

MAPEAMENTO DE SUSCETIBILIDADE A DESLIZAMENTOS A PARTIR DE MINERAÇÃO DE DADOS E DO MODELO SHALSTAB

Mapping of Susceptibility the Landslides from Data Mining and Shalstab Model

**Eduardo Samuel Riffel¹, Luis Fernando Chimelo Ruiz²
& Laurindo Antonio Guasselli^{1,2}**

**¹Universidade Federal do Rio Grande do Sul – UFRGS
Programa de Pós-Graduação em Geografia**

Avenida Bento Gonçalves, 9500, Agronomia, 91540-000, Porto Alegre, RS, Brasil
edriffel@gmail.com, laurindo.guasselli@ufrgs.com.br

**²Universidade Federal do Rio Grande do Sul – UFRGS
Programa de Pós-Graduação em Sensoriamento Remoto**

Avenida Bento Gonçalves, 9500, Agronomia, 91540-000, Porto Alegre, RS, Brasil
ruiz.ch@gmail.com

*Recebido em 26 de Julho, 2016/ Aceito em 28 de Outubro, 2016
Received on July 26, 2016/ Accepted on October 28, 2016*

RESUMO

Diversos modelos têm sido utilizados para mapear, modelar, quantificar e identificar áreas suscetíveis a deslizamentos. Os mapeamentos de suscetibilidade surgem como uma alternativa à cultura forense existente no Brasil em relação aos movimentos de massa, pois consistem em uma previsão de áreas onde é necessária maior atenção em relação à possibilidade de ocorrência de desastres. Assim, o objetivo do presente estudo foi realizar o mapeamento de suscetibilidade de movimentos de massa na bacia hidrográfica do rio Forromeco-RS. Para isso empregou-se duas metodologias, o modelo Shalstab e o modelo Árvore de Decisão, que posteriormente foram comparadas. Para implementar os modelos foram utilizadas 47 amostras de deslizamentos, identificadas a partir de cicatrizes de escorregamentos na região da bacia, mapeados por meio de imagens de satélite de alta resolução espacial. Como dado de entrada dos modelos foram utilizados parâmetros morfométricos, extraídos de imagem SRTM. A comparação dos resultados das duas metodologias e a análise da acurácia dos modelos, baseado no índice kappa, mostrou que o modelo Árvore de Decisão apresentou uma acurácia de 0,98, enquanto para o modelo Shalstab foi de 0,82. Ressalta-se que a diferença entre os resultados dos modelos é pouco significativa. Ambos podem representar de forma satisfatória o objetivo final de geração de mapa de suscetibilidade. No entanto, o modelo Shalstab apresenta mais limitações, devido à necessidade de dados de grande resolução, para uma boa resposta da análise.

Palavras chave: Deslizamentos, Shalstab, Árvore de Decisão.

ABSTRACT

Several models have been used to map, model, quantify and identify susceptible areas to landslides. The susceptibility mappings arose as an alternative to the forensic culture existing in Brazil in relation to mass movements because consist in a forecast of areas where it is needed most attention on the possibility the occurrence of disaster. So, the objective of this article was to accomplish the susceptibility mapping the mass movements in the basin of Forromeco river - RS. For this, two methodologies were used, the Shalstab model and the Decision Tree model, which posteriorly were

compared. To implement the model were used 47 samples of landslides, identified from landslide scars in the region of the basin, mapped by high spatial resolution satellite images. As input data of the models were used morphometric parameters extracted from the SRTM image. The comparison the results of the two methodologies and analysis of the accuracy of the models, based on the kappa index, showed that the Decision Tree model presented an accuracy of 0.98, while for the Shalstab model was 0.82. It is noteworthy that the difference between the results of the models is small. Satisfactorily, both can represent the ultimate goal of generation the susceptibility map. However, the Shalstab model has others limitations due to the need of high resolution data, for a good answer the analysis.

Keywords: Landslides, Shalstab, Decision Tree Model.

1. INTRODUÇÃO

Os desastres constituem hoje um dos principais problemas socioeconômicos mundiais. Entre os fenômenos de maior ocorrência estão os deslizamentos, inundações e os terremotos (MAFFRA & MAZOLLA, 2007; TOMINAGA, 2009; RIFFEL, 2012; LICCO, 2013). Durante o século XX, ocorreram modificações importantes no ambiente natural em decorrência do desenvolvimento econômico e do acelerado processo de urbanização verificado nas últimas décadas (TOMINAGA *et al.*, 2012).

Segundo o EM-DAT (2016), no período de 1991 a 2015, foram registrados no mundo 7.602 desastres, totalizando 1.620.093 óbitos e 4.496.909.737 pessoas afetadas. A análise dos dados para os anos de 2000 e 2010, mostra um aumento na frequência de ocorrência dos desastres, registrando 526 e 533 desastres, respectivamente.

No caso do Brasil, o Atlas Brasileiro de Desastres Naturais indica um aumento expressivo no número de ocorrências de desastres, no período 1991-2010, totalizando 8.671 ocorrências na década de 1990, e 23.238 na década de 2000. Nesse sentido, o Centro de Estudos e Pesquisas sobre Desastres (CEPED) alerta sobre os cuidados necessários ao afirmar que os desastres aumentaram em 73% nos últimos 10 anos (CEPED, 2012).

Os desastres naturais que ocorrem no Brasil são, na sua maioria, de origem atmosférica. A precipitação, seja em excesso ou em escassez, é o elemento atmosférico que mais contribui para a ocorrência de desastres na região Sul do Brasil (inundações e secas), (MARCELINO *et al.*, 2006; BARBIERI *et al.*, 2009).

Os movimentos de massa estão entre os principais tipos de desastres que ocorrem no Brasil atualmente, juntamente com as inundações

(PARIZI *et al.*, 2010; TOMINAGA *et al.*, 2012; RECKZIEGEL, 2012). As corridas de detritos ou de massa, deslizamentos planares e quedas de blocos são os mais frequentes no sul do Brasil (ZUQUETTE, 2005; TOMINAGA, 2012).

Os mapeamentos de suscetibilidade surgem como uma alternativa à cultura forense existente no Brasil em relação aos movimentos de massa, pois consistem em uma previsão de áreas onde é necessária maior atenção em relação à possibilidade de ocorrência de desastres (RIFFEL, 2012).

Vários modelos têm sido utilizados para mapear, modelar, quantificar e identificar áreas suscetíveis a movimentos de massa, como, os modelos de mineração de dados (RODRIGUES, 2002; CARDOSO, 2005; MARTINI *et al.*, 2006; SILVEIRA *et al.*, 2014) e o Shalstab (FERNANDES *et al.*, 2001; OLIVEIRA, 2009; PECHINCHA & ZAIDAN, 2013; SANTOS, 2015).

Mineração de Dados (MD) pode ser entendido como um processo de busca e descoberta de informações em grande volume de dados (WITTEN *et al.*, 2011). Dentre os modelos de MD avaliados no mapeamento da suscetibilidade a deslizamentos estão, Redes Bayesianas (SONG *et al.*, 2009), Redes Neurais Artificiais (FALASCHI *et al.*, 2009), Máquina de Vetores Suporte (BALLABIO & STERLACCHINI, 2012) e Árvore de Decisão (BUI *et al.*, 2012). Comparando os modelos de MD, a Árvore de Decisão (AD) possibilita a interpretação dos modelos resultantes, como também, demanda menos tempo na sua construção. E ainda, alcançam resultados semelhantes ou até superiores aos demais (TSAI *et al.*, 2013).

