

## Mapeamento da suscetibilidade a movimentos de massa na Quarta Colônia (RS) por meio de aprendizado de máquina

### *Mapping landslide susceptibility in Quarta Colônia (RS) using machine learning techniques*

**Marco Antônio da Rosa Soares** 

Doutorando em Geografia  
Universidade Federal de Santa Maria, Rio Grande do Sul, Brasil  
[ma-arco@hotmail.com](mailto:ma-arco@hotmail.com)

**Haline Dugolin Ceccato** 

Doutoranda em Engenharia Civil e Ambiental  
Universidade Federal de Santa Maria, Rio Grande do Sul, Brasil  
[haline.ceccato@gmail.com](mailto:haline.ceccato@gmail.com)

**Romario Trentin** 

Doutor em Geografia  
Universidade Federal de Santa Maria, Rio Grande do Sul, Brasil  
[romario.trentin@gmail.com](mailto:romario.trentin@gmail.com)

### Resumo

Os movimentos de massa são processos frequentemente desencadeados por eventos pluviométricos, potencializados por condições específicas do relevo. A presente pesquisa avaliou a aplicabilidade de métodos de aprendizado de máquina na Quarta Colônia, Rio Grande do Sul, a partir de um inventário com 2.318 cicatrizes construído com imagens de satélite, bases secundárias e verificação manual. Foram testados algoritmos supervisionados e não supervisionados, empregando variáveis geomorfológicas, topográficas e hidrológicas, como declividade, elevação, curvatura, orientação, Índice Topográfico de Umidade e Índice de Concentração de Rugosidade. Entre os modelos supervisionados, o Gradient Boosting apresentou o melhor desempenho geral e foi adotado como referência. Nos métodos de clustering, K-Means e Louvain demonstraram potencial para identificar áreas mais suscetíveis, especialmente com a inclusão das variáveis hidrológicas, embora com concordância apenas moderada em relação ao modelo supervisionado. Os resultados indicam que as classes mais altas de suscetibilidade concentram a maior parte das cicatrizes, em ambos os métodos, evidenciando um padrão de resultado. Conclui-se que as



<https://doi.org/10.28998/contegeo.11i.25.20964>

Artigo publicado sob a [Licença Creative Commons 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

Submetido em: 13/04/2026

Aceito em: 18/05/2026

Publicado: 06/07/2026

e-Location: 20964

abordagens são complementares, sendo os modelos supervisionados mais precisos e os não supervisionados mais úteis para análises exploratórias

**Palavras-chave:** clustering; movimentos de massa; aprendizado de máquina; aprendizado não supervisionado; Quarta Colônia.

### **Abstract**

*Mass movements are processes frequently triggered by intense rainfall events and intensified by specific relief conditions. This research evaluated the applicability of machine learning methods in the Quarta Colônia region, Rio Grande do Sul, based on an inventory of 2,318 scars constructed from satellite imagery, secondary datasets, and manual verification. Supervised and unsupervised algorithms were tested using geomorphological, topographic, and hydrological variables, such as slope, elevation, curvature, aspect, Topographic Wetness Index, and Roughness Concentration Index. Among the supervised models, Gradient Boosting showed the best overall performance and was adopted as a reference. In the clustering approaches, K-Means and Louvain demonstrated potential for identifying more susceptible areas, especially with the inclusion of hydrological variables, although with only moderate agreement compared to the supervised model. The results indicate that the highest susceptibility classes concentrate most of the mapped scars in both methods, revealing a consistent spatial pattern. It is concluded that the approaches are complementary, with supervised models providing greater predictive accuracy and unsupervised methods being more suitable for exploratory analyses.*

**Keywords:** clustering; landslides; machine learning; unsupervised learning; Quarta Colônia.

## **INTRODUÇÃO**

De acordo com Maciel Filho e Nummer (2011), os movimentos de massa envolvem o deslocamento coletivo de porções de solo e rocha. Eles podem ser classificados isoladamente ou em conjunto, combinando diferentes mecanismos em um único evento. A diferenciação entre essas ocorrências depende da velocidade do movimento e do tipo de material envolvido (Guidicini e Nieble, 1983).

Esse processo natural é desencadeado principalmente por terremotos, chuvas intensas ou ações humanas (Shang et al., 2003). Sua ocorrência em regiões habitadas representa alto risco devido ao elevado potencial destrutivo, indo além de prejuízos materiais e podendo resultar, em situações extremas, em óbitos.

No Rio Grande do Sul, apesar de episódios anteriores de chuvas intensas com impactos graves nas décadas de 1970 e 1980, o evento de 2024 se sobressaiu pelos

acumulados pluviométricos recordes, superando a média histórica do estado. Em Porto Alegre, por exemplo, o mês foi o mais chuvoso desde 1910.

A suscetibilidade a movimentos de massa refere-se à probabilidade de ocorrência desses eventos em uma dada região e pode ser avaliada por meio de métodos estatísticos ou computacionais (Liu et al., 2021). Esse mapeamento é fundamental, pois aprimora o entendimento dos fenômenos, minimiza riscos à sociedade e auxilia gestores na tomada de decisões mais assertivas (Silveira et al., 2014).

A maior parte dos estudos sobre zoneamento de suscetibilidade a movimentos de massa adota abordagens supervisionadas, com técnicas da inteligência artificial (Moayedi et al., 2018; Kawabata e Bandidas, 2009; Dou et al., 2015) e estatísticas integradas a sistemas GIS (Kamp et al., 2008; Silveira et al., 2014; Roccati, 2021). Essas técnicas requerem um conjunto consistente de dados espaciais para elaboração dos modelos de suscetibilidade.

Entre os maiores desafios dessa abordagem destaca-se a necessidade de construção de um inventário de movimentos de massa. Essa etapa é fundamental no processo metodológico, pois compila e sistematiza informações sobre a localização, a tipologia e as características dos eventos registrados. Conforme destacado por Wieczorek (1984), trata-se de um procedimento essencial para o mapeamento, uma vez que fornece a base necessária para a análise da distribuição espacial e dos padrões do fenômeno.

Embora a maioria dos estudos sobre suscetibilidade a movimentos de massa priorize técnicas supervisionadas, existe uma lacuna notável na literatura quanto a algoritmos não supervisionados. Nesse contexto, o presente trabalho propõe a adoção de técnicas não supervisionadas, como os algoritmos de *clustering*.

