

## Mapeamento e caracterização da estabilidade das encostas na Sub-bacia do Ribeirão Araranguá através do modelo Random-Forest.

### Mapping and characterization of slope stability in the Sub-basin of Ribeirão Araranguá through the Random-Forest model.

Gabriel G. Guaragna<sup>1\*</sup>, Rafael R. Higashi<sup>1</sup>, Thiago D. Vieck<sup>2</sup>

---

#### RESUMO

O presente estudo buscou avaliar a eficácia do modelo *Random-Forest* para o auxílio de mapeamentos de movimentos gravitacionais de massa. Neste trabalho o modelo foi definido em linguagem de programação R com o auxílio do pacote SDM. O algoritmo através deste pacote se mostrou de fácil implementação e possibilita a fácil interação com modelagens robustas, como o particionamento dos dados de entrada através de *bootstrapping*. Com o intuito de reduzir incertezas relacionados a um único modelo, definimos 20 modelos *Random-Forest*, dos quais foi realizado um *ensemble* – consenso. Como resultado obtivemos um mapa preditivo probabilístico caracterizado como a média ponderada entre os modelos, onde o AUC alcançado foi de 0,963. Ademais, a geração de curvas de resposta perante os pontos de ruptura nas encostas possibilitou uma análise mais profunda de como as diferentes variáveis utilizadas no modelamento de fato interagem e desenvolvem instabilidades na área de estudo.

**Palavras-chave:** Random-Forest; Mapeamento; Escorregamento; Estabilidade; Incerteza.

---

#### ABSTRACT

The present study sought to evaluate the effectiveness of the Random-Forest model to aid in mapping gravitational mass movement. In this work, the model was defined in R programming language with the help of the SDM package. The algorithm through this package proved to be easy to implement and enables easy interaction with robust modeling, such as partitioning the input data through bootstrapping. In order to reduce uncertainties related to a single model, we defined 20 Random-Forest models, from which an ensemble - consensus was performed. As a result, we obtained a probabilistic predictive map characterized as the weighted average among the models, where the AUC achieved was 0.963. In addition, the generation of response curves to the points of rupture in the slopes allowed a deeper analysis of how the different variables used in the modeling actually interact and develop instabilities in the study area.

**Keywords:** Random-Forest; Mapping; Landslide; Stability; Uncertainty.

---

---

<sup>1</sup> Universidade Federal de Santa Catarina, Florianópolis, Brasil.

\*E-mail: [gabriel\\_gguara@terra.com.br](mailto:gabriel_gguara@terra.com.br).

<sup>2</sup> Não possui vínculo no presente momento.

## INTRODUÇÃO

A identificação e o mapeamento das áreas com predisposição à ocorrência de processos naturais catastróficos, que resultam em danos e ameaça a sociedade, é uma demanda de elevada importância, principalmente pelo papel que exerce em ações de planejamento e gestão ambiental, territorial e de riscos (SILVEIRA *et al.*, 2014).

De acordo com Cerri e Nogueira (2014) nas áreas de assentamentos urbano precários há uma alta vulnerabilidade associada à forma ou localização inadequada da ocupação, pela ausência de infraestrutura urbana (drenagem, pavimentação, saneamento), de serviços básicos (coleta de lixo, redes elétrica e hidráulica, etc.) e pela degradação do ambiente. De acordo com os autores alguns tópicos são essenciais à gestão de riscos como a identificação, análise e mapeamento das situações de risco, em escala adequada; o planejamento de intervenções estruturais para redução ou erradicação dos riscos; o monitoramento permanente e prevenção de acidentes, especialmente nos períodos críticos; e informação pública e capacitação para autodefesa e prevenção.

Os métodos disponíveis para a identificação de áreas com predisposição a instabilidades são variados, como os determinísticos e estatísticos. Estes métodos geram como resultados mapas em formato matricial (*raster*) em que cada pixel indica uma probabilidade ou suscetibilidade aos movimentos gravitacionais de massa.

Métodos matemáticos e estatísticos para o mapeamento e classificação vem sendo experimentados e aplicados em diversas áreas do conhecimento e, com o avanço do aprendizado de máquinas, essas aplicações se tornam cada vez mais fáceis. São diversos os métodos, capazes de lidar com grandes volumes de dados. Alguns a serem mencionados são as Redes Neurais Artificiais (UEHARA *et al.*, 2020; YILMAZ, 2009; QUEVEDO *et al.*, 2019), Regressão Logística (YILMAZ, 2009; WANG; NIU, 2017), *Support Vector Machines* (UEHARA *et al.*, 2020; POURGHASEMI *et al.*, 2013) e *Random-Forest* (UEHARA *et al.*, 2020; DOU *et al.*, 2019; QUEVEDO *et al.*, 2019).

O presente trabalho aplica o método *Random-Forest* com base na linguagem de programação R para tentar definir áreas com alto risco (probabilidade) de ruptura das encostas através da ligação causa e efeito que as diferentes variáveis ambientais têm perante as encostas.

## MATERIAIS E MÉTODO

### SOFTWARE E FRAMEWORK

As análises e resultados que serão apresentados neste artigo foram gerados a partir de simulações realizadas com *scripts* próprios, escritos em linguagem R com auxílio da interface de desenvolvimento RStudio, versão 4.1.2.

Os modelos foram gerados utilizando a biblioteca SDM (NAIMI; ARAÚJO, 2016). O pacote SDM foi originalmente concebido para o modelamento da distribuição espacial de espécies, facilitando a aplicação de modelos matemáticos e estatísticos para a determinação de habitats potenciais. Da mesma forma que em mapeamentos de movimentos gravitacionais de massa, a modelagem da distribuição espacial das espécies se dá a partir de variáveis preditoras (independentes) e a presença ou ausência da determinada espécie em uma determinada área (variável dependente, a ser determinada pelos modelos).

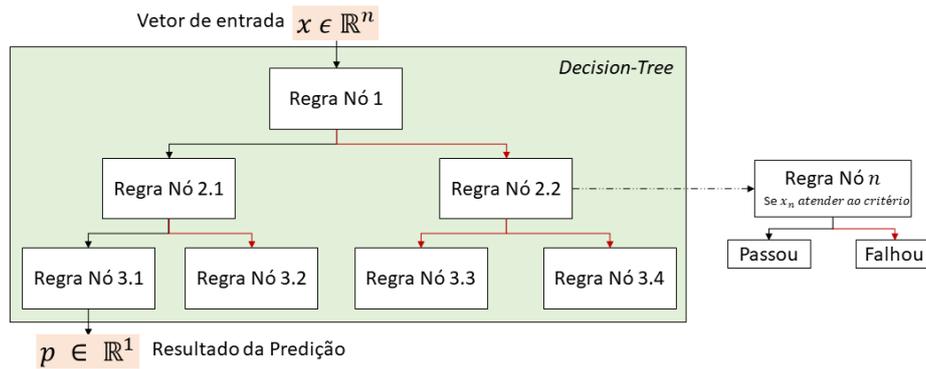
Os movimentos gravitacionais de massa podem ser tratados da mesma forma, relacionam-se diversas variáveis como a declividade, a altimetria ou a curvatura das encostas à presença ou ausência de movimentos gravitacionais de massa (dados de treinamento), relações estas que aferirão na previsão de áreas onde não ocorreram movimentações, mas que possuem potencial para tal.