O modelo Shalstab consiste na abreviação de Shallow Landslide Stability Analysis (análise de estabilidade de deslizamentos rasos). É

um modelo determinístico que leva em conta a combinação de outros dois modelos: um hidrológico de estado uniforme e outro de estabilidade de encosta infinita (DIETRICH & MONTGOMERY, 1998).

Devido a recorrência de eventos adversos na bacia, este artigo tem por objetivo mapear a suscetibilidade a deslizamentos, e comparar o resultado da aplicação de duas metodologias, o Shalstab e a Árvore de Decisão, para identificar e analisar os padrões de mapeamento.

1.1 Área de estudo

A bacia hidrográfica do arroio Forromeco está inserida na bacia do rio Caí. Possui recorrência de processos associados a enxurradas e movimentos de massa. Situa-se a nordeste do Estado do Rio Grande do Sul, entre as coordenadas geográficas 29°10' a 29°30' de latitude Sul e 51°00' a 51°30' de longitude Oeste (Figura 1). Possui superfície de 290,70 km² e população estimada em 96.000 habitantes

(IBGE, 2014), abrangendo municípios como Bom Princípio, São Vendelino, Carlos Barbosa e Farroupilha.

A exploração agrícola intensa, o desmatamento das encostas declivosas e a poluição hídrica no curso médio e inferior são os maiores problemas enfrentados nesta bacia (SEMA, 2010).

De acordo com a classificação climática de Rossato (2011) a bacia do arroio Forromeco encontra-se no tipo climático denominado Subtropical III, caracterizado por um clima úmido com variação longitudinal das temperaturas médias. Possui uma menor influência dos sistemas polares e maior interferência dos sistemas tropicais conjugados com o efeito do relevo (escarpa e vales da borda do Planalto Basáltico), da continentalidade, da maritimidade e das áreas urbanizadas. Os sistemas frontais são responsáveis pela maior parte das precipitações. Chove entre 1700-1800 mm ao ano, em 100-120 dias de chuva.

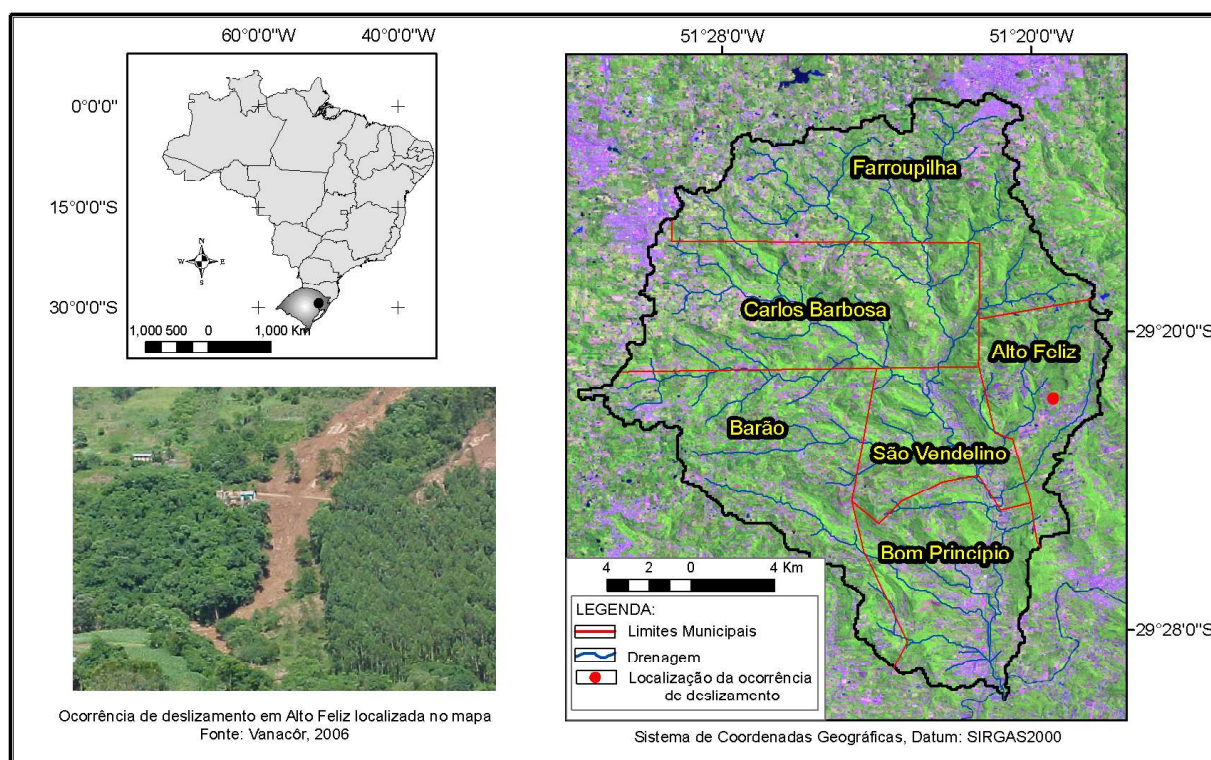


Fig. 1 - Localização da área de estudo.

Segundo Oliveira *et al.* (2010), a bacia do arroio Forromeco possui amplitude altimétrica de aproximadamente 750 metros sendo composta por vales bem encaixados com encostas de elevada declividade. Abrange as Unidades Geomorfológicas Planalto dos Campos Gerais,

Serra Geral, Patamares da Serra Geral e Planícies Alúvio Coluvionares. A bacia hidrográfica do arroio Forromeco - RS é frequentemente atingida por diversos tipos de desastres (OLIVEIRA, 2010; SCHENKEL, 2014). O principal registro de inundação da cidade remete ao ano de 1982,

quando parte da área urbana foi invadida e destruída pelas águas do arroio Forromeco. E, em 2000 foi registrada a ocorrência de um grande fluxo de detritos, formado com o material oriundo de escorregamentos de terra, próximo a uma das cabeceiras da bacia, fazendo algumas vítimas fatais (REIS *et al.*, 2014). Segundo Silveira (2008) no evento da corrida de detritos de 2000, o posto pluviométrico, junto a cidade de São Vendelino, indicou uma precipitação de 148 mm, em um período de duas horas, entre as 17 e 19 horas. Ao todo foram contabilizados 47 escorregamentos importantes de taludes, entre os quais 10 corridas de detritos. Seis rodovias vicinais foram interrompidas e a principal rodovia que liga a região de Caxias do Sul à região metropolitana de Porto Alegre foi interrompida completamente em 10 pontos.

Conforme o Sistema Brasileiro de Classificação de Solos (EMBRAPA, 2014), a litologia da região inclui a Formação Botucatu, com presença de arenito; a Fácies Gramado caracterizada por derrames basálticos; e a Fácies Caxias caracterizada por derrames de composição intermediária a ácida. Os solos que compõem a bacia são: Chernossolos, Neossolos Litólicos Eutróficos, Nitossolos e Luvisolos.