Essa abordagem deve ser considerada, pois o *clustering* não demanda um inventário com rótulos que indicam a localização exata dos movimentos de massa, eliminando assim uma fase crítica do processo de mapeamento. Essa abordagem não só simplifica o mapeamento em si, mas também possibilita análises em escala ampliada com redução expressiva no tempo de processamento, o que amplia consideravelmente o escopo de aplicação da metodologia proposta.

## Clustering: Definições e aplicações

A técnica de *clustering* pertence ao campo do aprendizado de máquina, mais precisamente ao aprendizado não supervisionado, no qual se opera sem a presença de rótulos associados aos objetos, ou seja, o conjunto de dados assume a forma  $(x_1, x_2, \dots, x_n)$ . Ademais, trata-se de um método de análise de dados que organiza elementos com base em suas semelhanças, possibilitando a detecção de padrões relevantes, mas também, dependendo do algoritmo utilizado, a separação entre informações úteis e ruídos (Kriegel et al., 2011).

O *clustering* pode ser categorizado em diversas abordagens, conforme o critério usado para formar os agrupamentos. Nas técnicas baseadas em centroides, como *K-Means* e *CLARA*, os *clusters* são gerados em torno de um ponto central, real ou virtual, do *dataset*, com a métrica de distância atuando como um fator decisivo na composição dos grupos. Por outro lado, os métodos baseados em densidade, como aqueles que contam eventos por unidade de área ou empregam o *Kernel Density Estimator*, detectam grupos em zonas de alta concentração de pontos, delimitados por regiões de baixa densidade, sem assumir uma distribuição pré-definida dos dados.

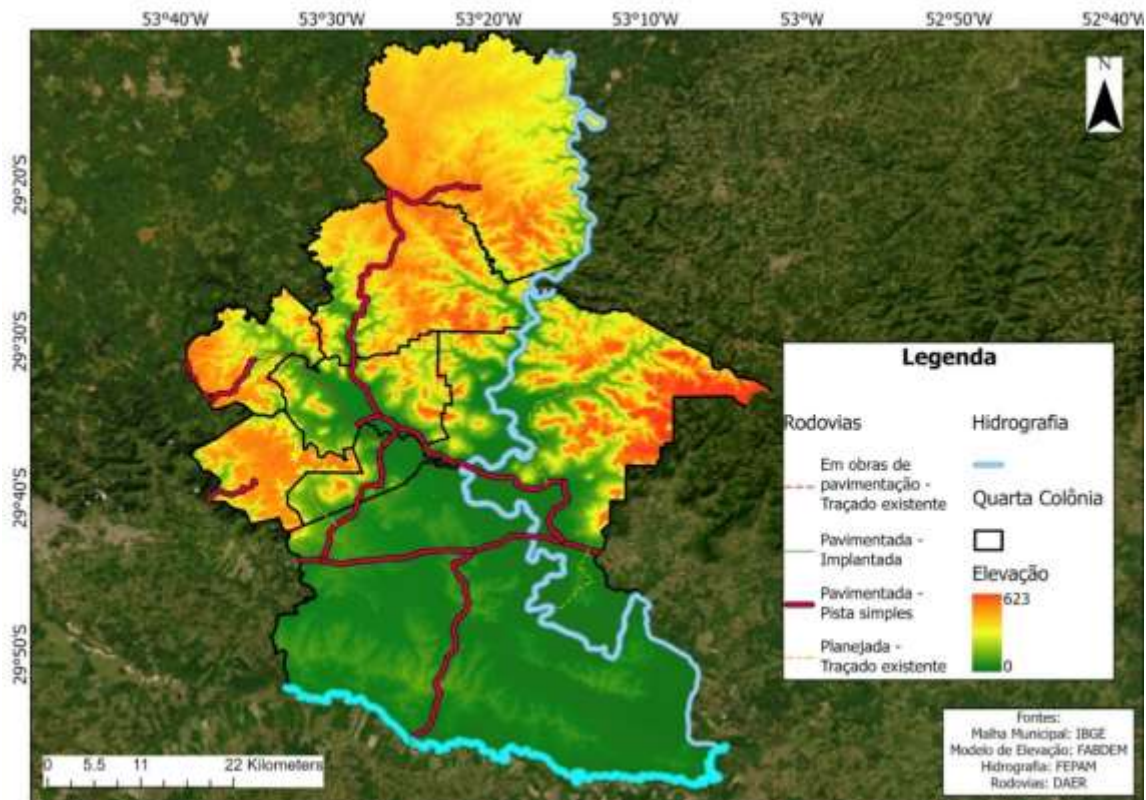
Ademais, distinguem-se entre métodos paramétricos e não paramétricos: os primeiros visam otimizar uma função de custo ligada ao agrupamento, enquanto os segundos evitam supor uma distribuição probabilística pré-estabelecida. Finalmente, os algoritmos de partição, a exemplo do *K-Means*, demandam a especificação antecipada do número de clusters e prosseguem por iterações até atingir a convergência, enquanto os hierárquicos, que podem ser aglomerativos ou divisivos, geram agrupamentos de forma progressiva e possibilitando a avaliação de soluções parciais via estruturas como o dendrograma.

## Área em estudo

A Quarta Colônia (Figura 1), localizada na região central do estado do Rio Grande do Sul, abrange nove municípios: Agudo, Dona Francisca, Faxinal do Soturno, Ivorá, Nova Palma, Pinhal Grande, Restinga Seca, São João do Polêsine, Silveira Martins. Selecionada como área de estudo devido ao seu elevado potencial do ponto de vista geocientífico e geoturístico, contando com fósseis de dinossauros no período

Triássico (250 Ma – 199 Ma), os mais antigos conhecidos (Godoy et al., 2012). Por fim, a região ainda registra um número expressivo de movimentos de massa.

**Figura 1** - Localização da área de estudo



Fonte: Autores, 2026.

A Quarta Colônia se enquadra no subtipo Tropical II, caracterizado como “medianamente úmido, com variação longitudinal de temperaturas médias. Entre maio e agosto, a região registra chuvas intensas e geadas, associadas à queda de temperatura provocada pela Massa Polar Atlântica (MPA) (Ziemann, 2016). Os meses de menores precipitações são julho, agosto e setembro (Ferrari, 2008). As temperaturas médias variam de 23° a 25° nas estações quentes e de 14° a 15° nas frias (Fantinelli e Becker, 2011).

Quanto à geologia, a Quarta Colônia ocupa as bordas da Bacia Sedimentar do Paraná, cuja formação remonta ao final do Paleozoico, no Rio Grande Do Sul. Esse processo ocorreu no período Permiano ao Cretáceo (Milani, 1997). As unidades geológicas locais integram o registro sedimentar da Bacia do Chaco-Paraná, abrangendo o intervalo do Triássico Inferior ao Cretáceo Inferior, na porção leste da bacia.