### MÉTODO *RANDOM FOREST* (RF)

Modelos *Random-Forest* são algoritmos estatísticos de regressão e classificação caracterizados pela construção em *ensembles*, i.e., modelos que contém múltiplos preditores e seus resultados são combinados de acordo com uma métrica estatística, como média simples no caso de regressão ou moda no caso de classificação (BROWN, 2010). No caso de *Random-Forests*, os modelos são formados por uma combinação de preditores árvore, conhecidos como *Decision-Trees* (Figura 1), em que cada árvore otimiza sua predição de acordo com uma partição dos dados de treinamento, tornando a predição de cada árvore independente e não correlata em relação aos outros preditores (BREIMAN, 2001).

*Decision-Trees* utilizam um conjunto de regramentos binários para processamento e filtragem das entradas numéricas e categóricas (KOTSIANIS, 2013). Estes modelos são treinados de forma supervisionada empregando funções que avaliam a impureza informacional do vetor de entrada, como: Índice de Gini, Entropia de Informação e redução de variância (BREIMAN, 2001).

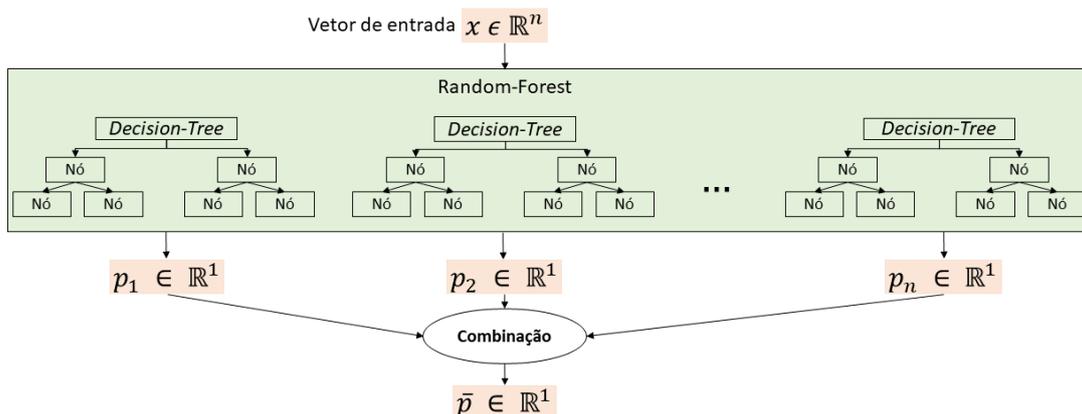
**Figura 1** – Esquema demonstrando o conceito de uma *Decision-Tree*.



Fonte: Dos autores (2022)

No entanto, para aprendizado em problemas que envolvem maior nível de complexidade, modelos *Decision-Trees* sofrem comumente de problemas de *overfit*, nos quais o modelo não atinge o nível desejado de generalização devido ao encontro de mínimos locais no processo de otimização. Isto se deve ao fato de que a performance da árvore depende extremamente da escolha dos seus hiperparâmetros estruturais, e.g. profundidade (BREIMAN, 2001). Como solução deste problema, surgem os modelos *Random-Forests* (Figura 2), que geram predições a partir da avaliação da predição de múltiplas *Decision-Trees*, treinadas com diferentes conjuntos de dados e que possuem diferentes estruturas. Isto gera maior acurácia na previsão final devido à Lei dos Grandes Números (BREIMAN, 2001).

**Figura 2** – Esquema demonstrando o conceito de um modelo *Random-Forest*.



Fonte: Dos autores (2022)

Breiman (2001) descreve o seguinte procedimento para o treinamento de modelos *Random-Forest*:

1. Selecionar  $n$  amostras *bootstrap*;
2. Dos  $p$  preditores existentes,  $n < p$  são selecionados aleatoriamente para treinamento;
3. O aprendizado de segmentação dos dados de entrada em cada nó é realizado através de uma função objetivo.

Sendo  $p(i|n)$  a probabilidade de os dados de entrada pertencerem à classe  $i$  no nó  $n$  de uma *Decision-Tree*, a otimização pode acontecer seguindo as seguintes funções objetivo:

$$Gini(t) = 1 - \sum [p(i|t)]^2$$

$$Entropia(t) = - \sum p(i|t) \log_2 p(i|t)$$

$$Erro(t) = 1 - \max [p(i|t)]$$

## VARIÁVEIS INDEPENDENTES

No estudo da instabilidade de encostas a característica principal que determina a estabilidade ou não é a resistência ao cisalhamento (TERZAGHI *et al.*, 1996), governada por dois componentes – a coesão e o ângulo de atrito interno.

As propriedades físicas do solo, como a resistência ao cisalhamento ou as suas propriedades hidrológicas podem ser determinadas em campo ou em laboratório, todavia é um processo trabalhoso, demorado e altamente incerto. Estas incertezas se dão por alguns aspectos, de acordo com Fiori (2016), as propriedades físicas do solo podem variar muito em função de condições externas, do tempo ou do método de análise.

A incerteza pode advir também da falta de tempo ou de recursos financeiros e até mesmo da inexperiência de quem está executando o ensaio (COROMINAS *et al.*, 2011). O autor ainda se refere a variabilidade espacial, a qual dificilmente poderá ser totalmente contornada, pois jamais será possível amostrar todos os possíveis pontos de uma determinada área de estudo, pois o solo se comporta de forma contínua.

A variabilidade espacial não é um processo aleatório, há alguma forma de correlação que controla e relaciona as propriedades dos solos com a sua localização no espaço (COSTA, 2005). Partindo-se deste pressuposto pode-se relacionar, portanto, variáveis proxy que influenciam positiva ou negativamente nestas propriedades altamente incertas.

