

Aplicação do Modelo Aditivo Generalizado espacial para a modelagem da susceptibilidade a ocorrência de deslizamentos

Tatiana Dias Tardelli Uehara¹, Eduardo Celso Gerbi Camargo¹,
Camile Sothe², Thales Sehn Körting¹

¹Divisão de Observação da Terra e Geoinformática (DIOTG)
Instituto Nacional de Pesquisas Espaciais (INPE)
Caixa Postal 515 – 12227-010 – São José dos Campos – SP – Brasil

²School of Earth, Environment and Society – McMaster University
Hamilton – ON – Canadá

{tatiana.uehara, eduardo.camargo, thales.korting}@inpe.br,

sothec@mcmaster.ca

Abstract. *The susceptibility mapping is a fundamental task to prevent social and environmental impacts resulting from landslides, which makes it possible to identify areas with greater or lesser probability of occurring the event. In this context, this work uses a spatial approach based on Generalized Additive Models (GAM) associated to driving factors, to estimate and analyze the spatial distribution of landslide susceptibility in the Rio Rolante Hydrographic Basin (RS/Brazil). The results reveal that the model was able to satisfactorily distinguish areas with greater or lesser susceptibility of landslide occurrence. All points of landslides occurrence have estimated probabilities greater than 0.999. The non-occurrence class showed greater variability in estimated values.*

Resumo. *O mapeamento de susceptibilidade é uma tarefa fundamental para prevenir impactos sociais e ambientais decorrentes de deslizamentos de terra, que possibilita identificar as áreas com maior ou menor probabilidade deste tipo de evento ocorrer. Este trabalho emprega uma abordagem espacial baseada em Modelos Aditivos Generalizados (GAM) associado a fatores potencializadores, para estimar e analisar a distribuição espacial da susceptibilidade a deslizamentos de terra na Bacia Hidrográfica do Rio Rolante (RS/Brasil). Os resultados revelam que o modelo foi capaz de distinguir satisfatoriamente as áreas com maior ou menor susceptibilidade. Todos os pontos de ocorrência de deslizamentos apresentaram probabilidades estimadas superiores a 0,999. O padrão de não ocorrência apresentou maior variabilidade de valores estimados.*

1. Introdução

Deslizamentos são um dos tipos de movimentos de massa, caracterizados por processos geomorfológicos naturais que ocorrem em diversas partes do planeta. Tais processos consistem em movimentos descendentes de material provindo da encosta que podem ser provocados por terremotos, degelo de neve ou chuvas intensas, podendo também ser causados ou intensificados por atividades antrópicas [Guzzetti et al. 2012]. Tal fenômeno causa grandes perdas econômicas e sociais, principalmente quando ocorre em áreas densamente povoadas. O objetivo desta pesquisa foi analisar os fatores potencializadores

à ocorrência de deslizamentos de terra empregando um Modelo Aditivo Generalizado (*Generalized Additive Model - GAM*) espacial e elaborar um mapa de suscetibilidade indicando os locais com maior e menor probabilidade de ocorrência.

2. Material e métodos

2.1. Área de Estudo

A área de estudos (Figura 1) inclui toda extensão da bacia hidrográfica do Rio Rolante (RS/Brasil). Sua área de drenagem abrange 828 km² e os valores de elevação variam entre 20m e 1040m. A área é predominantemente ocupada por floresta nativa, silvicultura, agricultura, pastagem e ocupação antrópica rural. No que tange aos aspectos geomorfológicos e climáticos, a área encontra-se na unidade Serra Geral com clima subtropical úmido. Trabalhos anteriores detectaram por volta de 300 cicatrizes de deslizamentos no local. Todas as cicatrizes foram geradas após um evento extremo de precipitação ocorrido no dia 5 de janeiro de 2017, em que foram registrados entre 90 e 272mm de chuva em quatro horas [SEMA 2017].

