

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E
METEOROLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO

Renata Pacheco Quevedo

**MAPEAMENTO DE SUSCETIBILIDADE A MOVIMENTOS DE MASSA A PARTIR
DE APRENDIZADO DE MÁQUINA**

Porto Alegre

2019

Renata Pacheco Quevedo

**MAPEAMENTO DE SUSCETIBILIDADE A MOVIMENTOS DE MASSA A PARTIR
DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Sensoriamento Remoto da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do Título de Mestra em Sensoriamento Remoto e Geoprocessamento.

Área de Concentração: Sensoriamento Remoto aplicado.

Orientador: Laurindo Antonio Guasselli

Co-Orientador: Guilherme Garcia de Oliveira

Porto Alegre

2019

CIP - Catalogação na Publicação

Quevedo, Renata Pacheco

Mapeamento de suscetibilidade a movimentos de massa a partir de aprendizado de máquina / Renata Pacheco Quevedo. -- 2019.

71 f.

Orientador: Laurindo Antonio Guasselli.

Coorientador: Guilherme Garcia de Oliveira.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre, BR-RS, 2019.

1. ALOS-PALSAR. 2. ASTER GDEM. 3. Redes Neurais Artificiais. 4. Random Forest. 5. Bacia Hidrográfica do Rio Rolante. I. Guasselli, Laurindo Antonio, orient. II. Oliveira, Guilherme Garcia de, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Renata Pacheco Quevedo

**MAPEAMENTO DE SUSCETIBILIDADE A MOVIMENTOS DE MASSA A PARTIR
DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Sensoriamento Remoto da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do Título de Mestra em Sensoriamento Remoto e Geoprocessamento.

Aprovada em 27 de fevereiro de 2019.

Prof. Dr. Laurindo Antonio Guasselli - Orientador

Prof. Dr. Guilherme Garcia de Oliveira – Coorientador

Prof. Dr. Clódís de Oliveira Andrades Filho - UERGS

Prof. Dr. Eduardo Samuel Riffel – 1º CGEO – DSG

Prof. Dr. Victor Fernandez Nascimento – PPGSR/UFRGS

AGRADECIMENTOS

Agradeço inicialmente aos meus pais, Silvana e Milton, por todo apoio e incentivo dados, por sempre estarem presentes e por fazerem todo o possível para garantir um futuro melhor para mim e meus irmãos. E aos meus irmãos, Rui e Rafa, por sempre estarem dispostos a ajudar quando precisei.

Aos professores Laurindo e Guilherme, por toda a orientação, trabalhos de campo, ajudas, conversas, conselhos, ensinamentos e apoio dados. Ambos sempre me incentivaram a ir mais longe e a explorar meu potencial, sou imensamente grata por toda a aprendizagem nesses 7 anos de convivência. Em especial, ao Laurindo, por me incentivar a ir para a Espanha e a buscar novos horizontes, e ao Guilherme por todo o apoio no final do mestrado, recebendo minhas finalizações um pouco atrasadas, mas sempre me tranquilizando.

Ao Luís Fernando, por toda a ajuda com meu segundo artigo e com as orientações, explicações e conselhos dados.

À Fabi, por toda a ajuda com os *abstracts*, por todo companheirismo, conselhos, abraços e por ter sempre esse coração enorme.

Ao Tales, pela ajuda com a formatação do trabalho e revisões finais, por mostrar que tudo na vida tem que ser equilibrado, momentos para focar e momentos para espalhar, além de todo o apoio dado e de sempre acreditar no meu potencial.

Aos meus queridos colegas do LAGAM: Ceci, Tássia, Delapasse, João, Pet, Samuel e Victor por todo esse tempo de convívio, conversas, apoio, trocas de conhecimentos, cafés, churrascos e cervejinhas que compartilhamos.

Aos meus queridos colegas de mestrado (Turma da Ressaca/Cachaça), com quem compartilhei mais de perto esses dois anos. Cada um me transmitiu conhecimentos distintos, pelos quais, sou grata. Riene foi um dos grandes apoios que tive nos momentos mais difíceis, Suzi e Pitanga com sua parceria, uma agitada, outra tranquila, mostrando outros modos de ver o mundo.

À Fena e à Fer, pois foram pessoas presentes e me deram grande apoio nos momentos difíceis que tive no começo do meu intercâmbio acadêmico.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de mestrado (Código de Financiamento 001), sem a qual, seria improvável a realização do mestrado.

À Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul – FAPERGS, pelo apoio financeiro via Edital 01/2017 – ARD, processo 17/2551-0000894-4.

À Universidade Federal do Rio Grande do Sul (UFRGS), ao Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia (CEPSRM) e ao Laboratório de Geoprocessamento e Análise Ambiental (LAGAM) pela infraestrutura disponibilizada para o desenvolvimento das pesquisas e desta dissertação.