Ao todo foram consideradas inicialmente 17 variáveis que potencialmente poderiam explicar a distribuição espacial das cicatrizes existentes na área de estudo,

destas, 11 foram selecionadas para compor o modelo (Quadro 1), de acordo principalmente com a sua importância relativa, mas também em relação à correlação entre as variáveis. As variáveis foram construídas através de dois planos de informação preliminares, o MDT e a ortofotografia da Secretaria de Estado do Desenvolvimento Econômico Sustentável de Santa Catarina, desenvolvidos em 2010 (acesso em: <http://sigsc.sc.gov.br/>)

Ressaltamos a importância da geologia e pedologia nos estudos de estabilidade de encostas, neste trabalho, denominado unidades geotécnicas. Estas unidades refletem a união da geologia com a pedologia e indicam comportamentos geomecânicos. Todavia esta variável não tem grande importância estatística nesta área de estudo especificamente pois, como um todo, é relativamente homogênea – Geologia: Arenitos.

**Quadro 1** – Lista das variáveis consideradas no modelamento.

Variável	Descrição	Variável	Descrição
<b>Aspecto</b>	Indiretamente considera algumas condições microclimáticas tais como exposição solar, ventos úmidos ou secos e intensidade da chuva. Todos estes fatores controlam o aspecto da encosta e causam diferenças na umidade do solo, o que pode levar ao rompimento da mesma. (SHIRANI et al. 2018).	<b>Convexidade</b>	Calculado no software SAGA Gis, V.7.8.2. A variável representa a curvatura positiva da superfície, cada valor de célula da matriz (raster) representa a porcentagem de células convexas para cima dentro de um raio constante de dez células (IWAHASHI; PIKE, 2007).
<b>Distância das Ruas</b>	Os cortes de estrada removem a massa estabilizadora na base das encostas e, quando construídos nas elevações das encostas, alteram a geometria das delas e podem causar instabilidades.	<b>Área de captação</b>	Esta variável é a contagem cumulativa do número de pixels que drenam naturalmente para fora do sistema e mede a área de uma bacia hidrográfica que contribui para o escoamento de cada pixel (CANAVESI et al., 2020). Em outras palavras, indica áreas no terreno onde a tendência é concentrar a água após a precipitação.
<b>Declividade</b>	O peso dos materiais faz com que uma massa instável de solo ou rocha se mova em favor da gravidade pela encosta abaixo. Ao aumentar a declividade, as forças de cisalhamento também aumentam (YILMAZ 2009).	<b>Distância das Falhas</b>	Lineamentos (falhas) são estruturas de deformação de rochas onde há deslocamento, especialmente em áreas inclinadas, de rochas ou massas de solo. Plano de informação construído através da análise do relevo sombreado.
<b>TRI*</b>	Variável morfológica que indica a rugosidade do terreno (WANG et al., 2017).	<b>Distância das Drenagens</b>	É uma variável diretamente relacionada à estabilidade das encostas, proximidade age saturando o solo e erodindo a base das encostas (WANG et al. 2016). A drenagem também atua no aumento das poropressões, tornando o estado de estresse efetivo do solo mais próximo das condições de ruptura (Silveira 2016).

### Continuação do Quadro 1:

<b>Profundidade do Solo</b>	Método adotado: MOORE et al. (1993). As camadas mais espessas do solo tendem a se concentrar em locais planos, enquanto as camadas mais finas tendem a se concentrar em declividades mais acentuadas, portanto, apresentando maior potencial de erosão.	<b>Altimetria</b>	Os pesos das camadas superiores do solo e das rochas atuam como forças cisalhantes, e quando as forças resistentes não as contrapõe, há a possibilidade da encosta romper.
<b>TWI**</b>	Variável topográfica hidrológica que indica a tendência da água a se acumular em qualquer ponto da área de captação, bem como a tendência da gravidade a forçar a água a descer a encosta (POURGHASEMI et al., 2012).	-	-

\* Índice topográfico de rugosidade; \*\* Índice topográfico de umidade

### DADOS DE TREINAMENTO E TESTE

Na área de estudo há um registro por parte da Defesa Civil do município de Blumenau dos movimentos gravitacionais de massa, o *dataset* apresenta 100 polígonos onde ocorreram movimentações do terreno. Nestes, a equipe técnica da defesa civil avaliou algumas características do terreno, como por exemplo a profundidade do solo e gerou um relatório técnico de uso interno (BLUMENAU, 2010). Esta informação foi utilizada na verificação da coerência do mapa de profundidade dos solos.

Vale ressaltar que não consideramos a área dos polígonos como um todo para o treinamento ou validação do modelo, um movimento de massa é essencialmente dividido em três partes, a zona de ruptura, de transporte do material e de acumulação. Cada fase possui um controle geomorfológico diferente (CHUNG; FABBRI, 2003). Considerar a área completa de uma cicatriz no modelamento pode levar a resultados errôneos e isto se deve, pois, a desestabilização de apenas uma pequena área, representada por poucas células instáveis, pode acarretar na desestabilização de um volume de solo muito maior devido ao relaxamento das tensões atuantes no maciço, fazendo com que um grande volume de solo presente nos arredores desta área também se desestabilize (MICHEL *et al.*, 2012). Desta forma, cada cicatriz foi dividida em uma área de ruptura e outra de transporte e acumulação. Da área de ruptura foram extraídos os seus centroides que posteriormente foram relacionados às variáveis preditoras (independentes).

Além destes 100 locais onde ocorreram movimentações de massa foram também definidos 100 locais estáveis dentro da área de estudo. Estes 100 pontos estáveis foram definidos através da ortofotografia realizada pela Secretaria de Estado do

Desenvolvimento Econômico Sustentável de Santa Catarina em 2010, deste *dataset* final com 200 pontos foram separados dados de treinamento e teste randomicamente, em uma proporção de 70:30, respectivamente.

Além da separação dos dados utilizamos a técnica *bootstrap* com o fim de contornar incertezas envolvidas com os dados de entrada no modelo. *Bootstrap* é um método estatístico onde a amostra inicial é re-amostrada e inferências estatísticas podem ser derivadas. São duas as possibilidades de aplicação do método, paramétrico e não paramétrico, sendo o primeiro aplicado quando a amostra é pequena e a distribuição de probabilidade da amostra é conhecida. Já o segundo, utilizado neste trabalho, baseia-se na amostragem com substituição da amostra original (THASNEEM *et al.*, 2021).