2. MATERIAL E MÉTODOS

As etapas metodológicas que possibilitaram a realização do presente estudo estão apresentadas na Figura 2, seguidas de uma breve descrição das principais etapas:

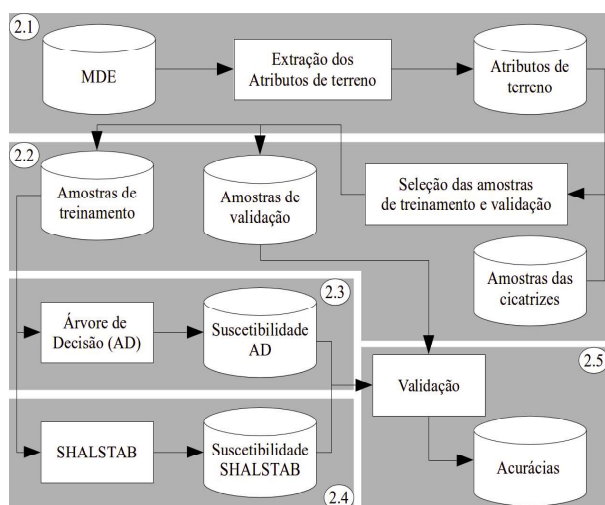


Fig. 2 - Fluxograma com etapas metodológicas.

2.1 Extração dos atributos de terreno

O Modelo Digital de Elevação (MDE) foi elaborado a partir de imagens do Shuttle Radar Topography Mission (SRTM), com resolução espacial de 30 metros, adquiridas no repositório do Serviço Geológico dos Estados Unidos (United States Geological Survey - USGS). Extraíram-se os seguintes atributos de terreno: a elevação da superfície (ELEV), a orientação das vertentes (OV), a declividade (DECL), a classificação das formas do terreno (GEOFORMAS), no qual é originado da curvatura de vertical e da horizontal. O Fator Topográfico (LS), o índice de umidade topográfica (IUT), a distância vertical da rede de drenagem (DVR) e a profundidade do vale (PV) (CONRAD, 2015). A extração dos atributos preditivos foi executada no programa SAGA-GIS.

2.2 Seleção das amostras de treinamento e validação

As áreas suscetíveis à ocorrência de movimento de massa foram determinadas, por meio de um inventário, elaborado a partir de interpretação de imagens de satélite de alta resolução. Já as áreas improváveis a ocorrência foram obtidas por interpretação dos atributos do terreno, selecionando locais planos e/ou próximos as redes de drenagens. Dos 2.836 pixels amostrados, 70% foram selecionados aleatoriamente como amostras de treinamento e o restante (30%), compôs as amostras de validação. A Figura 3 apresenta as amostras de treinamento distribuídas nos atributos LS e declividade, as amostras da classe de não ocorrência estão distribuídas na região central do atributo de terreno LS (Figura 3a). As amostras da classe ocorrência estão concentradas na região leste do atributo declividade (Figura 3b).

2.3 Mineração de dados

Modelos de AD conseguem identificar relações entre a variável dependente e as múltiplas variáveis independentes por meio de um particionamento recursivo binário, em que os dados são divididos em grupos cada vez mais homogêneos, utilizando para isso combinações das variáveis que melhor explicam a variação da variável dependente (QUINLAN, 1996).

2.1 Extração dos atributos de terreno

O Modelo Digital de Elevação (MDE) foi elaborado a partir de imagens do *Shuttle Radar Topography Mission* (SRTM), com resolução espacial de 30 metros, adquiridas no repositório do Serviço Geológico dos Estados Unidos (*United States Geological Survey* - USGS). Extraíram-se os seguintes atributos de terreno: a elevação da superfície (ELEV), a orientação das vertentes (OV), a declividade (DECL), a classificação das formas do terreno (GEOFORMAS), no qual é originado da curvatura de vertical e da horizontal. O Fator Topográfico (LS), o índice de umidade topográfica (IUT), a distância vertical da rede de drenagem (DVR) e a profundidade do vale (PV) (CONRAD, 2015). A extração dos atributos preditivos foi executada no programa *SAGA-GIS*.

2.2 Seleção das amostras de treinamento e validação

As áreas suscetíveis à ocorrência de movimento de massa foram determinadas, por meio de um inventário, elaborado a partir de interpretação de imagens de satélite de alta resolução. Já as áreas improváveis a ocorrência foram obtidas por interpretação dos atributos do terreno, selecionando locais planos e/ou próximos as redes de drenagens. Dos 2.836 pixels amostrados, 70% foram selecionados aleatoriamente como amostras de treinamento e o restante (30%), compôs as amostras de validação. A Figura 3 apresenta as amostras de treinamento distribuídas nos atributos LS e declividade, as amostras da classe de não ocorrência estão distribuídas na região central do atributo de terreno LS (Figura 3a). As amostras da classe ocorrência estão concentradas na região leste do atributo declividade (Figura 3b).

2.3 Mineração de dados

Modelos de AD conseguem identificar relações entre a variável dependente e as múltiplas variáveis independentes por meio de um particionamento recursivo binário, em que os dados são divididos em grupos cada vez mais homogêneos, utilizando para isso combinações das variáveis que melhor explicam a variação da variável dependente (QUINLAN, 1996).

O número de partições, ou seja, a profundidade da árvore é um parâmetro ajustável do modelo, quanto maior esse valor, maior e mais complexa a árvore resultante. Modelos por AD complexos decoram os dados de entrada e são menos generalistas (QIAN *et al.*, 2015). AD também podem ser utilizados na redução da dimensionalidade dos dados e na seleção de atributos mais informativos, calculando a importância do atributo para o modelo. A importância dos atributos é obtida quantificando a redução da impureza dos subconjuntos de amostras selecionados em cada nó da árvore e somando essa redução para cada atributo utilizado no nó.

Foram utilizados os modelos Classificação e Regressão por Árvore (*Classification And Regression Tree* – CART) (BREIMAN *et al.*, 1984) e o *Random Forest* (BREIMAN, 2001). Com intuito de selecionar o melhor valor de profundidade da árvore, foram avaliados os valores, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50 para ambos os modelos. O método *Random Forest* é um classificador do tipo comitê ou “*ensemble*”, constituído de várias AD. Cada uma dessas árvores dá um voto que indica sua decisão sobre a classe à qual pertencerá uma determinada amostra. A amostra pertencerá à classe com o maior número de votos.

O número de árvores é um parâmetro ajustável, neste estudo foram avaliados os seguintes valores, 2, 5, 10, 15, 20, 25 e 30. Para tanto foram desenvolvidas rotinas de programação na linguagem Python, e os métodos de classificação, CART e *Random Forest*, foram acessados por meio da biblioteca de programação *SciKit - Learn*.

2.4 SHALSTAB

Esse modelo foi rodado em *Arcview*, inicialmente com dados de elevação obtidos a partir do DEM. O *Shalstab* é organizado em 3 módulos, inicialmente no primeiro módulo são gerados mapas morfométricos como: declividades, área de contribuição e fluxos. Para a entrada do modelo utilizou-se valores estabelecidos por Silveira (2008) como: coesão de 7,2 kPa, ângulo de atrito interno de 31,6°, peso específico úmido 15.7 kN/m³, profundidade do solo 2 metros.