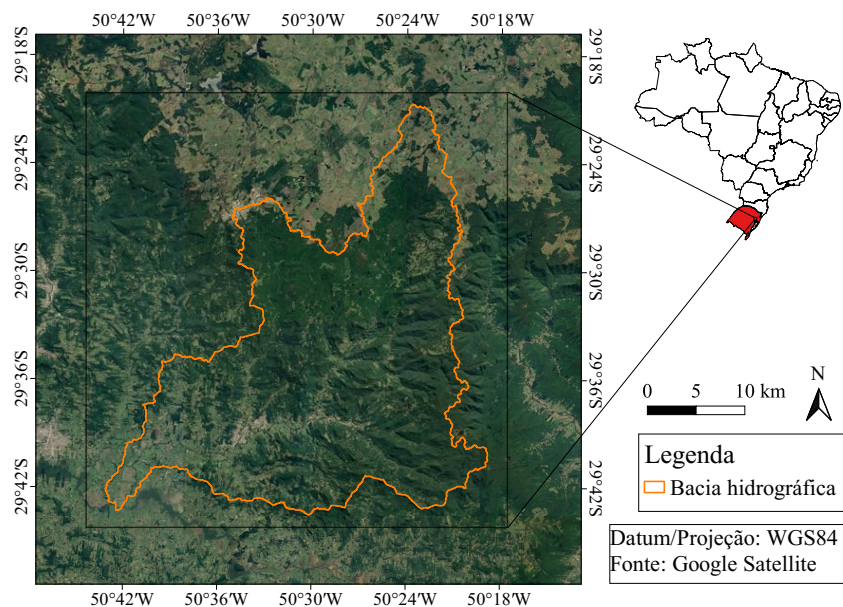


Figura 1. Localização da área de estudo.

2.2. Dados

A partir do inventário de cicatrizes de deslizamento estabelecido na região [Quevedo et al. 2020], foi gerado o centróide de cada cicatriz, o que resultou em 335 *pontos de ocorrência* de deslizamento. A mesma quantidade de *pontos de não ocorrência* foi gerada aleatoriamente, respeitando a distância delimitada por um buffer de 5m ao redor das cicatrizes, para que não houvesse sobreposição entre os *pontos de ocorrência* e *não ocorrência*. Da totalidade do conjunto amostral, 20% dos *pontos de ocorrência* e 20% dos *pontos de não ocorrência* foram retirados e guardados para a validação do modelo.

Para cada ponto amostral há um conjunto de variáveis potencializadoras associadas ao evento investigado. Um valor médio para cada uma das variáveis foi estabelecido dentro de um buffer de 5m em torno de cada ponto amostral. A seleção das variáveis potencializadoras consistiram na extração de atributos a partir do Modelo Digital de Elevação (MDE), mensuração da distância entre as cicatrizes e canal de drenagem mais próximo e a atribuição de pesos às classes de solo, conforme os trabalhos realizados por [Sothe et al. 2017, Saro et al. 2016]. O MDE utilizado provém de dados disponibilizados pelo projeto *Alaska Satellite Facility (ASF)*, no qual, a partir do MDE do *Shuttle Radar Topography Mission (SRTM)* obtem-se um produto de 12,5m de resolução espacial. Por meio da ferramenta SAGA disponível no software QGIS, foram geradas as seguintes variáveis geomorfológicas: declividade, curvatura horizontal e curvatura vertical.

Os tipos de solos foram obtidos do mapa de solos do Brasil, produzido pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em escala 1:5.000.000. Foram identificados quatro tipos de solos e a cada um deles atribuído um peso (0,0 a 0,9) com base no seu grau de erodibilidade [da Silva and Alvares 2007]. Os tipos de solo com seus respectivos graus de erodibilidade e peso são: Cambissolo Húmico e Neossolo Litólico (Muito Alto - 0,8); Argissolo Vermelho-Amarelo (Alto - 0,7) e Nitossolo Vermelho (Médio - 0,5). Os valores da variável distância dos rios foram extraídos a partir da mensuração da distância euclidiana entre os pontos amostrais e o vetor de hidrografia mais próximo.