No contexto estratigráfico regional, destaca-se o Grupo São Bento, formado pela Formação Serra Geral (constituída por rochas vulcânicas efusivas associadas a derrames magmáticos fissurais, com intercalações basais de arenitos eólicos intertrapes) e Botucatu (representado por arenitos finos a médios, consolidados em sistemas deposicionais continentais) (Zerfass, 2007).

Sotoposto a esse conjunto, o Grupo Rosário do Sul engloba as Formações Sanga do Cabral (unidade mais antiga, de idade triássica inferior), Santa Maria e Caturrita (ambas exclusivas do Rio Grande do Sul), caracterizado por sequências litológicas diversificadas, incluindo arenitos, siltitos e lamitos de origem fluvial a lacustre.

A Formação Serra Geral é constituída por derrames basálticos da idade jurássico-cretácea, associados ao magmatismo da Província Mágmatca do Paraná. Em adição, é composta por basaltos e latitos da Fácies Caxias e riodacitos e quartzo latitos da Fácies Gramado.

A Formação Tupanciretã, conforme descrito por Menegotto, Sartori e Maciel Filho (1968), assenta-se em contato discordante sobre as rochas vulcânicas da Formação Serra Geral. Esta unidade sedimentar apresenta predomínio conglomerados polimíticos e arenitos arcosianos, com estratificação cruzada e seleção moderada. Em sua porção superior, observa-se a intercalação progressiva de lâminas argilosas delgadas, o que indica variações nas condições do ambiente de deposição durante a fase final de sua sedimentação.

Por fim, recobrando essas rochas, ocorrem depósitos sedimentares de idade mais recente, associados às vertentes e aos sistemas fluviais atuais, representados por colúvios e alúvios. Os colúvios consistem em materiais inconsolidados, de granulometria variável, formados predominantemente por processos gravitacionais em encostas, enquanto alúvios são constituídos por sedimentos arenosos, de composição majoritariamente arcoseana, acumulados ao longo dos cursos de água e em planícies aluviais adjacentes (Silva, 2019).

Diante das condições e dos fatores geológicos presentes na Quarta Colônia, a região assume grande relevância científica em razão da frequência de novas descobertas. Além disso, a área apresenta, de forma recorrente, eventos associados a desastres naturais, destacando-se os deslizamentos (Reckziegel, 2007).

## PROCEDIMENTOS METODOLÓGICOS

### Elaboração do inventário

A fase inicial do estudo envolveu, principalmente, a construção de um inventário dos movimentos de massa na região da Quarta Colônia – RS, incluindo apenas aqueles que ocorreram no evento de precipitação extrema de maio de 2024. Essa etapa revela-se essencial, uma vez que serviu de base para a validação das conclusões obtidas com os algoritmos de aprendizado não supervisionado. Apesar de sua descrição parecer direta, representa uma tarefa metodologicamente complexa, a qual exige uma quantidade significativamente alta de tempo. O processo foi estruturado em dois procedimentos principais:

- 1) Identificação e mapeamento sistemático dos eventos de movimentos de massa na área de estudo, com base na análise de imagens de satélite, registros históricos e dados secundários.
- 2) Coleta de variáveis morfométricas, geológicas e ambientais associadas aos eventos mapeados, visando caracterizar suas condições de ocorrência e mecanismos.

A detecção dos movimentos de massa baseou-se em três fontes primárias de dados. Inicialmente, obtiveram-se as coordenadas fornecidas pela equipe de pesquisa *Goddard Space FlightCenter* (GSFC) da *National Aeronautics and Space Administration* (NASA), que registrou os eventos ocorridos após 29 de abril de 2024 por meio de imagens do *PlanetScope*, com resolução espacial de 3 metros.

A segunda consulta de dados foi o *WebMapa* de Movimentos de Massa para equipes de apoio na situação de calamidade que ocorreu em maio de 2024 no RS, elaborado pelo Laboratório Latitude – Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia (CEPSRM)/ Programa de Pós-Graduação em Sensoriamento Remoto (PPGSR) / Departamento de Geodésia / Instituto de Geociências (IGeo) | Universidade Federal do Rio Grande do Sul (UFRGS) (Andrades-Filho et al., 2025).

A terceira e final validação foi conduzida manualmente pelo autor, abrangendo toda a área da Quarta Colônia. O objetivo foi confirmar as posições previamente identificadas e detectar novos movimentos de massa que poderiam não ter sido vistos pelas análises anteriores das outras equipes.

Para esse mapeamento, empregaram-se imagens do satélite *PlanetScope*, com resolução espacial de 3 metros, capturadas aproximadamente um mês após o início dos eventos. Complementarmente, utilizaram-se imagens pré-evento para discernir se as cicatrizes observadas resultavam especificamente desse episódio ou de ocorrências anteriores.

Para além disso, utilizou-se a hidrografia fornecida pela Fundação Estadual de Proteção Ambiental (FEPAM), na escala 1:25.000, e as curvas de nível do Rio Grande do Sul, disponibilizadas pela Infraestrutura Estadual de Dados Espaciais (IEDE-RS), também na escala 1:25.000. A validação das cicatrizes partiu dessa comparação entre imagens e da análise visual da textura, forma e contexto.

As variáveis empregadas para explicar a ocorrência do fenômeno foram definidas com base nos resultados da revisão da literatura. A seguir, descrevem-se os métodos de obtenção dos condicionantes utilizados como variáveis de entrada nos algoritmos, bem como aqueles que subsidiaram análises complementares apresentadas na seção de resultados.

- Litologia: Utilizou-se o mapa litológico disponibilizado pela Companhia de Pesquisa de Recursos Minerais (CPRM), na escala 1:750.000.
- Elevação: Obtida a partir do Modelo Digital de Elevação (MDE) do produto FABDEM, com resolução espacial de 30 metros.
- Orientação da Encosta: A orientação foi inicialmente calculada utilizando a ferramenta Aspect do ArcGIS Pro. Posteriormente, em razão da natureza circular dessa variável, os valores foram transformados em seus componentes seno e cosseno, em radianos, a fim de evitar distorções no cálculo das distâncias empregadas pelos algoritmos de *clustering*.
  - A curvatura foi obtida com a ferramenta “Curvature” do ArcGIS Pro.
  - O ângulo de inclinação da encosta foi calculado em graus, com o mesmo MDE utilizado para a geração da orientação da encosta.
  - O Topographic Wetness Index (TWI) foi calculado no ArcGis Pro utilizando a equação  $TWI = \ln\left(\frac{A_s}{\tan(\beta)}\right)$  onde  $A_s$  é a área de contribuição específica e o  $\beta$  é a declividade.
  - O *Index of Convergence* (ICR) foi calculado no software SAGA GIS utilizando a ferramenta *Convergence Index*. Em ambos, utilizou-se o produto FABDEM para o processamento.