Ao Santander Universidades pela bolsa concedida para realização de um semestre do mestrado na Espanha, a partir de uma parceria com a UFRGS e a Universidad Autónoma de Madrid.

À Universidad Autónoma de Madrid, pois através da parceria existente com o Santander Universidades, permite a vários estudantes a oportunidade de estudar um semestre em uma universidade europeia de renome.

A todos aqueles passaram pela minha vida nestes dois últimos anos e que não foram citados aqui, pois certamente todos tiveram sua medida de contribuição para que eu me tornasse a pessoa que sou atualmente e que pudesse concluir este ciclo que foi o Mestrado em Sensoriamento Remoto.

Meu muito obrigada a todos!

EPÍGRAFE

O saber apresenta-se por si mesmo como um obstáculo ao delírio da onipotência do dinheiro e do utilitarismo.

É bem verdade que tudo se pode comprar. De parlamentares a juízes, do poder ao sucesso, tudo tem seu preço.

Mas não o conhecimento: o preço a ser pago para conhecer é de outra natureza.

Nem mesmo uma assinatura num cheque em branco poderá nos dar a permissão de adquirir mecanicamente aquilo que é fruto exclusivo de um esforço individual e de uma paixão que não se esgota.

Resumindo, ninguém, senão nós mesmos, poderá realizar o cansativo percurso que nos permitirá aprender.

Sem grandes motivações interiores, o título de maior prestígio adquirido com o dinheiro não trará nenhum verdadeiro conhecimento, não favorecerá nenhuma autêntica metamorfose do espírito.

Nuccio Ordine
A utilidade do inútil

RESUMO

Os movimentos de massa são fenômenos naturais que, ao atingir um conjunto social, pode vir a causar danos e prejuízos. No Brasil, foram os principais desencadeantes de desastres com registro de óbitos entre os anos de 1991 e 2012. Nesse sentido, o mapeamento de áreas suscetíveis a movimentos de massa atua como uma importante ferramenta para os gestores locais, a fim de localizar as áreas mais suscetíveis e a pensar estratégias para mitigar riscos. As redes neurais artificiais (RNA) e o *Random Forest* (RF) se destacam na modelagem e mapeamento de suscetibilidade por sua elevada acurácia, capacidade de aprendizagem e generalização dos resultados. Assim, este estudo teve como objetivo analisar diferentes técnicas para modelar e mapear áreas suscetíveis a movimentos de massa na Bacia Hidrográfica do Rio Rolante. Foram utilizados RNA e RF, considerando como dados de entrada sete atributos do terreno extraídos dos modelos digitais de elevação (MDE) ALOS-PALSAR (AP) e ASTER (AS): elevação, declividade, curvaturas longitudinal e vertical, relevo sombreado, fator LS e profundidade do vale. As amostras foram compostas de áreas de ocorrência e de não ocorrência de movimentos de massa. Amostras de não ocorrência foram obtidas a partir de duas áreas amostrais distintas (restritiva/abrangente) e de ocorrência foram extraídas por elaboração de inventário de cicatrizes. Foram realizados processos de reamostragem do conjunto de treinamento, a fim de diminuir o tempo e a exigência de capacidade do processamento. De modo geral, os atributos do terreno que apresentaram maior importância foram: elevação, declividade, fator LS e profundidade do vale. Em contrapartida, os menos relevantes foram as curvaturas. Ressalta-se que quando há dois parâmetros com informações similares, como declividade e fator LS, a tendência é que um apresente maior importância, enquanto o outro apresente um valor menor de relevância. Foi observado que a restrição de área para coleta de amostras aleatórias de não ocorrência pode afetar a capacidade de generalização do modelo. Pode-se concluir que a redução do conjunto amostral de treinamento diminui o tempo de processamento, sem interferir significativamente na acurácia. Todas as configurações de modelos resultaram em acurácias entre 0,88 e 0,94, demonstrando que os modelos RNA e RF, combinados com os MDEs AP e AS, atendem ao objetivo de identificar áreas suscetíveis a movimentos de massa.