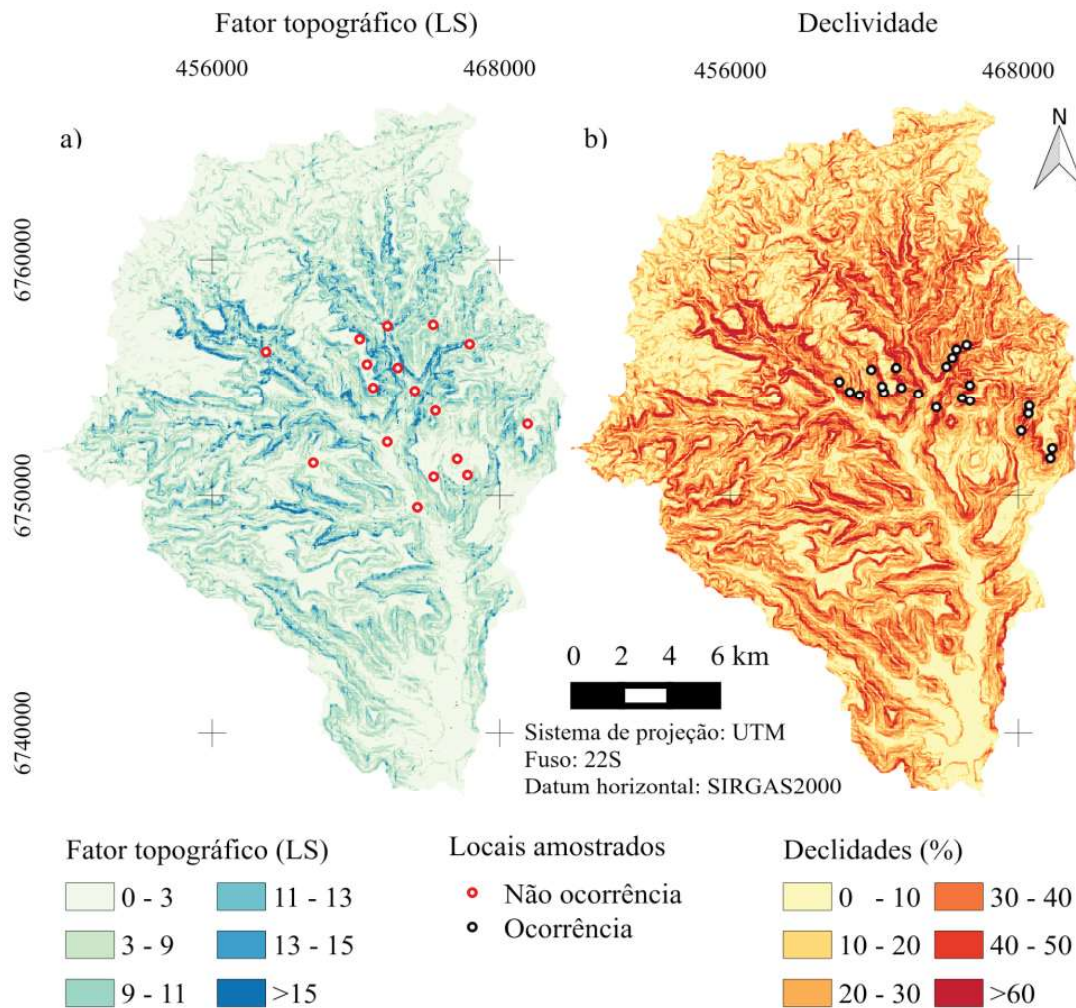


Fig. 3 - Amostras de treinamento distribuídas por: a) Fator topográfico (LS); e b) declividade.

O modelo original gera um mapa de suscetibilidade em 7 classes, sendo duas extremas, uma cronicamente instável e outra incondicionalmente estável (Quadro 1). Foram adicionadas amostras de escorregamentos, para identificar em quais classes é maior a sua ocorrência.

Como os valores são muito baixos, o modelo apresenta as classes em logaritmos de Q/T que variam de $<-3,1$ a $>-2,2$, essas classes são padronizadas, baseados em valores usualmente encontrados em deslizamentos planares. A última etapa consiste na classificação das classes com Alta e Baixa ocorrência de escorregamentos, a fim de gerar o mapa de suscetibilidade e permitindo a comparação entre os resultados obtidos por MD e *Shalstab*. Para realizar a classificação em duas classes, foram avaliados os seguintes limiares: -2; -1,5; -1; -0,5; 0; 0,5; 1; 1,5 e 2.

2.5 Validação

Os mapas de suscetibilidades gerados pelos modelos foram elaborados com duas classes: alta suscetibilidade e baixa suscetibilidade, o que permitiu comparar os resultados obtidos. Foram calculados o índice *Kappa* e sua variância, e a Acurácia do Produtor (AP) para cada modelo. Esses cálculos foram determinados por meio da biblioteca de programação *StatsModels*, essa biblioteca permite explorar dados, estimar modelos estatísticos e executar testes estatísticos.

3. RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados 3 itens contendo resultados e discussões, sendo o primeiro com resultados de modelos baseados em árvores de decisão, o segundo com resultados de modelos do *Shalstab* e no terceiro item será feita uma comparação dos dois modelos.

3.1 Modelos Baseados em Árvores de Decisão

O modelo *Random Forest* resultou em um valor de *Kappa* próximo a 0,98, para número de árvores igual a 5 e 15. Quanto menor esse valor, menor será o tempo da classificação, devido a isso, foi selecionado o número de árvores igual a 5, como o mais adequado neste trabalho. A Figura 4 ilustra o desempenho das classificações em relação aos valores de profundidades das árvores, o maior *Kappa* para o *Random Forest*, foi obtido para o número de árvores igual a 5. Observa-se na Figura 4, que o modelo *Random Forest* foi superior ao CART em todos os casos, isso ocorre devido ao modelo considerar várias árvores na classificação. Ao avaliar o comportamento do *Random Forest* em relação

aos valores de profundidade, pode-se identificar um declínio da acurácia a partir do valor 25. Para o modelo CART esse declínio iniciou no valor de profundidade igual a 15. O desempenho dos modelos para maiores valores de profundidades das árvores confirmam que árvores complexas superajustam-se as amostras de treinamento. Como consequência, resultam em classificações equivocadas para amostras não utilizadas no treinamento. Poudyal (2012) em seu trabalho para determinar a suscetibilidade a movimento de massa no Phidim, uma área ao leste do Nepal, obteve o valor 12 como o mais apropriado para a profundidade. Esse número está entre o intervalo que resultou nos maiores valores de *Kappa* deste presente estudo.

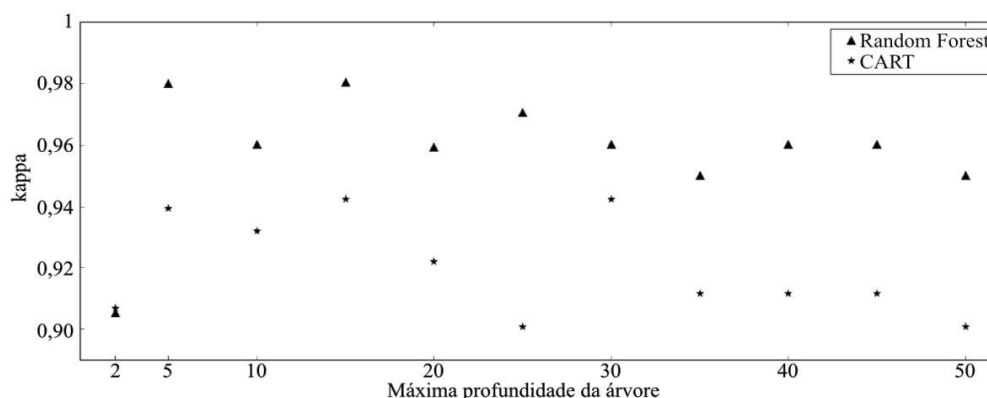


Fig. 4 - Acurácia das classificações dos modelos baseados em Árvores em relação aos valores de profundidades das árvores.