## Programação e aplicação do *clustering*

A implementação e a execução dos algoritmos de *clustering* foram efetuadas em *Python*, com o auxílio da biblioteca de machine learning *Scikit-Learn*. As análises cobriram todo o território da Quarta Colônia, após a divisão da área de estudo em uma grade regular de 30 metros de espaçamento, onde cada célula atua como unidade de análise e incorpora os valores dos fatores previamente estabelecidos.

A densidade de movimentos (número de eventos por km<sup>2</sup>) foi usada como indicador da suscetibilidade. Nessa abordagem, calculou-se a razão entre o número de eventos do inventário contidos no cluster e a área ocupada por ele. *Clusters* com maior densidade foram interpretados como zonas de maior suscetibilidade.

O algoritmo *K-Means* foi aplicado via implementação padrão do *Hard C-Means*, com 100 reinicializações para otimização inicial e limite de 500 iterações por execução. A determinação do número ótimo de *clusters* *k* baseou-se em índices de qualidade de agrupamento, como o Índice de Silhueta e o Índice de Davies-Bouldin. O processamento ocorreu de duas maneiras, sem e com a inclusão dos condicionantes TWI e ICR.

Para interpretar os resultados, computaram-se as médias das variáveis por cluster em dois cenários. O primeiro considera apenas os pontos do inventário de movimentos de massa e o segundo todos os pixels da área de estudo. Essa comparação revela se as cicatrizes mapeadas se concentram espacialmente em clusters específicos e se essas representam frações relevantes do território.

A seleção do melhor resultado integrou critérios estatísticos, observações empíricas e métricas de qualidade, incluindo densidade de eventos e médias das variáveis. Essa abordagem permitiu identificar a configuração superior, tendo como base os valores de referência do inventário previamente elaborado.

O algoritmo *Louvain* foi implementado de forma análoga ao *K-Means*. Entretanto, em sua forma mais simples, desconsiderando integrações com dendrogramas ou cortes hierárquicos em etapas específicas, esse algoritmo não permite a identificação prévia do número de agrupamentos. Por conseguinte, o *Louvain*, nesse trabalho, foi utilizado para segmentar inicialmente o conjunto de dados em agrupamentos iniciais, cujas variáveis foram resumidas por estatísticas descritivas para produzir centroides representativos.

Posteriormente, aplicou-se o *K-Means* sobre esses centroides, originando os *meta-clusters*. Essa etapa visou reduzir o número inicial de agrupamentos, simplificar a interpretação e viabilizar comparações com outros métodos de classificação utilizados.

O número final de meta-clusters foi determinado testando vários valores de *k*, com base no Índice da Silhueta e no Índice de Davies-Bouldin. Dentre as opções, *k* = 8 demonstrou boa separabilidade, sendo selecionado para ambas as configurações, por equilibrar tanto compactação interna quanto interpretabilidade.

### **Programação e aplicação dos métodos supervisionados – *Random Forest* e *Gradient Boosting***

O pré-processamento iniciou com a geração de pontos aleatórios em quantidade equivalente aos movimentos de massa mapeados na área de estudo, para compor o conjunto balanceado a ser processado para criar o modelo. Adicionalmente, todas as camadas foram reprojetadas para o sistema de referência SIRGAS2000 UTM Zone 22S. Após isso, removeram-se os pontos de borda com valores nulos.

A divisão entre treino e teste empregou todos os pontos de deslizamento mais pontos aleatórios para treinar os modelos. Em seguida, aplicou-se o modelo treinado a todos os *pixels* da área de estudo. Além disso, nenhum critério de parada adicional foi definido para os algoritmos.

A avaliação de performance utilizou métricas como *Area Under Roc Curve* (AUC), Acurácia (CA), *F1-Score*, precisão, *recall* e *Matthews Correlation Coefficient* (MCC). Os valores finais derivaram do conjunto de teste independente, com validação cruzada no treino guiando a otimização de hiperparâmetros. A relevância das variáveis foi quantificada por permutação com AUC como métrica principal.

O modelo de maior desempenho foi rasterizado pixel a pixel, produzindo um mapa contínuo de probabilidade de ocorrência de movimento de massa, variando entre 0 e 1. Para classificação temática em suscetibilidade foi adotada a divisão em cinco classes de intervalos iguais da escala de probabilidade, sendo elas: Muito Baixa, Baixa, Moderada, Alta e Muito Alta. Embora quantis ou outras ideias fossem viáveis, optou-se por intervalos iguais para promover comparabilidade e reprodutibilidade.

## Comparação dos resultados gerados pelas duas abordagens de algoritmo

A comparação entre os mapas finais de suscetibilidade a movimentos de massa envolveu a análise da sobreposição espacial, tomando o mapeamento supervisionado como referência e confrontando-o com os produtos dos métodos não supervisionados. O foco principal foi medir o grau de coincidência espacial entre técnicas que particionam a área sob premissas distintas, além de examinar forças e fraquezas de cada uma para aplicações em planejamento e gestão ambiental.

A quantificação da concordância recorreu a índices estatísticos como Concordância Global (Overall), Coeficiente Kappa, Recall e Precisão. As métricas foram calculadas individualmente para cada classe de suscetibilidade, o que permitiu uma avaliação granular do desempenho.

Complementarmente, avaliou-se a eficácia dos métodos em concentrar as cicatrizes dos inventários nas classes de maior suscetibilidade, por meio do número absoluto de eventos por classe e da proporção da área correspondente.

## ANÁLISES E DISCUSSÕES

No inventário elaborado, foram identificados 2.318 movimentos de massa, distribuídos em nove municípios que integram a Quarta Colônia. As variáveis utilizadas no inventário e na caracterização dos movimentos de massa abrangem tanto a geometria da encosta, expressa por declividade, curvatura e orientação, quanto as condições topográficas e hidrológicas, representadas pela elevação, TWI e ICR.

As estatísticas descritivas básicas associadas a cada uma dessas variáveis são apresentadas no Quadro 1, sendo os valores obtidos exclusivamente nos pontos de ruptura dos movimentos de massa mapeados.