### ***RANDOM FOREST ENSEMBLE***

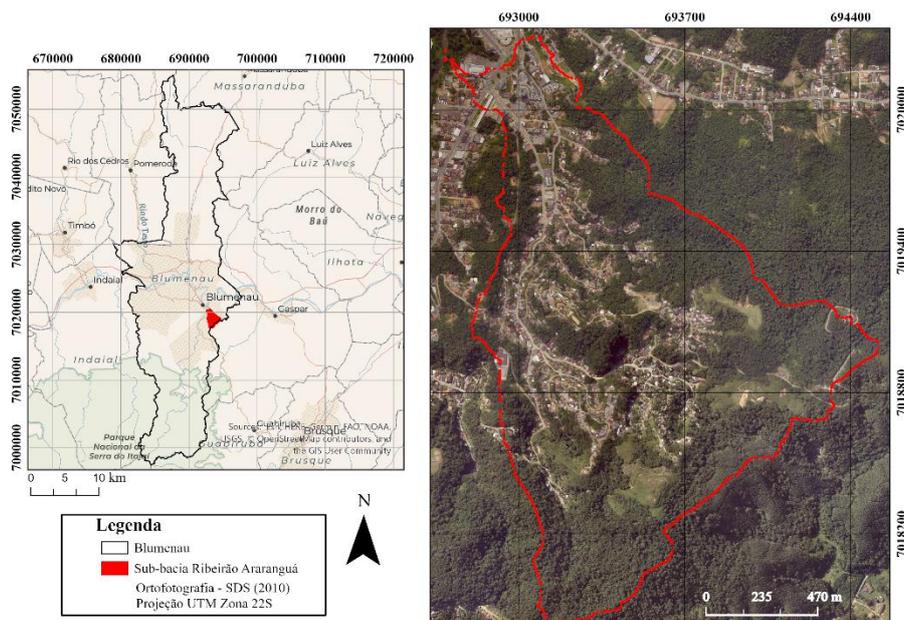
Como mencionado, RF é uma combinação de árvores de decisões. Todavia, apesar de evitar incertezas relacionadas à uma única árvore de decisão, cada modelo irá se diferenciar de outro caso, por exemplo, os dados de treino e teste mudem. Além disto, em cada etapa de *bootstrap* os dados serão relativamente diferentes. Desta forma, com o fim de contornar incertezas quanto ao modelo, geramos 20 RF's onde cada modelo diferenciou-se do outro devido ao exposto.

*Ensambling* é um método no qual gera-se um modelo consensual entre todas as simulações individuais. O pacote SDM disponibiliza diversos métodos para tal, neste caso, aplicamos o método “média-ponderada”, definido como a média de duas etapas, que é quando várias replicações são ajustadas para cada método de modelagem (*bootstrap*), usando este método uma média não ponderada é tomada sobre os valores previstos de diferentes replicações de cada método (ou seja, dentro da média do modelo), então uma média ponderada é usada para combiná-las em valores finais do conjunto (NAIMI; ARAÚJO, 2016).

### **ÁREA DE ESTUDO**

A área de estudo, denominada Sub-bacia do Ribeirão Araranguá (Figura 3) está localizada no município de Blumenau, estado de Santa Catarina, Brasil. A localidade situa-se na porção centro sul do município, caracterizada por relevos acidentados e, em relação ao seu tamanho, é uma das áreas que mais sofrem com movimentos gravitacionais de massa no município.

**Figura 3** – Localização da Sub-bacia do ribeirão Araranguá.



O relatório técnico de Blumenau (2010) caracteriza a área de estudo geomorfologicamente como sendo composta por vales em “V” fechados com rede de drenagem multidirecional ordenada, obedecendo ao padrão retangular-dentrítico.

Diversos fatores condicionantes e potencializadores, tanto naturais quanto antrópicos, foram delineados por Vieira (2004) na área de estudo, a serem citados:

- Condicionadores:
  1. Geologia - Falhas, fraturas, bandamentos, foliações e tipo de rocha.
  2. Pedologia - Manto de intemperismo (granulometria, porosidade, permeabilidade, resistência ao cisalhamento).
  3. Geomorfologia - Forte dissecação e forma da encosta (côncava, retilínea, convexa).
  4. Declividade - Média a alta (30% a 143%).
  5. Clima - Água de chuvas distribuídas (saturação do solo) e concentradas, observando-se a quantidade, intensidade, duração e distribuição nos setores da encosta.
- Potencializadores:
  1. Uso e ocupação - Desestabilização das encostas através de cortes e aterros; lançamento de águas pluviais e servidas; vazamentos na rede de água.
  2. Lixo - Deposição de entulhos e lixo a céu aberto.
  3. Cobertura vegetal - Retirada da cobertura e plantio de espécies que potencializam os deslizamentos.

Quanto a geologia local, trata-se de uma variedade monótona, exclusivamente sedimentar. Em sua grande maioria, caracterizada como arenitos médios, por vezes arenitos finos e grosseiros também podem ser encontrados, além de siltes.

A região é uma plataforma dobrada e é representada por sedimentos do Grupo Itajaí, além de depósitos de sedimentos continentais quaternários (BLUMENAU, 2010). O grupo Itajaí é representado pela formação Gaspar, que por sua vez é representado por arenitos feldspáticos de cor bordô com estratificação plano paralela e cruzada (VIEIRA, 2004)

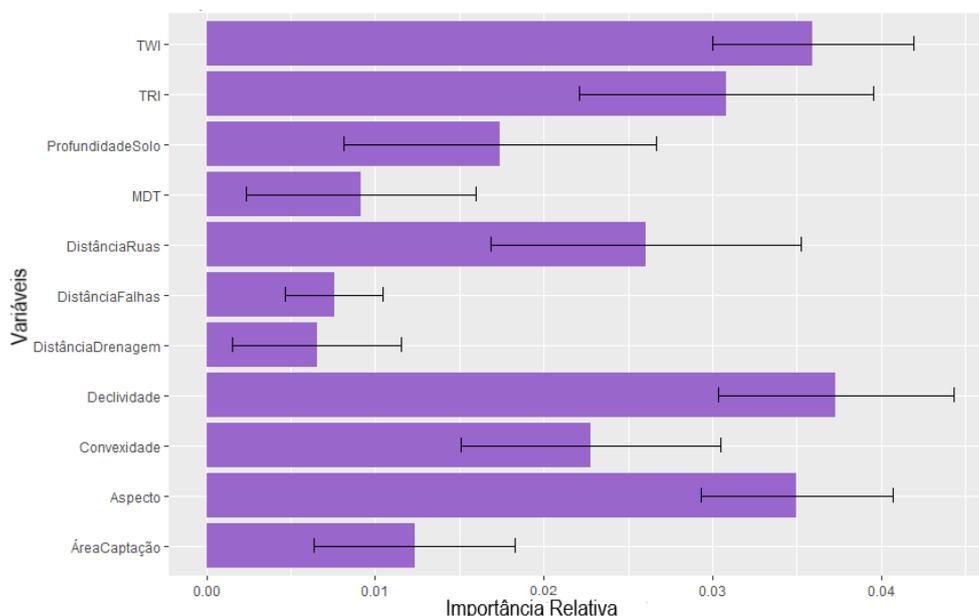
## RESULTADOS E DISCUSSÕES

### SELEÇÃO DAS VARIÁVEIS

As variáveis preditoras são peças fundamentais no modelamento RF, portanto, faz-se necessário testar dentre todas as consideradas, quais mais impactam positivamente o poder preditivo do modelo. Para tanto, rodam-se modelos RF's, verifica-se a importância das variáveis, retira-se a menos importante e repete-se o processo até restarem apenas as mais importantes.