Nesse estudo o modelo de classificação mais apropriado foi o *Random Forest*, com número de árvores igual a 5 e profundidade igual 5, obtendo uma acurácia superior a 0,98 Entretanto, outros trabalhos relatam acurácias inferiores, como Saito *et al.* (2009), os autores alcançaram valores de *Kappa* próximos a 0,61, utilizando o modelo por AD C4.5, descrito no programa *WEKA* como J48, na análise da susceptibilidade a movimento de massa em uma área das Montanhas de Akaishi, Japão. Já Bui *et al.* (2012) utilizando o modelo C4.5, obtiveram uma acurácia igual a 0,86 na classificação de riscos de movimento de massa na província Hoa Binh, Vietnã.

A Figura 5 mostra a importância das variáveis utilizadas (atributos do terreno) para a discriminação das classes, com base no modelo *Random Forest*.

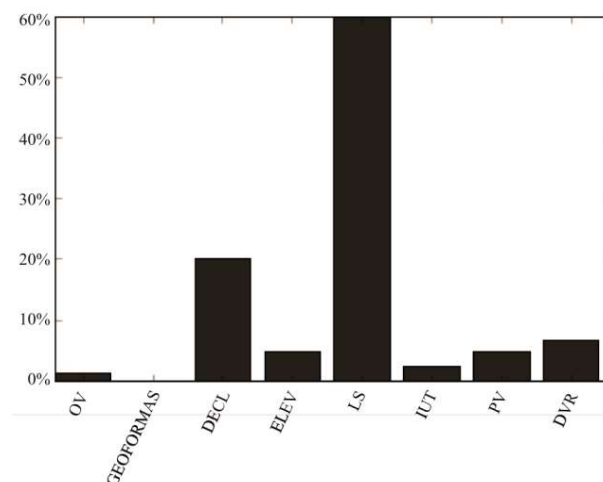


Fig. 5 - Importância dos atributos de terreno.

O atributo do terreno mais importante foi o fator topográfico (LS), com valor de importância próximo a 60%, seguido do modelo digital de elevação (ELEV), importância superior a 20%.

Esses valores indicam que os atributos LS e ELEV foram mais utilizados nos nós da árvore e proporcionaram uma maior redução da impureza dos subconjuntos.

Os demais atributos do terreno resultaram em importâncias inferiores a 10%. GEOFORMAS obteve o menor valor de importância, próximo a 0%, podendo ser desconsiderado da análise.

Dentre os atributos utilizados por Youssef *et al.* (2015), a declividade e a elevação do terreno se mostraram atributos mais importantes na análise de suscetibilidade a movimento de massa com o modelo *Random Forest*, em que a declividade resultou no maior valor de importância. Para Nefeslioglu *et al.* (2010), utilizando um modelo AD disponível no *Microsoft Server 2008 Analysis Servicess*, para avaliar a suscetibilidade a movimento de massa, os maiores valores foram para declividade e curvatura de perfil, e a elevação do terreno ficou entre os menos importantes. A curvatura vertical, juntamente com a curvatura horizontal, neste trabalho, foram utilizadas para construir o atributo GEOFORMAS, desse modo, futuros trabalhos podem considerar o uso dos atributos separadamente.

3.2 SHALSTAB

Para a geração do mapa de suscetibilidade através do *Shalstab* foram utilizados dados extraídos do MDE e dados de solo obtidos em campo por Silveira (2008), conforme descrito na metodologia.

O mapa elaborado pelo *Shalstab* contém 7 classes de suscetibilidade, baseado apenas nos dados do relevo extraídos do MDE sendo duas classes extremas, uma cronicamente instável e uma estável. A etapa da validação do mapa de log q/t consiste na inserção das amostras de escorregamentos nas classes geradas pelo modelo e, a partir da relação com o mapa do log Q/T são estabelecidos os limiares do mapa de suscetibilidade. O Quadro 1 mostra os valores de log q/t as áreas e cicatrizes para cada classe de mapeamento.

Conforme Quadro 1, a maioria das ocorrências de escorregamentos está nas classes entre os valores -3,1 e -2,8 a -2,5, que correspondem a aproximadamente 76% das ocorrências totais. Duas classes correspondem a aproximadamente 7% da área total da bacia.

Quadro 1: Escorregamentos e áreas de suscetibilidade a movimentos de massa por classe de estabilidade (*SHALSTAB*)

Classe (log Q/T)	Área das classes (%)	Cicatriz	Cicatriz (%)
<i>Cronicamente Instável</i>	0,18	2	4,25
<-3.1	2,70	20	42,55
-3,1 a -2,8	3,91	16	34,05
-2,8 a -2,5	4,39	6	12,75
-2,5 a -2,2	2,07	0	0
>-2,2	0,23	0	0
<i>Estável</i>	86,52	3	6,40

Para a elaboração do mapa de suscetibilidade foi considerado o limiar mínimo de log q/t de 2. Essa foi a melhor resposta do modelo com o mais alto valor de *Kappa* (0,82). Para o valor de -2,5, limiar que seria escolhido baseado na classificação das ocorrências de deslizamentos em relação às classes de estabilidade, o *Kappa* foi mais baixo (0,64).

As sete classes foram simplificadas para apenas duas classes, uma que corresponde às áreas estáveis, e outra de áreas críticas, para fins de comparação. Dessa forma, 3 cicatrizes ficam fora da área de instabilidade. Essas 3 amostras, possivelmente demonstram um erro do modelo em relação à qualidade necessária do MDE e das amostras. O modelo possui uma resolução de 30 metros e, nessas áreas, embora sejam de alta declividade, o mapa do *Shalstab* indicou que essa área é estável.

Analisando os mapas de declividade (Figura 3b) e de elevação (Figura 6a) gerados pelo *Shalstab* percebe-se que a maior interação entre os mapas se deu a partir da classe de -2,8 a -2,5 que apresenta a maioria da sua área nas declividades entre 40 e 60% e está na faixa entre os 400 e 600 metros de altitude. A classe de -3,1 a -2,8 ficou classificada nas áreas de declividade entre 70 e 80% e entre os 200 e 300 metros de altitude na sua maioria, enquanto que a classe cronicamente instável está na sua maior parte acima de 80% de declividade e entre os 500 e 700 metros.

Vieira e Ramos (2015) ao aplicarem o *Shalstab* em Caraguatatuba na Serra do Mar

(SP) identificaram que cerca de 55% das cicatrizes dos escorregamentos se concentraram nas classes mais instáveis, sendo elas: 16,2% (incondicionalmente instável), 12,7% ($\log Q/T < -3,1$), 11,7% ($\log Q/T -3,1$ à $-2,8$) e 14,4% ($\log Q/T -2,8$ à $-2,5$). 45% das cicatrizes se concentraram nas classes “estáveis”, sendo elas; 16,4% (incondicionalmente estável), 10% ($\log Q/T > -2,2$) e 18% ($\log Q/T -2,5$ à $-2,2$). Segundo os autores, as diferenças entre o planalto e as escarpas podem explicar os resultados, uma vez que o planalto concentrou uma parte da área classificada como incondicionalmente estável com a presença de cicatrizes de escorregamentos. Outra explicação seria a existência, nessa área de planalto, de cicatrizes de escorregamentos rotacionais não previstos pela estrutura matemática do *Shalstab*.