2.3. Modelo

A estrutura de modelagem empregada é baseada na ideia de processos pontuais espaciais, pela qual se pode estimar uma superfície de suscetibilidade que varia continuamente na região de interesse, conforme proposto por [Kelsall and Diggle 1998]. Neste trabalho, o processo pontual subjacente é caracterizado a partir de um conjunto de pontos espacialmente distribuídos sobre a área de estudo, cuja variável resposta é do tipo binária, isto é, *pontos suscetíveis* com valor 1, e 0 em caso contrário. Isso caracteriza a uma abordagem de regressão binária com a inclusão de fatores potencializadores, um tipo de GAM [Hastie and Tibshirani 1990] que pode ser entendido como um modelo linear generalizado estendido por uma componente aditiva espacial que, por suposição, varia suavemente na região de estudo. Sob estas considerações [Kelsall and Diggle 1998] propõem um GAM, com uma função de ligação *logit*, conforme a Equação 1:

$$\text{logit}[p(s)] = \log \frac{p(s)}{1 - p(s)} = \beta \mathbf{x} + g(s) \quad (1)$$

em que: \mathbf{x} é o vetor de covariáveis, β são seus efeitos e g é uma componente espacial (uma função suave), porém desconhecida, das coordenadas espaciais s . Se a suscetibilidade é considerada constante na região de estudo, então $g(s) = 0$ e o modelo reduz-se a um modelo de regressão logística usual [Hosmer et al. 1989]. O procedimento de estimação de β e $g(s)$ baseia-se em métodos iterativos usuais de modelos aditivos generalizados, conforme descritos em [Hastie and Tibshirani 1990]. Neste procedimento, $g(s)$ é estimado usando regressão Kernel ponderada [Wand and Ripley 2006].

O modelo proposto (Equação 1) estima uma superfície de suscetibilidade, porém é importante avaliar se esta superfície varia significativamente na região de estudo, ou seja, se existem evidências estatísticas suficientes para rejeitar a hipótese nula de suscetibilidade constante ($H_0: g(s) = 0$). Além disso, é de interesse a construção de contornos

de tolerância que auxiliam na identificação de áreas onde a suscetibilidade é significativamente superior ou inferior à média global. Para tal, o teste global da suscetibilidade e a identificação de áreas de baixa e alta suscetibilidade foram realizados empregando o método de simulação de Monte Carlo proposto por [Kelsall and Diggle 1998]. A aplicação do modelo foi realizada no software R (versão 2.10), utilizando o pacote “SPGAM”.

3. Resultados e Discussão

Inicialmente, a seleção de variáveis foi realizada pelo método *backwards* [Derksen and Keselman 1992]. A variável solo, por apresentar um *p-valor* não significativo, foi removida do modelo. Uma possível explicação pode ser a escala dos dados pedológicos que generalizou os tipos de solo em quatro grandes categorias que não detalham as variações entre áreas mais ou menos susceptíveis. A Equação 2 revela o modelo final. As variáveis potencializadoras incluídas, bem como suas respectivas significâncias estatísticas (*p-valor*), foram as seguintes: β_0 (1,08E-12), elevação (4,07e-04), declividade (1,96e-30), curvatura vertical (2,59e-04) e curvatura horizontal (1,47e-03).

$$\text{logit}[p(s)] = \beta_0 + \beta_{Elev.} + \beta_{Decliv.} + \beta_{Curv.vertical} + \beta_{Curv.horizontal} + g(s) \quad (2)$$

A partir do modelo expresso na Equação 2 estimou-se a suscetibilidade associada à ocorrência de deslizamento de terra, conforme ilustra a Figura 2. Em seguida, foram realizadas 500 simulações via método de Monte Carlo, cujos resultados observados foram: i) houve uma variação espacial global significativa da suscetibilidade (*p-valor* = 0,001992), portanto a hipótese nula de suscetibilidade constante na região ($g(s) = 0$) foi descartada; ii) a identificação de áreas delimitadas por linhas de contorno de 2,5% e 97,5%, de aproximadamente 95% de tolerância, apontam os locais onde a suscetibilidade é significativamente inferior ou superior à média global, respectivamente. A partir da Figura 2, identifica-se que as áreas com maiores probabilidades de ocorrência de deslizamento (apresentadas em tons vermelhos) localizam-se nas regiões com alta declividade, onde se concentraram as cicatrizes geradas em 2017.