**Quadro 1** – Estatísticas descritivas das variáveis geomorfológicas e hidrológicas.

	Valor Máximo	Valor Mínimo	Média	Mediana
Curvatura	11,5805	-7,6178	0,1053	0,0716
Declividade	58,7612	1,2488	30,1414	30,8823
Elevação	551	60,8499	305,89	313,25
TWI	19,6713	2,7698	5,2075	5,2074
ICR	89,2481	-60,3322	2,9810	2,8420

Fonte: Autores, 2026.

## Análise e discussão dos resultados dos algoritmos não supervisionados

A aplicação do algoritmo *K-Means* com  $k = 10$  proporcionou uma discriminação consideravelmente refinada das unidades morfológicas, mantendo a concentração dos deslizamentos restrita a poucos agrupamentos. No cenário sem a inclusão das variáveis TWI e ICR, observou-se predominância dos eventos nos *clusters* C9, C6 e C10, todos caracterizados por declividades médias próximas ou superiores a  $30^\circ$ . Apesar da coerência geomorfológica desses grupos, particularmente em relação a influência da declividade, detectou-se elevada heterogeneidade interna no *cluster* C9, cuja curvatura média próxima de zero decorreu da compensação estatística entre formas côncavas e convexas, o que limita a interpretação da suscetibilidade representativa desse conjunto.

Com a inclusão das variáveis TWI e ICR, verificou-se uma redistribuição significativa dos eventos, evidenciando maior poder discriminatório do algoritmo. Nesse novo cenário, destacaram-se sobretudo os *clusters* C8 e C6, associados a declividades elevadas e curvaturas expressivas, respectivamente convexas e côncavas. O *cluster* C9, que anteriormente concentrava o maior número de registros, passou a apresentar densidade notavelmente inferior, indicando que a inclusão desses condicionantes permitiu a separação de padrões antes não capturados pelo método de funcionamento de algoritmos de *clustering* baseados em centroides. O Quadro 2 sintetiza os principais resultados obtidos para os *clusters* mais representativos em ambos os cenários.

**Quadro 2** – Principais resultados do algoritmo K-Means

	Cluster	Declividade (°)	Curvatura	Área (%)	Nº de eventos	Densidade
Sem TWI/ICR	C6	30,34	+2,3918	9,52	411	3,536
	C9	32,81	+0,0252	14,59	1275	5,785
	C10	30,59	-2,1557	7,63	371	3,098
Com TWI/ICR	C6	33,24	-1,2655	5,93	839	2,562
	C8	32,46	+1,2900	6,78	990	5,052
	C9	22,57	+0,1066	12,37	251	0,982

Fonte: Autores, 2026.

Em resumo, ainda que o aumento do número de partições possa refinar a compartimentação espacial, a inclusão das novas variáveis mostrou-se mais relevante, no presente trabalho, para a melhoria da discriminação da suscetibilidade. Ainda assim,

persiste um contraste significativo entre áreas de alta e baixa densidade de eventos, o que evidencia as limitações do método em distinguir classes intermediárias quando aplicado a um inventário associado a um evento pluviométrico extremo.

O algoritmo de *Louvain* foi empregado como etapa inicial de segmentação espacial, produzindo 163 *clusters* no cenário sem a inclusão de TWI e ICR. A análise da distribuição dos pontos revelou forte concentração em poucos agrupamentos, o que justificou a agregação posterior por meio do *K-Means*, resultando em 8 *meta-clusters* mais interpretáveis.

No primeiro cenário (sem TWI e ICR), os *meta-clusters* 1,7 e 8 concentraram a maior parte dos deslizamentos e apresentaram as mais elevadas densidades, correspondendo a áreas com declividades maiores e curvaturas mais acentuadas. Os demais agrupamentos exibiram densidades reduzidas, sugerindo menor relevância para a previsão da suscetibilidade.

Com a inclusão das variáveis TWI e ICR, o algoritmo passou a gerar 74 *clusters iniciais*, com uma distribuição dos pontos mais homogênea ao longo da área de estudo. Após a agregação em *meta-clusters*, verificou-se uma reorganização dos padrões espaciais, na qual os *clusters* 5 e 8 passaram a concentrar as maiores densidades. Diferentemente do cenário anterior, não se observa um único agrupamento que concentre simultaneamente superfícies côncavas e convexas, o que sugere maior capacidade de discriminação morfológica por parte do método. O Quadro 3 sintetiza os agrupamentos mais representativos em ambos os casos.

**Quadro 3** – Principais resultados do algoritmo Louvain

	Cluster	Declividade (°)	Curvatura	Área (%)	Nº de eventos	Densidade
Sem TWI/ICR	C1	31,46	+0,2853	10,96	1223	3,86
	C7	31,66	-1,5992	6,90	618	3,10
	C8	30,71	+2,8050	2,99	288	3,33
Com TWI/ICR	C5	31,92	+1,1710	8,11	1012	4,32
	C8	31,28	-1,0181	13,84	949	2,37

Fonte: Autores, 2026.

A partir desse quadro, nota-se que a inclusão do TWI e ICR promoveu uma redistribuição dos deslizamentos entre os agrupamentos, reduzindo a sobreposição

entre padrões morfológicos distintos e reforçando a capacidade da abordagem em identificar zonas mais críticas de suscetibilidade.

### Análise e discussão dos resultados dos algoritmos supervisionados

Inicialmente, foram treinados modelos apenas com as variáveis geomorfológicas e topográficas previamente utilizadas na análise, configurando o cenário de referência. Os resultados obtidos são apresentados no Quadro 4.

**Quadro 4** – Métricas de qualidade obtidas no aprendizado supervisionado

Modelo	AUC	CA	F1	Precisão	Recall	MCC
RF	0,812	0,739	0,755	0,712	0,805	0,482
GB	0,823	0,758	0,785	0,706	0,885	0,533

Fonte: Autores, 2026.

Os resultados evidenciam desempenho satisfatório para ambos os classificadores, com destaque para o *Gradient Boosting*, que supera o *Random Forest* na maioria das métricas. Observa-se maior capacidade discriminativa (AUC = 0,823), maior sensibilidade na identificação de ocorrência (Recall = 0,885) e melhor equilíbrio global na classificação (MCC = 0,533). O RF apresenta desempenho levemente inferior, ainda assim, robusto e aceitável para o contexto da análise.