A Figura 4 apresenta a importância das variáveis após o processo descrito anteriormente, as barras de variação indicam a influência das variáveis em todas os 20 modelos. Definimos como aceitável qualquer variável com importância superior a 0.01.

**Figura 4** – Importância das variáveis consideradas.

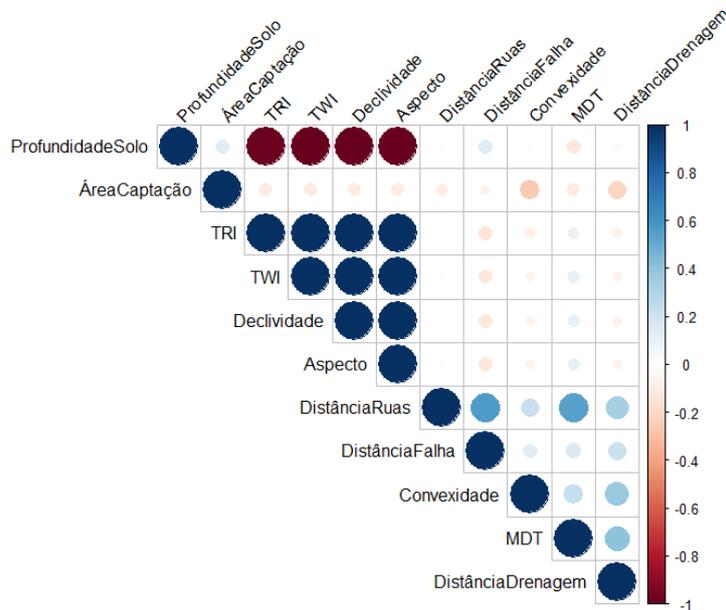


Verifica-se que três são as fontes principais de instabilidades na área de estudo, a hidrologia, a topografia e a influência antrópica. O quanto afeta a hidrologia na estabilidade das encostas é demonstrada através das variáveis TWI, distância das

drenagens e área de captação, indiretamente o aspecto das encostas também se enquadra neste grupo. O aspecto topográfico é relacionado com as variáveis declividade, TRI, profundidade do solo, MDT e convexidade e distância das falhas, este é o aspecto mais influente na estabilidade global da área. Por fim, o fator antrópico também parece ser decisivo, a vasta maioria das rupturas que ocorreram na área de estudo estão associadas a proximidade de ruas e cortes para a passagem destas.

*Random-Forests* são sensíveis à correlação entre variáveis pois impacta a habilidade do modelo em identificar as melhores variáveis preditoras (DARST et al., 2018). Com o fim de verificar a correlação entre as variáveis utilizadas foi construído um gráfico de correlação apresentado na Figura 5.

**Figura 5** – Correlação entre as variáveis selecionadas a compor o modelo.

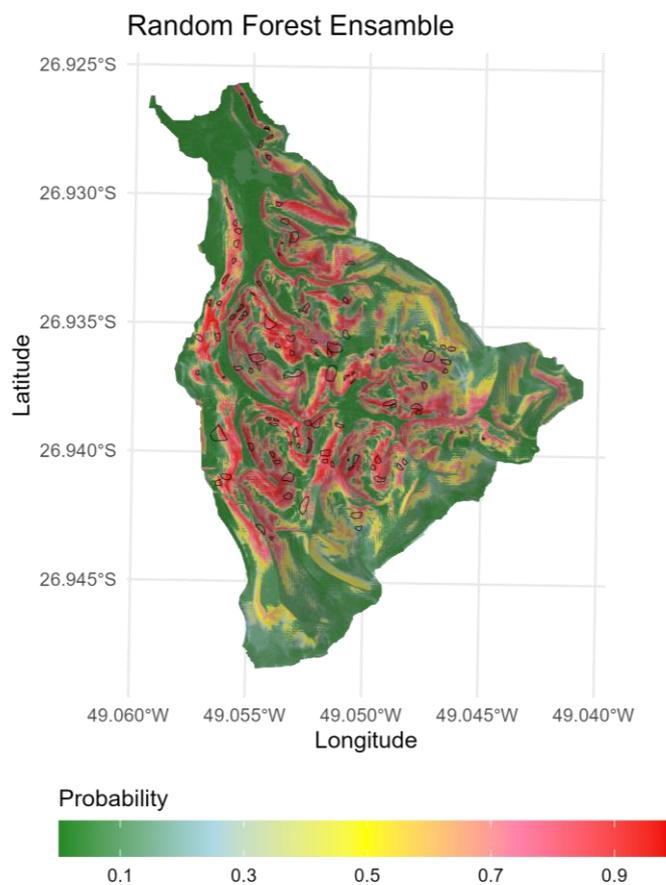


Apesar das variáveis TRI, TWI, declividade e Aspecto apresentarem alta correlação, ainda são as variáveis mais importante para o modelo, as restantes variáveis não apresentaram correlações elevadas.

### MAPA PREDITIVO

Como mencionado, foram gerados 20 modelos RF, os quais passaram por um processo de *ensembling*. Após tal processo foi gerado o mapeamento apresentado na Figura 6, onde a probabilidade máxima de ruptura foi de 0,994 e mínima de 0,00185. Os polígonos em preto são às áreas de ruptura dos movimentos gravitacionais de massa.