Um estudo semelhante pode ser visto em Sothe et al. (2017), que aplicou um modelo GAM para a análise de suscetibilidade na Bacia do Rio Luís Alves (SC/Brasil). Seus resultados revelaram que 100% das *amostras de ocorrência* foram estimadas com probabilidades acima de 0,9 e 75% das *amostras de não ocorrência*, abaixo de 0,5.

A adequação do modelo empregado foi realizada por meio da comparação: i) entre o conjunto de *pontos de ocorrência* reservado para validação e o conjunto de *pontos de ocorrência* extraído da superfície estimada, ambos sobre a mesma localização geográfica; ii) de forma similar, entre o conjunto de *pontos de não ocorrência* reservado para validação e o conjunto de *pontos de não ocorrência* estimado. O que se espera é que a distribuição estatística de tais conjuntos de pontos (*de ocorrência* e *não ocorrência*) sejam distintas, com média estimada para o conjunto de *pontos de ocorrência* próxima de um (1) e para o conjunto de *pontos de não ocorrência* em torno de zero (0). Isto pode ser avaliado graficamente através da construção de um boxplot, conforme mostra a Figura 3.

Na Figura 3 observa-se que as distribuições das probabilidades estimadas para os *pontos de ocorrência* e *não ocorrência* são bastantes distintas. Pode-se afirmar que

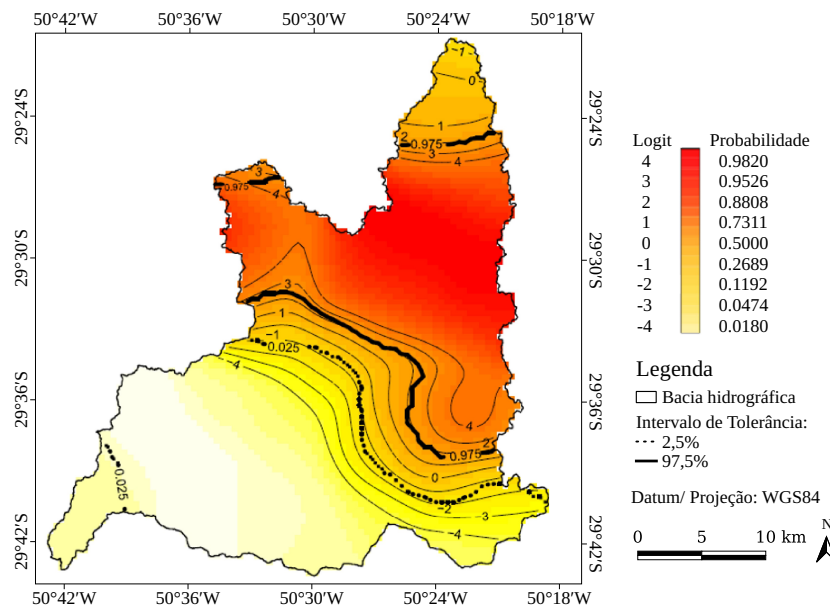


Figura 2. Mapa de susceptibilidade estimada.

100% dos *pontos de ocorrência* apresentam probabilidades estimadas superiores a 0,999. Por outro lado, nas áreas não suscetíveis a deslizamentos de terra, os *pontos de não ocorrência* apresentaram uma maior variação de valores de probabilidade estimado, sendo que cerca de 50% das amostras desta classe apresentaram valores inferiores a 0,472. Esta maior incerteza associada às estimativas dos *pontos de não ocorrência* deve ser objeto de investigação de futuros trabalhos, como por exemplo, a inclusão de outros fatores potencializadores no modelo.