Em etapa subsequente, incluíram-se os índices TWI e ICR no conjunto de variáveis preditoras, com o objetivo de avaliar potenciais ganhos no desempenho dos modelos. Os resultados obtidos são apresentados no Quadro 5.

**Quadro 5** – Métricas de qualidade com a inclusão do TWI e ICR

Modelo (TWI/ICR)	AUC	CA	F1	Precisão	Recall	MCC
RF	0,807	0,740	0,738	0,747	0,740	0,487
GB	0,822	0,760	0,756	0,779	0,760	0,538

Fonte: Autores, 2026.

A comparação entre os cenários indica que a inclusão do TWI e do ICR trouxe mudanças no desempenho dos classificadores. No caso do RF, registrou-se uma leve queda na AUC e no F1-Score, acompanhada de pequenas flutuações nas demais métricas. O GB manteve desempenho estável, com variação mínima na AUC e moderado no MCC, mas apresentou redução considerável no Recall.

Essa proximidade de desempenho entre os cenários pode ser atribuída, em parte, à redundância informacional das variáveis adicionadas. O TWI é derivado

matematicamente a partir da declividade e da área de contribuição, o que indica que uma porção da sua variabilidade já estava representada no modelo pelas variáveis morfométricas previamente inseridas. Assim, a incorporação do TWI e ICR tende a gerar incrementos de desempenho marginalmente pequenos, sem impacto decisivo na discriminação final da suscetibilidade.

De maneira geral, o modelo *Gradient Boosting* apresentou um desempenho ligeiramente superior ao *Random Forest*, com maior consistência entre os dois cenários de processamento, o que indicou ser o classificador mais adequado no contexto do trabalho. Por fim, diante da queda acentuada do Recall quando o TWI e o ICR foram incluídos, optou-se por adotar o modelo treinado sem essas variáveis, priorizando a capacidade de detectar o maior número possível de ocorrências. A análise de importância por permutação indicou que a declividade e a elevação foram as variáveis mais relevantes para a predição da suscetibilidade, sendo seguidas, em menor grau, por orientação e curvatura da encosta.

### Mapeamento da suscetibilidade por métodos supervisionados e não supervisionados

Entre os algoritmos não supervisionados avaliados, o *K-Means* com a inclusão das variáveis TWI e ICR apresentou o melhor resultado em termos de sensibilidade para a identificação das classes de Alta e Muito Alta Suscetibilidade. A comparação espacial entre o modelo supervisionado e o *K-Means*, em porcentagem, é apresentada no Quadro 6.

**Quadro 6** – Comparação pixel a pixel do *Gradient Boosting* e *K-means*

GB / K-Means	Muito Baixa	Baixa	Moderada	Alta	Muito Alta	Total
Muito Baixa	24,98	4,12	13,99	3,18	0,08	46,34
Baixa	0,45	2,27	2,88	4,38	0,84	10,82
Moderada	0,15	3,25	4,29	8,49	4,72	20,89
Alta	0,11	3,25	2,75	8,18	6,05	20,34
Muito Alta	0,02	0,23	0,10	0,23	1,02	1,60
Total	25,66	13,13	24,00	24,36	12,71	100

Fonte: Autores, 2026.

A maior concordância absoluta foi observada na classe Muito Baixa, indicando que áreas de menor suscetibilidade apresentam maior homogeneidade morfométrica e são mais facilmente reconhecidas por ambos os métodos. As classes intermediárias,

por sua vez, demonstraram maior dispersão, evidenciando sobreposição entre padrões morfométricos adjacentes.

Na classe Muito Alta, observou-se elevada concordância relativa, com aproximadamente 63% da área classificada como Muito Alta pelo *K-Means* coincidindo com a mesma classe no modelo supervisionado. Além disso, o *K-Means* concentrou cerca de 77% das cicatrizes inventariadas na classe Muito Alta, sugerindo alta sensibilidade na identificação de áreas associadas aos deslizamentos mapeados. Esse comportamento, contudo, também implicou a ampliação territorial dessa classe.

Já o algoritmo *Louvain*, com a inclusão das variáveis TWI e ICR, apresentou o melhor desempenho após a reclassificação em cinco níveis de suscetibilidade. A incorporação dessas variáveis contribuiu para uma discriminação espacial mais refinada das áreas de maior suscetibilidade.

A concordância global entre o mapeamento supervisionado e o não supervisionado foi de 51,27%, conforme exibido no Quadro 7. Observou-se que a maior correspondência ocorreu na classe Muito Baixa, o que indica boa concordância na delimitação de áreas menos suscetíveis. Por outro lado, as classes intermediárias apresentaram maior dispersão, evidenciando sobreposição entre níveis adjacentes de suscetibilidade.