**Figura 6** – Resultado final: Mapa *Random-Forest*.



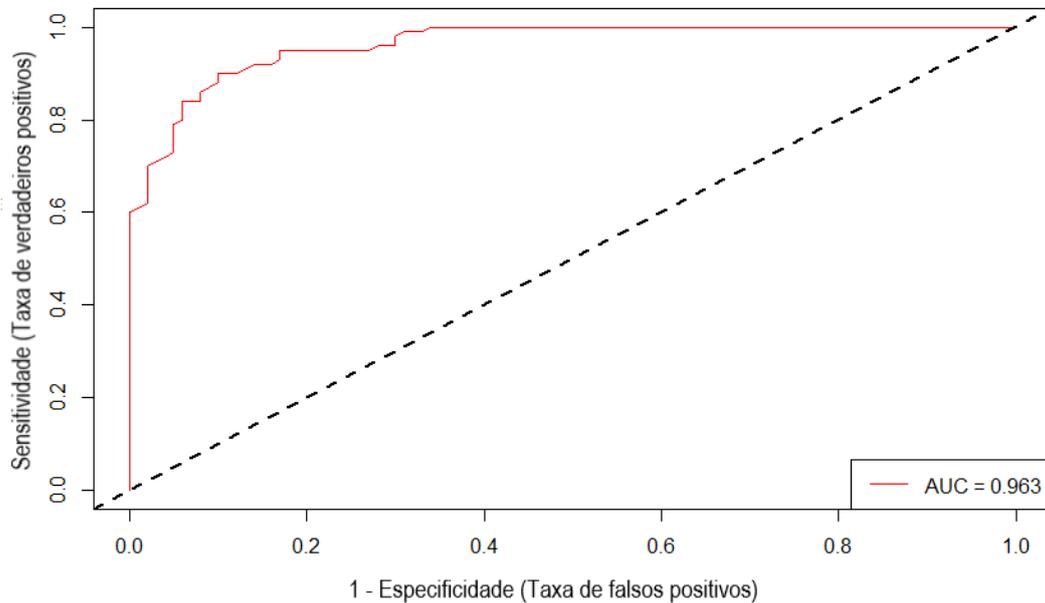
Visualmente o modelo conseguiu classificar com altas probabilidades locais onde de fato ocorreram movimentações de massa. Com o fim de averiguar e atestar esta acurácia foi aplicado o método de validação ROC/AUC.

A curva ROC compara cada pixel da análise com as cicatrizes visualizadas em campo. Se um pixel computado instável estiver dentro do perímetro de uma cicatriz, será considerado “verdadeiro positivo” (*vp*), se estiver fora, “falso positivo” (*fp*). Da mesma forma, se um pixel computado estável estiver dentro do perímetro de uma cicatriz, será considerado “falso negativo” (*fn*) e, se estiver fora, “verdadeiro negativo” (*vn*) (ZIZIOLI, 2013; FAWCETT, 2006). Segundo Gorsevski *et al.* (2006), a curva é a plotagem da probabilidade de se ter um *vp* versus a probabilidade de *fp*, onde o limiar de probabilidade aceitável varia. Segundo Zizioli (2013) a realização da análise se dá a partir do cálculo de dois parâmetros: Sensitividade (razão entre *vp* e *vp + fn*), no eixo y e especificidade (razão entre *vn* e *vn + fp*), no eixo x. Uma boa curva é aquela em que a mesma se aproxima do canto superior esquerdo do gráfico (FRATTINI *et al.*, 2010).

A “Area Under the ROC Curve” - AUC representa a taxa de sucesso ou acurácia global do modelo e é representada pela área abaixo da curva ROC (ZIZIOLI, 2013). Na

Figura 7 está apresentado o gráfico da curva ROC para o mapeamento exposto na Figura 6, onde alcançou-se um AUC de 0,963.

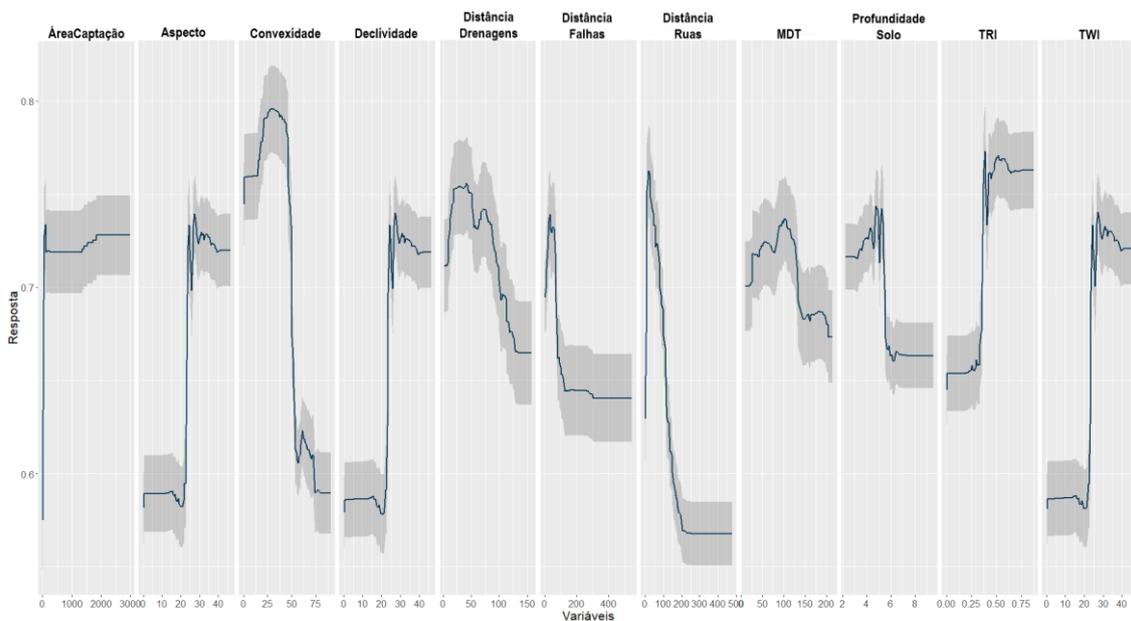
**Figura 7** – Validação do modelo através da curva ROC e AUC.



## RESPOSTA UNITÁRIA DAS VARIÁVEIS

Com o fim de averiguar a resposta individual das variáveis perante as cicatrizes foi construído gráficos de respostas das variáveis, os quais elucidaram de forma clara e objetiva como as diferentes variáveis se comportaram no decorrer da simulação (Figura 8). As áreas hachuradas acinzentadas representam a flutuação das respostas nos 20 modelos e as linhas centrais a resposta média.

**Figura 8** – Curvas de resposta dos 20 modelos RF.



A partir dos gráficos apresentados na Figura 8 fica clara a contribuição de cada variável. A área de captação indica que a tendência preferencial de haver uma probabilidade alta de ruptura seja próxima das áreas de drenagem, quando a contagem cumulativa do número de pixels que drenam naturalmente para fora do sistema é maior que 500. A distância das drenagens e TWI corroboram tal pressuposto, as maiores respostas estão em localidades a menos de 100m de uma rede de drenagem segundo a primeira e, de acordo com a variável TWI, observa-se claramente o domínio de valores mais elevados, indicando acúmulo de água no solo.

A partir da variável convexidade observa-se que os valores mais elevados, indicando maior convexidade do terreno, não tiveram grande influência nas predições, formas côncavas estão associadas a maior acúmulo de água, o que se relaciona as três variáveis anteriores.

