos são 12,4 vezes mais letais que eventos isolados. O principal resultado indica que, nos eventos acima do P_{90}, a precipitação acumulada é apenas 1,2 vezes maior que em eventos comuns, mas a população exposta é 4,7 vezes superior.

Os dados evidenciam que a mortalidade por deslizamentos no Sudeste está associada principalmente à vulnerabilidade urbana das populações expostas. Sistemas de alerta precoce são ferramentas necessárias, porém complementares a políticas de ordenamento territorial voltadas à redução da exposição em encostas de risco. Esta pesquisa contribui para a compreensão dos movimentos de massa como fenômeno resultante da interação entre forçantes meteorológicas e condicionantes sociais de vulnerabilidade.

AGRADECIMENTOS

Os autores agradecem aos centros que forneceram os dados para esse estudo e à FAPEMIG, UNIFEI e ao CNPq pelo suporte financeiro.

REFERÊNCIAS

ALCÂNTARA, E. et al. Machine learning approaches for mapping and predicting landslide-prone areas in São Sebastião (Southeast Brazil). **Natural Hazards Research**, Beijing, v. 5, n. 2, p. 247–261, 2025.

ALVALÁ, R. C. S. et al. Mapping characteristics of at-risk population to disasters in the context of Brazilian early warning system. **International Journal of Disaster Risk Reduction**, Amsterdam, v. 41, p. 101326, 2019.

ESCOBAR, G. C. J. Zona de Convergência do Atlântico Sul (ZCAS): Critério de detecção para uso em centros operacionais de previsão de tempo. **CEP**, v. 12, p. 10, 2019.

FERNANDES, N. F. et al. Topographic controls of landslides in Rio de Janeiro: field evidence and modeling. **Catena**, Amsterdam, v. 55, n. 2, p. 163–181, 2004.

FUNK, C. et al. The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. **Scientific Data**, London, v. 2, n. 1, p. 1–21, 2015.

RAMOS, L. M. et al. Landslide susceptibility mapping based on rainfall scenarios: a case study from São Paulo in Brazil. **AIMS Geosciences**, v. 8, n. 3, p. 438–451, 2022.

REICHSTEIN, M. et al. Deep learning and process understanding for data-driven Earth system science. **Nature**, London, v. 566, n. 7743, p. 195–204, 2019.