**Quadro 7** - Comparação pixel a pixel do *Gradient Boosting* e *Louvain*

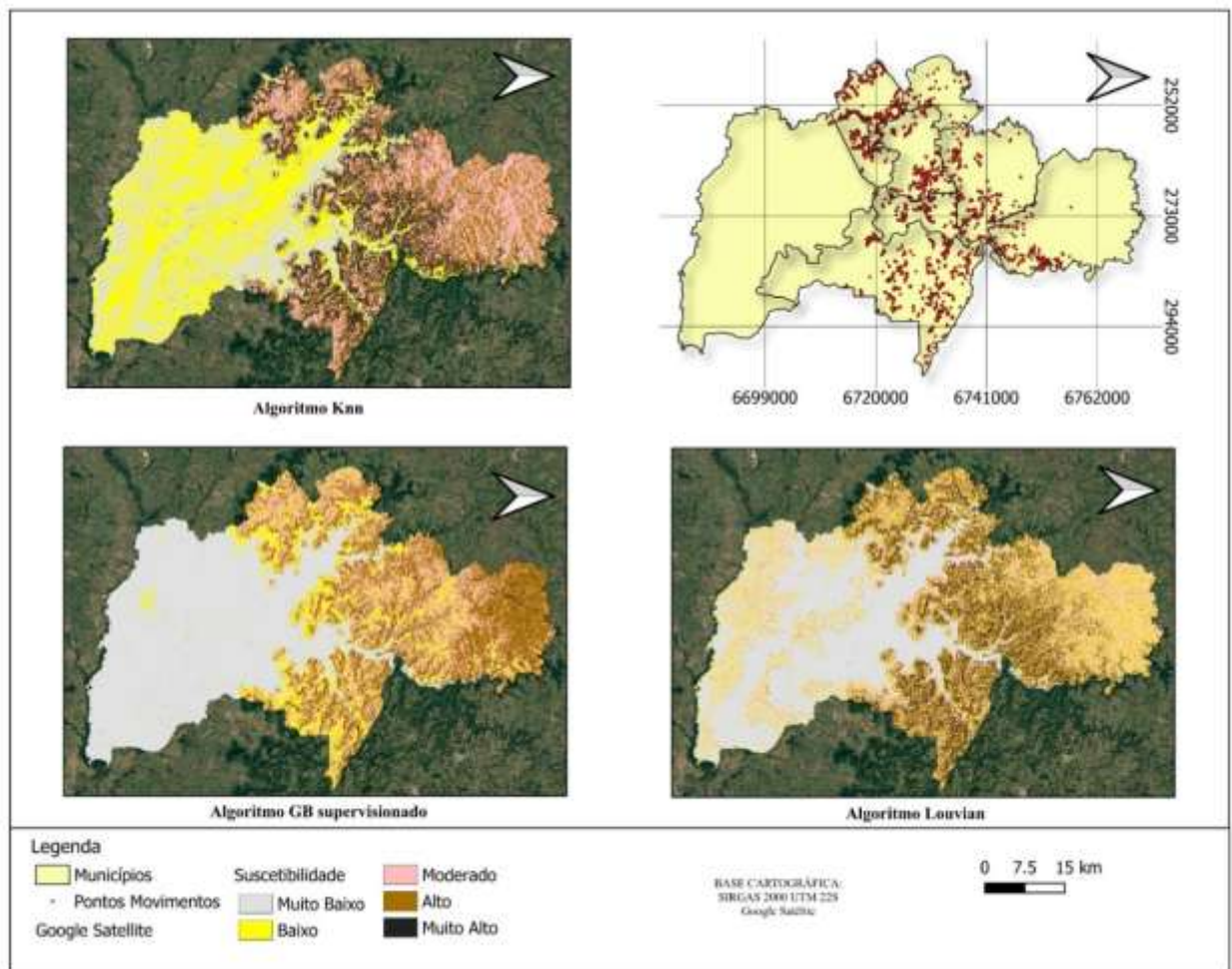
GB / Louvain	Muito Baixa	Baixa	Moderada	Alta	Muito Alta	Total
Muito Baixa	36,44	4,13	5,41	0,20	0,15	46,33
Baixa	1,46	2,16	4,32	2,12	0,77	10,82
Moderada	0,51	3,97	7,32	6,46	2,64	20,90
Alta	0,55	4,32	6,87	4,69	3,91	20,34
Muito Alta	0,09	0,20	0,29	0,38	0,65	1,60
Total	39,04	14,77	24,22	13,85	8,12	100

Fonte: Autores, 2026.

Ao comparar com os dados do inventário, verifica-se que ambos os métodos concentraram a maior parte das cicatrizes em uma fração relativamente reduzida da área de estudo, cerca de 22% da área total, correspondendo a 81% a 84% dos eventos localizados. Em contrapartida, o modelo supervisionado apresentou uma distribuição mais restrita para a classe Muito Alta, favorecendo maior especificidade espacial e menor extensão de áreas de alta suscetibilidade. O *Louvain*, por sua vez, ampliou consideravelmente essa classe, aumentando a sensibilidade na detecção de áreas potencialmente suscetíveis, ainda que com redução da precisão espacial.

De modo geral, os resultados indicam que os métodos supervisionados tendem a apresentar resultados mais precisos, enquanto os algoritmos de *clustering* possuem caráter mais exploratório, ampliando áreas prioritárias para análise e monitoramento. Assim, o modelo supervisionado mostra-se mais indicado para aplicações que demandam maior precisão espacial e menor taxa de falsos positivos, conforme pode ser observado na Figura 2.

**Figura 2** – Comparação entre os resultados gerados



Fonte: Autores, 2026.

## CONSIDERAÇÕES FINAIS

Esta pesquisa teve como objetivo avaliar a aplicabilidade de técnicas de aprendizado de máquina no mapeamento da suscetibilidade a movimentos de massa na região da Quarta Colônia, RS. Para isso, foram testados métodos supervisionados e não supervisionados, utilizando como referência um inventário de cicatrizes

construído a partir do evento de precipitação extrema ocorrido em maio de 2024 no Rio Grande do Sul. Os resultados mostraram coerência entre os padrões identificados pelos algoritmos e a distribuição observada no inventário, confirmando o potencial das técnicas de *clustering* como ferramentas exploratórias na análise de suscetibilidade.

Entre os condicionantes avaliados, a elevação e a declividade se destacaram como as variáveis de maior relevância, enquanto a orientação apresentou influência intermediária e a curvatura teve um papel menos influente. Nos métodos supervisionados, o *Gradient Boosting* apresentou o melhor desempenho geral, servindo como referência para a comparação com os resultados dos algoritmos não supervisionados.

Os resultados indicaram ainda que a inclusão dos índices TWI e ICR contribuiu para melhorar o desempenho dos métodos não supervisionados, especialmente na distinção de áreas com maior suscetibilidade. Embora a concordância pixel a pixel com o produto supervisionado tenha sido apenas moderada, os valores obtidos mostraram que os algoritmos *Louvain* e *K-Means* são capazes de representar as classes extremas de suscetibilidade, ainda que apresentem maior sobreposição nas classes médias. Em contrapartida, ambos os métodos tenderam a ampliar as áreas classificadas como mais suscetíveis, aumentando a sensibilidade em detrimento da precisão espacial.

Dessa forma, conclui-se que os métodos supervisionados e não supervisionados possuem finalidades complementares. O *Gradient Boosting* mostrou-se mais adequado para aplicações que exigem maior precisão e menor taxa de falsos positivos, enquanto os algoritmos de *clustering* se destacam por sua rapidez e capacidade exploratória, sendo úteis em etapas preliminares de análise e na formulação de hipóteses espaciais. Assim, os métodos não supervisionados não devem ser vistos como substitutos do mapeamento supervisionado, mas como instrumentos complementares para apoiar a interpretação dos padrões de suscetibilidade.