Montgomery e Dietrich (1994), reportaram que de 83% a 100% do total das cicatrizes

atingiram limites de $\log(Q/T) < -2.5$. Percentuais acumulativos um pouco menores foram obtidos por Dietrich e Montgomery (1998), que identificaram apenas 54% dos escorregamentos em valores de $\log(Q/T) < -3.1$, 68% em $\log(Q/T) < -2.8$ e 90% em $\log(Q/T) < -2.5$.

3.3 Comparação entre os modelos

O fator LS foi selecionado pelo modelo *Random Forest* como o mais importante na discriminação das classes, esse atributo combina o comprimento de rampa (L) e a declividade (S) do terreno, altos valores para esse atributo correspondem em maiores velocidades do escoamento superficial, uma vez que, a suscetibilidade a movimento de massa está diretamente relacionada a declividade do terreno e a precipitação, o LS pode ser um atributo a ser considerado em futuros estudos de movimentos de massa.

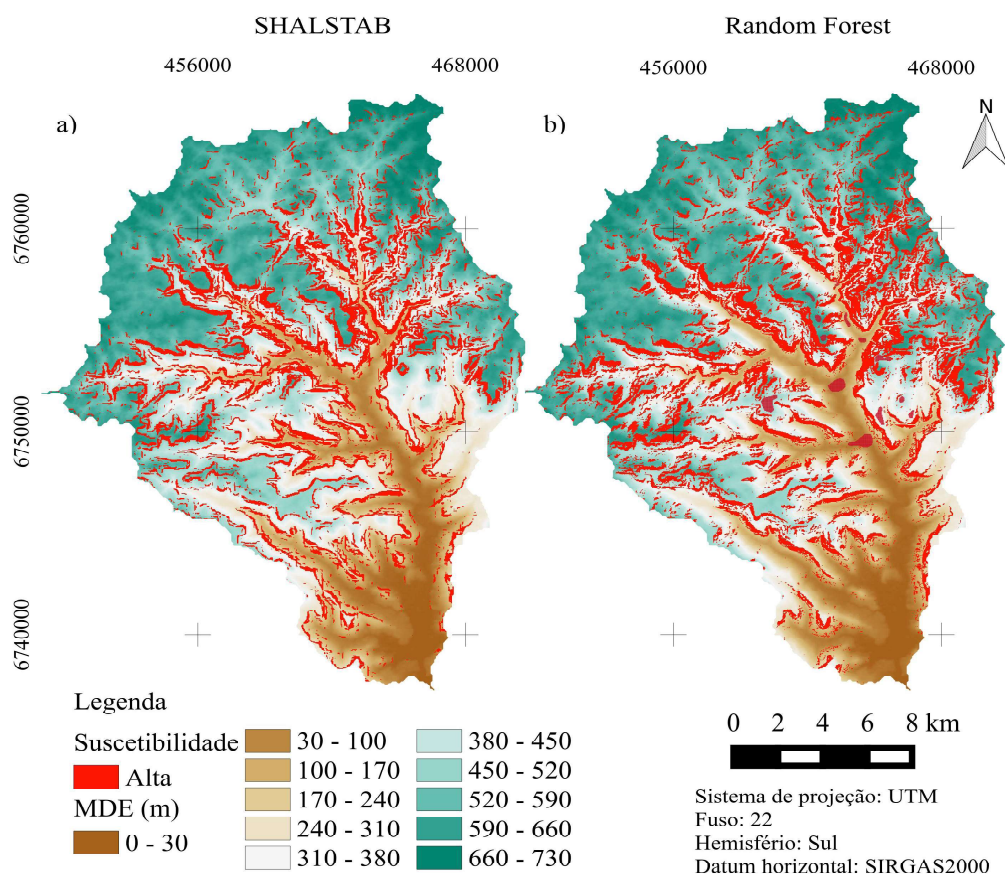


Fig. 6 – Classe de suscetibilidade: a) *Shalstab*; e b) *Random Forest*.

A declividade foi selecionada por ser o atributo de maior importância no modelo *Shalstab*, já que influencia diretamente na ocorrência de movimentos de massa. De modo

geral, áreas de maior declividade são mais suscetíveis a movimentos de massa. Percebe-se no mapa que grande parte das ocorrências amostradas estão localizadas em áreas com

declividades superiores a 40%, as mesmas áreas se sobrepõem nos mapas de suscetibilidade, pois tanto o modelo *Random Forest* quanto o *Shalstab* são influenciados pela declividade (figuras 3a e 3b).

A Figura 6 apresenta as áreas com alta suscetibilidade a movimento de massa, obtidas pelas classificações dos modelos *Shalstab* (Figura 6a) e *Random Forest* (Figura 6b). Também estão representados os mapas hipsométricos, separados em 11 classes com elevações entre 30 e 730 metros. As classificações de ambos os modelos foram selecionadas porque resultaram nos maiores valores de *Kappa*.

Para o *Shalstab* o maior valor de *Kappa* foi obtido com o limiar de seleção igual a 2 (Figura 6a), no qual obteve *Kappa* igual a 0,82 (Quadro 2). Já para o *Random Forest*, o melhor caso, resultou em um valor de *Kappa* igual a 0,98 (Quadro 2).

Quadro 2: Matriz de confusão dos modelos

Classes	<i>SHALSTAB</i>		<i>Random Forest</i>	
	Baixa	Alta	Baixa	Alta
Baixa	762	16	762	2
Alta	0	39	0	53
Total	762	55	762	55
AP	100%	70,9%	100%	96,4%
<i>Kappa</i>	0.82		0.98	
Variância	0.0020746		0.0001960	

A análise do *Shalstab* (Figura 6a) mostra que o modelo classificou áreas com altos valores de declividade em todas as encostas como suscetíveis, e também em áreas com forte convergência de linhas de drenagem, amparado pela influência que as vertentes convergentes possuem em relação à suscetibilidade a deslizamentos.

O modelo *Random Forest* (Figura 6b) considerou mais áreas com alta suscetibilidade, principalmente na região nordeste, e nas encostas orientadas para leste e norte não foram classificadas como alta suscetibilidade.

Como o modelo *Random Forest* considerou vários atributos de terreno, dentre eles a orientação das vertentes (OV), no momento em que as regras de classificação foram geradas, o

modelo desconsiderou as encostas orientadas para leste e norte. No referencial teórico utilizado no trabalho, percebe-se que o atributo de orientação de vertentes não é utilizado como variável para mapeamento de suscetibilidade.

Ao comparar os dois modelos percebe-se que o *Random Forest* classificou como de alta suscetibilidade uma área superior à classificada pelo *Shalstab*. O que pode ser explicado tendo em vista que o *Random Forest* relaciona diretamente as amostras de deslizamentos com as características das áreas suscetíveis, enquanto que o *Shalstab* realiza uma exclusão de áreas suscetíveis, definidas pelos dados de entrada do relevo a partir do cruzamento de informações com as amostras de deslizamentos, etapa compreendida no *Shalstab* como de validação.

Ao considerar a área mapeada, o modelo *Random Forest* classificou como de alta suscetibilidade um total de 90.457 pixels, o que corresponde a 8.141,13 ha da área da bacia, enquanto que o modelo *Shalstab* classificou 60.416 pixels, ou 5.437,44 ha.

O Quadro 2 apresenta as matrizes de confusão dos modelos. No total foram selecionados 762 pixels para a classe Baixa e 55 pixels para a classe Alta. Identifica-se que os modelos *Shalstab* e *Random Forest* classificaram a classe com baixa suscetibilidade em uma AP igual a 100%.