A declividade é um dos controles de estabilidade mais importantes, fica clara a dominância de declividades superiores a 30° na classificação de probabilidades altas de ruptura. É interessante notar como as respostas entre as variáveis aspecto, declividade e TWI são parecidas, as duas últimas são relacionadas, a declividade é utilizada no equacionamento da variável TWI. Entretanto o aspecto não, portanto, com estes resultados corroboramos o mencionado por Shirani et al. (2018) em que esta variável está indiretamente ligada com fatores microclimáticos. Haja vista a influência clara de valores elevados dessa variável, presume-se que sejam faces das encostas que não estejam voltadas para o sol na maior parte do tempo e, portanto, devem apresentar maior umidade no solo.

No sul do município de Blumenau, onde a área de estudo está inserida, a tectônica, o sistema de fraturas, a morfologia do terreno, somado ao avançado estado de alteração das rochas tornam algumas áreas inadequadas a ocupação humana (AUMOND, 2009). A variável Distância das falhas demonstra que há uma forte influência da proximidade com estas falhas na caracterização de probabilidades altas de rompimento das encostas, as maiores influências se dão para distâncias inferiores a 100m de alguma falha. Quanto a variável TRI, variável considerada morfológica, pois indica a rugosidade do terreno, demonstra que valores elevados, ou seja, terrenos mais rugosos, contribuem mais para a instabilidade das encostas.

Em se tratando de profundidade dos solos podemos correlacionar o apresentado no gráfico com o exposto no relatório técnico de Blumenau (2010). Como já mencionado

a Defesa Civil do município observou e catalogou diversas características da área de estudo em relação às movimentações gravitacionais de massa, como se houve ou não danos a equipamentos públicos ou privados, a geologia e também a profundidade do solo. Quanto às profundidades foi constatado a dominância de profundidades próximas a 5m, o que também foi observado no decorrer das simulações.

Por fim, observa-se a contribuição antrópica na instabilidade das encostas da área de estudo através da variável Distância das ruas. A sub-bacia do ribeirão Araranguá é uma área de exclusão social no município de Blumenau e também caracterizada como área de risco. Portanto, não há grandes investimentos por parte do município na localidade. Várias ruas são abertas sem cuidados perante a geotecnia, por exemplo executando cortes de estradas, além disto, a sobrecarga dos veículos pode agir como um acréscimo das forças atuantes desestabilizando as encostas. Observamos através de fotos aéreas, também, a prática de jogar lixo ao lado destas ruas, que se encontram quase sempre inseridas nas encostas. Está pratica age de duas formas principais na elevação da potencialidade de instabilidade, a primeira podendo gerar sobrecarga e a segunda, impedindo ou reduzindo a permeabilidade natural do solo. Portanto através desta variável observa-se que distâncias inferiores a 50m das ruas são os locais onde há a maior propensão de haver algum movimento gravitacional de massa no futuro.

## CONCLUSÕES

O modelo *Random-Forest* se mostrou altamente eficiente na modelagem e predição de instabilidades das encostas, como pode ser observado através do AUC. Apesar do pacote SDM ter sido feito com o intuito de modelar a distribuição espacial de habitats de espécies animais, este pode ser usado para a modelagem da distribuição espacial de movimentos gravitacionais de massa.

Haja vista a dificuldade de aquisição de dados e ainda mais a representatividade dos dados referentes a resistência ao cisalhamento dos solos para grandes áreas, métodos estatísticos como o *Random-Forest* se mostram potenciais grandes aliados para o mapeamento de áreas de risco, podendo agir como métodos preliminares de setorização destas áreas e guiar localidades que necessitem análises mais robustas e localizadas.

A importância das variáveis é bastante relativa de localidade para localidade, não existe um consenso em relação a quais devem ser empregadas ao se modelar a estabilidade de encostas, portanto, a métrica “importância” do RF permite aferir com maior certeza quais variáveis de fato explicam o fenômeno a ser modelado. Além da

importância das variáveis as curvas de resposta demonstram de forma clara como cada variável influencia o desenvolvimento das RF's possibilitando correlações importantes e verificações subsequentes de coerência do modelo com a teoria da estabilidade de encostas.

## REFERÊNCIAS

- AUMOND, J. J *et al.*; Condições naturais que tornam o vale do Itajaí sujeito aos desastres. 2009. In: FRANK, B.; SEVEGNANI, L. (Orgs.). Desastre de 2008 no Vale do Itajaí: água, gente e política. Blumenau: Agência de Água do Vale do Itajaí.
- BLUMENAU. **Carta de Uso Recomendado do Solo (Área 09) Região Rua Araranguá e Transversais**. Interno, Sub-bacia do Ribeirão Araranguá: Secretaria Municipal de Planejamento Urbano: Diretoria Geologia., 2010.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>.
- BROWN, G. Ensemble Learning. In: C. Sammut; G. I. Webb (Orgs.); **Encyclopedia of Machine Learning**. p.312–320, 2010. Boston, MA: Springer US. Disponível em: <[https://doi.org/10.1007/978-0-387-30164-8\\_252](https://doi.org/10.1007/978-0-387-30164-8_252)>.
- CANAVESI, V.; SEGONI, S.; ROSI, A.; et al. Different Approaches to Use Morphometric Attributes in Landslide Susceptibility Mapping Based on Meso-Scale Spatial Units: A Case Study in Rio de Janeiro (Brazil). **Remote Sensing**, v. 12, n. 11, p. 1826, 2020. Multidisciplinary Digital Publishing Institute. Disponível em: <<https://doi.org/10.3390/rs12111826>>.
- CHUNG, C.J.; FABBRI, A. Validation of Spatial Prediction Models for Landslide Hazard Mapping. **Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards**, v. 30, n. 3, p. 451–472, 2003. Springer & International Society for the Prevention and Mitigation of Natural Hazards.
- COROMINAS, J.; MAVROULI, O.C. Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies (No. 226479). SafeLand, 2011.
- COSTA, E. A. DA. **Avaliação de ameaças e risco geotécnico aplicados à estabilidade de taludes**, 2005. Dissertação, Porto Alegre: Universidade Federal do Rio Grande do Sul. Disponível em: <<https://lume.ufrgs.br/handle/10183/6150>>.
- DARST, B. F.; MALECKI, K. C.; ENGELMAN, C. D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. **BMC Genetics**, v. 19, n. 1, p. 65, 2018. Disponível em: <<https://doi.org/10.1186/s12863-018-0633-8>>.

DOU, J.; YUNUS, A. P.; TIEN BUI, D.; et al. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. **The Science of the Total Environment**, v. 662, p. 332–346, 2019. Disponível em: <<https://doi.org/10.1016/j.scitotenv.2019.01.221>>.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters, ROC Analysis in Pattern Recognition.**, v. 27, n. 8, p. 861–874, 2006. Disponível em: <<https://doi.org/10.1016/j.patrec.2005.10.010>>.

FIORI, A. P. **Fundamentos de mecânica dos solos e das rochas**. Oficina de Textos, 2016.