## REFERÊNCIAS

ANDRADES-FILHO, C. O., MEXIAS, L. F. S., QUEVEDO, R. P., HERRMANN, P. B., OLIVEIRA, G. G., CREMON, É. H., CARGNIN, B. R., REIS, M. S., NOVAKOSKI, K., dos Santos, D. C., SLUTER, C. R., IESCHECK, A. L., GUASSELLI, L. A., RIBEIRO JUNIOR, M., GRUBER, N. L. S., PHILIPP, R. P., MICHELIN, C. R. L., VIERO, A. P., GONZATTI, C., ROSA DA CÂMARA ROSA, M. L., BARBOZA-PINZON, E., SCHUMACHER, R. G., MOLANO CÁRDENAS, S. M., SOARES, V. M., NÚÑEZ, W. P., MENDONÇA, R. R., RIGHI DA SILVA, M. O., SILVA, M. C., BRESSANI, L. A., PERES, L. D., MEVEL, L. B., DORNELES, J. I., CACCIATORE, J. A., JACQUES, F.

de M., DUARTE, L. C., DINIZ, F. de C., PETRY, L., SCHWARZER, G., CARNEIRO, M. & GIACCOMI, B. The biggest landslide event in Brazil: preliminary analysis of the Rio Grande do Sul mega disaster in May 2024. **Landslides**, 22, 3615–3624, 2025.

DOU, J., YAMAGISHI, H., POURGHASEMI, H. R., YUNUS, A. P., SONG, X., XU, Y. & ZHU, Z. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. **Natural Hazards**, 78, 1749–1776, 2015.

FANTINELLI, D. & BECKER, E. A gastronomia italiana como atrativo turístico na região da Quarta Colônia – Rio Grande do Sul – Brasil. **Revista Disciplinarum Scientia**. Série: Humanas. Santa Maria, v. 12, n. 1, 1-117, 2011.

FERRARI, R. **Modelagem dinâmica do uso e cobertura da terra da Quarta Colônia**. 2008. 131 f. Dissertação (Mestrado em Geografia) – Universidade Federal de Santa Maria, Santa Maria, 2008.

GUIDICINI, G. & NIEBLE, C. **Estabilidade de taludes naturais e de escavação**. São Paulo: Edgard Blücher, 1983. 217 p.

GODOY, L. P.; FIGUEIRÓ, A. S.; DA ROSA, Átila, A. S.; HEINRICH, C.; GUARESCHI, A. P.; SCHERER, C. M. S. **Proposta de Geoparque Quarta Colônia (RS)**. In: SCHOBENHAUS, C.; SILVA, C. R. (org.). Geoparques do Brasil: propostas. Rio de Janeiro: CPRM, 2012. p. 483-514.

KAMP, U.; GROWLEY, B. J.; KHATTAK, G. A. & OWEN, L. A. GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. **Geomorphology**, 101(4), 631–642, 2008.

KAWABATA, D. & BANDIBAS, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an artificial neural network (ANN). **Geomorphology**, v. 113, n. 1–2, p. 97–109, 2009.

KRIEGLER, H.P.; PEER, K.; SANDER, J. & ZIMEK, A. Density-based clustering. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 1, n. 3, p. 231–240, 2011.

LIU, M.; LIU, J.; SHENGHUA X.; ZHOU, T.; MA, Y.; ZHANG, F. & KONEČNÝ, M. Landslide susceptibility mapping with the fusion of multi-feature SVM model based FCM sampling strategy: a case study from Shaanxi Province. **International Journal of Image and Data Fusion**, v. 12, n. 4, p. 349–366, 2021.

MENEGOTTO, E; SARTORI, P. & MACIEL FILHO, C. Nova sequência sedimentar sobre a Serra Geral no Rio Grande do Sul. **Publicação Especial do Instituto de Solos e Culturas**, v. 1, p. 1–19, 1968.

MILANI, E. **Evolução tectono-estratigráfica da Bacia do Paraná e seu relacionamento com a geodinâmica fanerozóica do Gondwana sul-ocidental**.

1997. 225 f. Tese (Doutorado em Geociências) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 1997.

MOAYEDI, H.; MOHAMMAD, M.; MOSALLANEZHAD, M.; RASHID, A.S.A. & PRADHAN, B. Modification of landslide susceptibility mapping using optimized PSO-ANN technique. **Engineering with Computers**, v. 35, p. 967–984, 2019.

NUMMER, A. & MACIEL FILHO, C. **Introdução à geologia de engenharia**. Santa Maria: Editora da UFSM, 2011. 456 p.

RECKZIEGEL, B. **Levantamento dos desastres desencadeados por eventos naturais adversos no estado do Rio Grande do Sul no período de 1980 a 2005**. 2007. 284 f. Dissertação (Mestrado em Geografia) – Universidade Federal de Santa Maria, Santa Maria, 2007.

ROCCATI, A.; PALIAGA, G.; LUINO, F.; FACCINI, F. & TURCONI, L. GIS-based landslide susceptibility mapping for land use planning and risk assessment. **Land**, v. 10, n. 2, p. 162, 2021.

SHANG, Y.; YANG, Z.; LI, L.; LIU, D.; LIAO, Q. & WANG, Y. A super-large landslide in Tibet in 2000: background, occurrence, disaster, and origin. **Geomorphology**, v. 54, n. 3–4, p. 225–243, 2003.

SILVA, G.C. **Modelo preditivo de suscetibilidade a escorregamentos planares no rebordo planalto entre Santa Maria e Candelária-RS, utilizando o método do valor informativo**. 2019. 211 f. Tese (Doutorado em Geografia) – Programa de Pós-Graduação em Geografia, Universidade Federal de Santa Maria, Santa Maria, 2019. 211 p.

SILVEIRA, C. T.; FIORI, A. B.; SCHILIPACK, P. & DIAS, S. M. Mapeamento preliminar da suscetibilidade natural a movimentos de massa da Serra do MAR Paranaense apoiado na análise digital do relevo. **Revista Brasileira de Geomorfologia**, v. 15, n. 1, 2014. em Geografia.

WIECZOREK, G. Preparing a detailed landslide-inventory map for hazard evaluation and reduction. **Bulletin of the Association of Engineering Geologists**, v. 21, n. 3, p. 337–342, 1984.

ZERFASS, H.; SANDER, A.; NORONHA, F. de L.; SCHERER, O. L. B.; DIAS, A. de A.; PARISI, G. N.; MACHADO, J. L. F.; RIBEIRO, A. M.; FERIGOLO, J.; ROSA, Á. A. S. & SILVA, R. C. **Geologia da Folha Agudo SH.22-V-C-V, estado do Rio Grande do Sul: escala 1:100.000**. Porto Alegre: CPRM, 2007. 1 CD-ROM.

ZIEMANN, D. R. **Estratégias de geoconservação para a proposta do Geoparque Quarta Colônia – RS**. 2016. 241 f. Dissertação (Mestrado em Geografia) – Universidade Federal de Santa Maria, Santa Maria, 2016