Os erros de classificação foram identificados para a classe com alta suscetibilidade, como pode ser constatado no valor da AP para o modelo *Random Forest*, que foi igual a 96,4%. Para o *Shalstab* a AP foi de 70,9%, indicando que as áreas com alta suscetibilidade estão classificadas como baixa.

Avaliando estatisticamente em um nível de significância de 5%, os valores de *Kappa* e as respectivas variâncias dos modelos, define-se que esses resultados são diferentes, ou seja, a classificação por meio do *Random Forest* resultou em uma acurácia superior ao do *Shalstab*. Ambos os modelos representaram resultados coerentes para identificar áreas suscetíveis a deslizamentos. O modelo *Random Forest*, no entanto, apresentou resultados mais satisfatórios, pois utiliza apenas parâmetros locais baseados nas características da área de estudo, enquanto que o *Shalstab* possui uma hierarquia em classes já estabelecidas no seu modelo. Além disso o *Shalstab* possui uma

capacidade menor de detectar as amostras de deslizamentos diferentes do *Random Forest* que classificou quase o total das amostras dentro de áreas suscetíveis.

4. CONCLUSÕES

O modelo *Random Forest* apresentou resultados mais satisfatórios em relação ao índice de acerto, se comparado com o modelo *Shalstab*, atestado pelos valores de *Kappa* e pela validação em relação às cicatrizes localizadas nas áreas classificadas como críticas. Apresentando melhor adaptação em relação às características da área.

Essa diferença nos resultados pode ter sido influenciada pelos seguintes motivos: (a) o *Shalstab* utiliza dados pré-estabelecidos em outras áreas para a geração dos algoritmos, que podem ter características diferentes da área de estudo e interferir nos resultados, enquanto que o modelo Árvore de Decisão utiliza apenas dados da área de estudo; (b) o modelo Árvore de Decisão estabelece as áreas de suscetibilidade a partir de dados de entrada, incluindo as amostras de deslizamentos e as características do relevo, enquanto que o *Shalstab* utiliza apenas dados do relevo; (c) as amostras de deslizamentos são utilizadas como dados de validação, para eliminar áreas que tenham sido consideradas como suscetíveis, no entanto, sem cruzamento de informações com as áreas de amostras; (d) o *Shalstab* é utilizado apenas para deslizamentos planares, sendo que alterações geomorfológicas podem trazer diferenças nos resultados, e ainda que devido ao tamanho do pixel, é necessário uma imagem de alta resolução para dados mais precisos.

Ressalta-se que a diferença entre os resultados dos modelos é pequena. Ambos podem representar de forma satisfatória o objetivo final de geração de mapa de suscetibilidade. No entanto, o modelo *Shalstab* apresenta mais limitações, devido à necessidade de dados de grande resolução, para uma boa resposta da análise.

REFERENCIAS BIBLIOGRÁFICAS

BALLABIO C.; STERLACCHINI S. Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy., **Mathematical Geosciences**, v. 44, n. 1, p. 47–70, 2014.

BARBIERI, D. W.; FERREIRA, C. C.; SAITO, S. M.; SAUSEN, T. M.; HANSEN, M. A. F. Relação entre os desastres naturais e as anomalias de precipitação para a região Sul do Brasil. **XIV Simpósio Brasileiro de Sensoriamento Remoto**, p. 3527-3534, 2009.

BUI, D. T.; PRADHAM, B.; LOFMAN, O.; REVHAUG, I. Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models. **Mathematical Problems in Engineering**, v. 2012, n. 1, 2012, 26p.

BREIMAN L Random Forest. **Machine Learning** v. 45, n.1, p. 5–32, 2001.

BREIMAN L.; FRIEDMAN JH.; OLSHEN RA.; STONE CJ. **Classification and regression trees**. Chapman & Hall, New York, 1984, 355p.

CARDOSO, D. **Análise da suscetibilidade a movimentos de massa no município de Peruíbe - SP, com o apoio de um sistema integrador de informações georreferenciadas**. Dissertação (Mestrado em Geociências). Universidade Estadual de São Paulo. Rio Claro, 2005, 128p.

CENTRO UNIVERSITÁRIO DE ESTUDOS E PESQUISAS SOBRE DESASTRES (CEPED). **Atlas Brasileiro de Desastres Naturais 1991 a 2010**. Brasil. Florianópolis: Universidade Federal de Santa Catarina, 2012, 88p.

CONRAD, O.; BECHTEL, B.; BOCK, M.; DIETRICH, H.; FISCHER, E.; GERLITZ, L.; WEHBERG, J.; WICHMANN, V.; BOHNER, J.; System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. **Geoscientific Model Development**, v. 8, n. 7, p. 1991-2007, 2015.

DIETRICH, W. E.; MONTGOMERY, D. R. **SHALSTAB: a digital terrain model for mapping shallow landslide potential**. NCASI (National Council of the Paper Industry for Air and Stream Improvement) Technical Report, 1998, 29p.

EMERGENCY DATABASE (EM-DAT). OFDA/CRED **The Office of Foreign Disaster Assistance/Centre for Research on the Epidemiology of Disasters** - Université Catholique de Louvain, Brussels, Belgium. (s/d.). Disponível em: <<http://www.emdat.be/>>

- disaster_trends/index.html>. Acesso: em 15 ago. 2016.
- EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. **Sistema Brasileiro de Classificação de Solos**. 4. ed. Brasília: EMBRAPA, 2014. 376 p.
- FERNANDES, N. F.; GUIMARÃES, R. F.; GOMES, R. A. T.; VIEIRA, B. C.; MONTGOMERY, D. R. Condicionantes Geomorfológicos dos Deslizamentos nas Encostas: Avaliação de Metodologias e Aplicação de Modelo de Previsão de Áreas Suscetíveis. **Revista Brasileira de Geomorfologia**, v. 2, n. 1, p. 51-71, 2001.
- FUAN TSAI; JHE-SYUAN LAI; WALTER W. CHEN; TANG-HUANG LIN. Analysis of topographic and vegetative factors with data mining for landslide verification. **Ecological Engineering**, 2013, 61p.
- LICCO, E. A. Vulnerabilidade social e desastres naturais: uma análise preliminar sobre Petrópolis, Rio de Janeiro. **Interface, Revista de Saúde, Meio Ambiente e Sustentabilidade**, SP, v. 8, n. 1, p 24-41, 2013.
- MAFFRA, C. Q. T.; MAZOLLA M. As razões dos desastres em territórios brasileiros. In: **Vulnerabilidade Ambiental** / Rozely Ferreira dos Santos, organizadora. – Brasília: MMA, 2007. 192 p.
- MARCELINO, E. V.; NUNES, L. H.; KOBAYAMA, M. Banco de dados de desastres naturais: Análise de dados globais e regionais. **Caminhos de Geografia**, Uberlândia, v. 6, n. 19, p. 130-149, 2006.
- MARTINI, L. C. P.; UBERTI, A. A. A.; SCHEIBE, L. F.; COMIN, J. J.; OLIVEIRA, M. A. T. Avaliação da suscetibilidade a processos erosivos e movimentos de massa: decisão multicriterial suportada em sistemas de informações geográficas. **Geologia. USP, Série Científica**. v. 5, n. 2, São Paulo, p. 41-52, 2006.
- NEFESLIOGLU, H. A.; SEZER, E.; GOKCEOGLU, C.; BOZKIR, A. S.; DUMAN, T. Y. Assessment of Landslide Susceptibility by Decision Trees in the Metropolitan Area of Istanbul, Turkey. **Mathematical Problems in Engineering**, v. 2010, n. 1, 2010, 15p.
- OLIVEIRA, G. G. **Modelos para previsão, espacialização e análise das áreas inundáveis na bacia hidrográfica do rio Caí, RS**. Dissertação (Mestrado em Sensoriamento Remoto). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2010, 150p.
- OLIVEIRA, G. G.; GUASSELLI, L. A.; SALDANHA, D. L. Influência de variáveis morfométricas e da distribuição das chuvas na previsão de enchentes em São Sebastião do Caí, RS. **Revista de Geografia (Recife)**, v. 27, n. 3. p. 140-155, 2010.
- OLIVEIRA, M. T. **Estudo de movimentos de massa gravitacionais no município de Belo Horizonte, MG**. Tese (Doutorado em Engenharia Civil). Universidade Federal de Viçosa. Viçosa, 2009, 279 p.
- PARIZI, M. G.; SEBASTIÃO, C. S.; VIANA, C. S.; PFLUEGER, M. C.; CAMPOS, L; C.; CAJAZEIRO, J. M. D.; TOMICH, R. S.; GUIMARÃES, R. N.; ABREU, M. L.; SOBREIRA, F. G.; REIS, R.; Correlações entre chuvas e movimentos de massa no município de Belo Horizonte, MG, **Geografias: Artigos Científicos**, v.2, n.6, p. 49-68, 2010.
- PECHINCHA, M. G. H.; ZAIDAN, R. T. Risco a ocorrência de movimento de massa e seu potencial na análise da qualidade ambiental. **Revista de Geografia**, v. 2, n. 2, p. 1-7, 2013.
- POUDYAL, C. P. Landslide susceptibility analysis using decision tree method, Phidim, Eastern Nepal. **Bulletin of the Department of Geology**, v. 15, n. 1, p. 69-76, 2012.
- QIAN, Y.; ZHOU, W.; YAN, J.; LI, W.; HAN, L. Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. **Remote Sensing**, v. 7, n. 1, p. 153-168, 2015.
- QUINLAN, R. Improved use of continuous attributes in C4.5. **Journal of Artificial Intelligence Research**, v. 4, n. 279, p. 77-90, 1996.
- RECKZIEGEL, E. W. **Identificação e mapeamento das áreas com perigo de movimento de massa no município de Porto Alegre, RS**. Dissertação (Mestrado em Geografia). Universidade Federal do Rio Grande