FRATTINI, P.; CROSTA, G.; CARRARA, A. Techniques for evaluating the performance of landslide susceptibility models. **Engineering Geology**, v. 111, n. 1–4, p. 62, 2010.

FURTADO, S. M. DE A. **Um olhar sobre a paisagem e o lugar como expressão do comportamento frente ao risco de deslizamento**, 2004. Phd, Florianópolis: Universidade Federal de Santa Catarina. Disponível em: <<https://repositorio.ufsc.br/xmlui/handle/123456789/87294>>.

GORSEVSKI, P.; GESSLER, P.; FOLTZ, R.; ELLIOT, W. Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis. **T. GIS**, v. 10, p. 395–415, 2006. Disponível em: <<https://doi.org/10.1111/j.1467-9671.2006.01004.x>>.

IWAHASHI, J.; PIKE, R. J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. **Geomorphology**, v. 86, n. 3–4, p. 409–440, 2007. Disponível em: <<https://doi.org/10.1016/j.geomorph.2006.09.012>>.

KOTSIANTIS, S. B. Decision trees: a recent overview. **Artificial Intelligence Review**, v. 39, n. 4, p. 261–283, 2013. Disponível em: <<https://doi.org/10.1007/s10462-011-9272-4>>.

MICHEL, G. P.; KOBIYAMA, M.; GOERL, R. F. Análise comparativa entre os modelos Shalstab e Sinmap na identificação de áreas susceptíveis a escorregamentos translacionais. In: POLETO, C.; PLETSCH, A. L.; MELLO, E. E.; CARVALHO, N. O. (orgs) X ENES Encontro Nacional de Engenharia de Sedimentos: artigos selecionados. Porto Alegre: ABRH, 2012. p 209 - 225.

MOORE, I. D.; GESSLER, P. E.; NIELSEN, G. A.; PETERSON, G. A. Soil Attribute Prediction Using Terrain Analysis. **Soil Science Society of America Journal**, v. 57, n. 2, p. 443–452, 1993. Disponível em: <<https://doi.org/10.2136/sssaj1993.03615995005700020026x>>.

NAIMI, B.; ARAÚJO, M. B. sdm: a reproducible and extensible R platform for species distribution modelling. **Ecography**, v. 39, n. 4, p. 368–375, 2016. Disponível em: <<https://doi.org/10.1111/ecog.01881>>.

POURGHASEMI, H. R.; JIRANDEH, A. G.; PRADHAN, B.; XU, C.; GOKCEOGLU, C. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. **Journal of Earth System Science**, v. 122, n. 2, p. 349–369, 2013. Disponível em: <<https://doi.org/10.1007/s12040-013-0282-2>>.

POURGHASEMI, H. R.; PRADHAN, B.; GOKCEOGLU, C. Remote Sensing Data Derived Parameters and its Use in Landslide Susceptibility Assessment Using Shannon's Entropy and GIS. **Applied Mechanics and Materials**, v. 225, p. 486–491, 2012. Trans Tech Publications Ltd. Disponível em: <<https://doi.org/10.4028/www.scientific.net/AMM.225.486>>.

QUEVEDO, R. P.; GUASSELLI, L. A.; OLIVEIRA, G. G. D.; RUIZ, L. F. C. MODELAGEM DE ÁREAS SUSCETÍVEIS A MOVIMENTOS DE MASSA: AVALIAÇÃO COMPARATIVA DE TÉCNICAS DE AMOSTRAGEM, APRENDIZADO DE MÁQUINA E MODELOS DIGITAIS DE ELEVAÇÃO. **Geosciences = Geociências**, v. 38, n. 3, p. 781–795, 2019. Disponível em: <<https://doi.org/10.5016/geociencias.v38i3.14019>>.

THASNEEM, S.A; N. R. CHITHRA; SANTOSH G. THAMPI. Assessment of Nonstationarity and Uncertainty in Precipitation Extremes of a River Basin Under Climate Change. **Environmental modeling and assessment**, v. 26, n. 3, p. 295–312, 2021. Disponível em: <<https://doi.org/10.1007/s10666-021-09752-y>>.

SHIRANI, K.; PASANDI, M.; ARABAMERI, A. Landslide susceptibility assessment by Dempster–Shafer and Index of Entropy models, Sarkhoun basin, Southwestern Iran. **Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards**, v. 93, n. 3, p. 1379–1418, 2018. Springer & International Society for the Prevention and Mitigation of Natural Hazards.

SILVEIRA, C. T. DA; FIORI, A. P.; 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. Disponível em: <<https://doi.org/10.20502/rbg.v15i1.366>>.

SILVEIRA R. M da.; Comportamento Geotécnico das Encostas. 2016. Disponível em: <<https://www.crea-pr.org.br/ws/wp-content/uploads/2016/12/comportamento-geotecnico-das-encostas.pdf>>

TERZAGHI, K.; PECK, R. B.; MESRI, G. **Soil Mechanics in Engineering Practice**. John Wiley & Sons, 1996.

UEHARA, T. D. T.; CORRÊA, S. P. L. P.; QUEVEDO, R. P.; et al. Landslide Scars Detection using Remote Sensing and Pattern Recognition Techniques: Comparison Among Artificial Neural Networks, Gaussian Maximum Likelihood, Random Forest, and Support Vector Machine Classifiers. **Revista Brasileira de Cartografia**, v. 72, n. 4, p. 665–680, 2020. Disponível em: <<https://doi.org/10.14393/rbcv72n4-54037>>.

WANG, Q.; WANG, Y.; NIU, R.; PENG, L. Integration of Information Theory, K-Means Cluster Analysis and the Logistic Regression Model for Landslide Susceptibility

Mapping in the Three Gorges Area, China. **Remote Sensing**, v. 9, p. 938, 2017.  
Disponível em: < [https://doi.org/ 10.3390/rs9090938](https://doi.org/10.3390/rs9090938)>.

YILMAZ, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). **Computers & Geosciences**, v. 35, n. 6, p. 1125–1138, 2009.  
Disponível em: < [https://doi.org/ 10.1016/j.cageo.2008.08.007](https://doi.org/10.1016/j.cageo.2008.08.007)>.

ZIZIOLI, D.; MEISINA, C.; VALENTINO, R.; MONTRASIO, L. Comparison between different approaches to modeling shallow landslide susceptibility: a case history in Oltrepo Pavese, Northern Italy. **Natural Hazards and Earth System Sciences**, v. 13, n. 3, p. 559–573, 2013. Copernicus GmbH. Disponível em: < [https://doi.org/ 10.5194/nhess-13-559-2013](https://doi.org/10.5194/nhess-13-559-2013)>.

*Recebido em: 12/09/2022*

*Aprovado em: 15/10/2022*

*Publicado em: 19/10/2022*