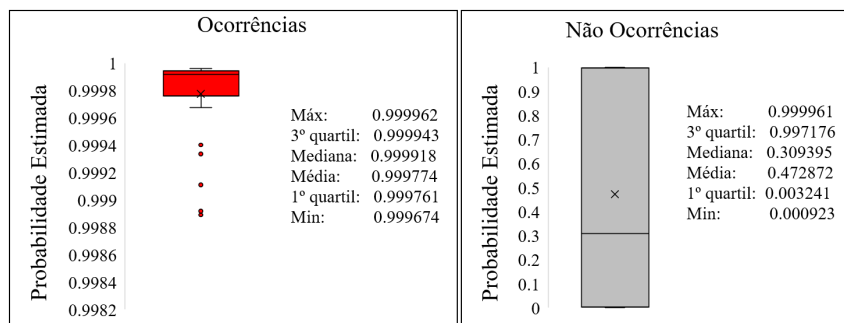


Figura 3. Boxplot da probabilidade estimada para os *pontos de ocorrência* e *não ocorrência*, associados a deslizamentos de terra.

4. Conclusão

Este trabalho apresentou uma abordagem para modelar e mapear a suscetibilidade a deslizamentos de terra a partir de informações pontuais. Para tal, empregou-se um método baseado em modelos aditivos generalizados espaciais, com a inclusão de fatores potencializadores correlacionados com o evento investigado. Neste estudo, as variáveis Elevação,

Declividade, Curvatura Horizontal e Curvatura Vertical foram fatores potencializadores significativos no modelo empregado. Quanto à avaliação do modelo, ficou evidente que as estimativas de suscetibilidade foram mais precisas nas áreas em que ocorreram deslizamentos de terra, todos os pontos de ocorrência apresentaram probabilidades estimadas superiores a 0,999. Já para os locais menos vulneráveis, as estimativas foram menos precisas, sendo que a metade dos pontos de não ocorrência apresentaram valores inferiores a 0,472. De um modo geral, trata-se de um resultado preliminar, que pode auxiliar no planejamento de ações mais específicas e dirigidas a essas áreas. Para trabalhos futuros recomenda-se investigar a influência da variação dos dados de entrada em diferentes escalas, bem como a exploração e inclusão de outros fatores potencializadores.

Referências

- da Silva, A. M. and Alvares, C. A. (2007). Levantamento de informações e estruturação de um banco dados sobre a erodibilidade de classes de solos no estado de são paulo. *Geociências (São Paulo)*, 24(1):33–41.
- Derksen, S. and Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2):265–282.
- Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K.-T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112(1-2):42–66.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC press.
- Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (1989). The multiple logistic regression model. *Applied logistic regression*, 1:25–37.
- Kelsall, J. E. and Diggle, P. J. (1998). Spatial variation in risk of disease: a nonparametric binary regression approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(4):559–573.
- Quevedo, R. P., Guasselli, L. A., De Oliveira, G. G., and Ruiz, L. F. C. (2020). Modelagem de áreas suscetíveis a movimentos de massa: avaliação comparativa de técnicas de amostragem, aprendizado de máquina e modelos digitais de elevação. *Geociências (São Paulo)*, 38(3):781–795.
- Saro, L., Woo, J. S., Kwan-Young, O., and Moun-Jin, L. (2016). The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: A case study of inje, korea. *Open Geosciences*, 8(1):117–132.
- SEMA (2017). Diagnóstico preliminar: Descritivo dos eventos ocorridos no dia 5 de janeiro de 2017 entre as regiões dos municípios de são francisco de paula e rolante/rs. *Secretaria do Ambiente e Desenvolvimento Sustentável, Porto Alegre*, page 26.
- Sothe, C., Camargo, E. C. G., Gerente, J., Rennó, C. D., and Monteiro, A. M. V. (2017). Uso de modelo aditivo generalizado para análise espacial da suscetibilidade a movimentos de massa. *Revista do Departamento de Geografia*, 34:68–81.
- Wand, M. and Ripley, B. (2006). Kernsmooth: Functions for kernel smoothing for wand & jones (1995). *R package version*, 2:22–19.