do Sul. Porto Alegre, 2012, 90p.

REIS, J. T.; SILVA, J. S.; MICHEL, G. P.; KOBAYAMA, M. Mapeamento da vulnerabilidade a desastres hidrológicos nos municípios de Alto Feliz e São Vendelino/RS como forma de contribuição à engenharia de sedimentos. In: **Encontro Nacional de Engenharia de Sedimentos (ENES)**, 2014, João Pessoa. Anais do evento, 2014, 17p.

RIFFEL, E. S. **Análise e mapeamento de áreas suscetíveis a movimentos de massa na bacia hidrográfica do Rio Paranhana (RS)**. Dissertação (Mestrado em Geografia). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2012, 105p.

RODRIGUES, B. B. **Proposta de sistemática para tomada de decisão relativa a movimentos de massa gravitacionais: aplicação em Ouro Preto (MG)**. Tese (Doutorado em Geociências). Universidade Estadual Paulista, 2002, 308p.

ROSSATO, M. S. **Os climas do Rio Grande do Sul: variabilidade, tendências e tipologia**. Tese (Doutorado em Geografia). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2011, 253p.

SANTOS, V. **Influência da escala no mapeamento de áreas suscetíveis a movimentos de massa com o SHALSTAB na microbacia do Ribeirão do Baú, Ilhota/SC**. Monografia (Graduação em Engenharia Civil). Universidade Federal de Santa Catarina. Florianópolis, 2015, 126p.

SAITO, H.; NAKAYAMA, D.; MATSUYAMA, H. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. **Geomorphology**, v. 109, n. 3, p. 108-121, 2009.

SCHENKEL, J. C. **Mapeamento das áreas de risco de escorregamentos translacionais na bacia do arroio Forromeco**. Monografia (Graduação em Engenharia Ambiental). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2014, 76p.

SEMA. **Banco de dados Bacia Hidrográfica do rio Caí** Disponível em: <http://www.sema.rs.gov.br/conteudo.asp?cod_menu=56&cod_conteudo=5864>. Acesso em: 23 ago. 2016.

SILVEIRA, H. L. S.; VETTORAZZI, C. A.;

VALENTE, R. A. Avaliação multicriterial no mapeamento da suscetibilidade de deslizamentos de terra. **Revista Árvore**, Viçosa-MG, v. 38, n. 6, p. 973-982, 2014.

SILVEIRA, R. M. **Comportamento geotécnico de um solo coluvionar de São Vendelino (RS)**. Tese (Doutorado em Engenharia Civil). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2008, 302p.

SILVEIRA, R. M. P.; SILVEIRA, C. T.; OKAFIORI, C. Emprego de técnicas de inferência espacial para identificação de unidades de relevo apoiado em atributos topográficos e Árvores de Decisão. **Revista Brasileira de Geomorfologia** v. 15, n. 1, p. 87-101, 2014.

SONG, Y.; GONG, J.; GAO, S.; WANG, D.; CUI, T.; LI, Y.; WEI, B. Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan, China. **Journal Computers & Geosciences**, v. 42, n. 1, p. 189-199, 2012.

SOUZA, C. R. G. Flooding in the São Sebastião region, northern coast of São Paulo state, Brazil. **Anais Academia Brasileira Ciências**, v. 70, n. 2, p. 353-366, 1998.

TOMINAGA, L. K. Escorregamentos. In: **Desastres naturais: conhecer para prevenir**. São Paulo: Instituto Geológico, p. 25-38, 2009.

TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. (Orgs) **Desastres naturais: conhecer para prevenir**. São Paulo: Instituto Geológico, 2009, 196p.

VANACÔR, R. N. **Sensoriamento Remoto e Geoprocessamento aplicados ao mapeamento das áreas suscetíveis a movimentos de massa na região nordeste do Estado do Rio Grande do Sul**. Dissertação (Mestrado em Sensoriamento Remoto). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2006, 132p.

VIEIRA, B. C. **Previsão de Escorregamentos Translacionais Rasos na Serra do Mar (SP) a Partir de Modelos Matemáticos em Bases Físicas**. Tese (Doutorado em Geografia). Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2007, 193p.

VIEIRA, B. C.; RAMOS, H. Aplicação do modelo *Shalstab* para mapeamento da suscetibilidade a

escorregamentos rasos em Caraguatatuba, Serra do Mar (SP), **Revista do Departamento de Geografia – USP**, v. 29, n. 1, p. 161-174, 2015.

WITTEN, I.H.; FRANK, E.; HALL, M.A. **Data Mining: Practica Machine Learning Tools and Techniques**, 3ed. San Francisco: Morgan Kaufmann, 2011, 558p.

YOUSSEF, A.M.; POURGHASEMI, H.R.;

POURTAGHI, Z.S.; AL-KATHEERI, M.; Landslides susceptibility mapping using random forest. **Landslides**. v. 13, n. 5, p. 839-856, 2016.

ZUQUETTE, L. V. **Movimentos de massa gravitacionais - proposta de um sistema de previsão: aplicação na área urbana de Campos de Jordão - SP**. Tese (Doutorado em Engenharia Civil). Universidade de São Paulo. São Carlos, 2005, 390p.