Palavras-chave: **ALOS-PALSAR. ASTER GDEM. Redes Neurais Artificiais. Random Forest. Bacia Hidrográfica do Rio Rolante.**

RESUMEN

Los movimientos de masa son fenómenos naturales que, al alcanzar un conjunto social, pueden causar daños y perjuicios. En Brasil, han sido los principales responsables por la ocurrencia de desastres con registro de muertes entre los años 1991 y 2012. En este sentido, el mapeo de áreas susceptibles a movimientos de masa actúa como una importante herramienta para los gestores locales, a fin de ubicar las áreas más susceptibles y pensar en estrategias para mitigar los riesgos. Redes neuronales artificiales (RNA) y *Random Forest* (RF) se han destacado en la modelización y mapeo de susceptibilidad debido a su elevada precisión, capacidad de aprendizaje y generalización de los resultados. Este estudio ha objetivado analizar diferentes técnicas para modelizar y mapear áreas susceptibles a movimientos de masa en la cuenca del río Rolante. Han sido utilizados RNA y RF, y considerados como datos de entrada siete atributos del terreno extraídos de los Modelos Digitales de Elevación (MDE) ALOS-PALSAR (AP) y ASTER (AS): elevación, declividad, curvas horizontales y verticales, relieve sombreado, factor-LS y profundidad del valle. Las muestras han sido compuestas por áreas de ocurrencia y de no ocurrencia de movimientos de masa. Muestras de no ocurrencia han sido obtenidas por dos áreas de muestreo distintas (restrictiva/amplia) y de ocurrencia han sido extraídas por elaboración de inventario de cicatrices. Han sido realizados dos procesos de reducción de muestreo para el conjunto de entrenamiento, para disminuir el tiempo y la exigencia de capacidad de procesamiento. En general, los atributos de terreno que presentaron más gran importancia han sido: elevación, declividad, factor-LS y profundidad del valle, los menos pertinentes han sido las curvas. Destacase que cuando hay dos parámetros con informaciones similares, como la declividad y el factor-LS, la tendencia es que uno de ellos presente más gran importancia, mientras tanto el otro presente valor menor de pertinencia. Se ha observado que el área restrictiva para la selección de muestras aleatorias de no ocurrencia puede afectar la capacidad de generalización del modelo. Se concluye que la reducción del conjunto de muestreo de entrenamiento disminuye el tiempo de procesamiento, sin afectar de manera significativa la precisión. Todas las configuraciones de modelos resultaron en precisiones (AUC) entre 0,88 y 0,94, demostrando que los modelos RNA y RF, sumados a los MDEs AP y AS, cumple el objetivo de identificar las áreas susceptibles a movimientos de masa.

Descripción: **ALOS-PALSAR. ASTER GDEM. Redes Neuronales Artificiales. Random Forest. Cuenca del Río Rolante.**

ABSTRACT

Landslides are one of the main phenomena responsible for natural disasters in Brazil. In order to reduce the risks of disasters, Geographic Information System (GIS) can help in the spatial planning of susceptible areas. Artificial neural networks (RNA) and Random Forest (RF) stand out in the susceptibility modeling due to their high accuracy, learning and generalization capacity of the results. Thus, this study aimed to analyze different techniques for mapping landslides susceptible areas in the Rolante River Basin. It were considered as input data seven terrain attributes extracted from ALOS-PALSAR (AP) and ASTER (AS) digital elevation models: elevation, slope, longitudinal and vertical curvatures, shaded relief, LS factor and valley depth. The samples were composed of occurrence and non-occurrence landslides areas. Non-occurrence samples were obtained from restrictive and wide sample areas, while occurrence samples were extracted by scars inventory. Sampling reduction processes were performed for training set in order to reduce the time and the processing capacity requirement. In general, most important attributes were: elevation, slope, LS factor and valley depth. However, the less relevant were the curvatures. It is emphasized that when there are two parameters with similar information, such as slope and LS factor, the tendency is that one presents greater importance while the other presents a lower relevance. It was observed that the non-occurrence sampling area restriction might affect the model generalization capacity. It can be concluded that the training sample set reduction decreases the processing time, without significantly change the accuracy. All model configurations resulted in accuracy between 0.88 and 0.94, demonstrating that RNA and RF models, combined with AP and AS DEM, assist the purpose of identifying landslide susceptible areas.

Keywords: ALOS-PALSAR. ASTER GDEM. Artificial Neural Network. Random Forest. Rolante River Basin.

LISTA DE FIGURAS

Figura 1.1. Material remanescente dentre aqueles deslocados desde a encosta. Fonte: PATRAM Canela, 13 de janeiro de 2017.	3
Figura 1.2. Cicatrizes de movimentos de massa na BHRR. Fonte: 2º Pelotão de Polícia Ambiental de Canela (PATRAM) de Canela, RS, 13 de janeiro de 2017.....	8
Figura 1.3. Cicatrizes de movimentos de massa na BHRR, evento de 05/01/2017. Fonte: 2º Pelotão de Polícia Ambiental de Canela (PATRAM) de Canela, RS, 13 de janeiro de 2017....	9
Figura 1.4. Identificação de cicatrizes a partir de imagem <i>Sentinel 2</i> (R1G2B3).	9
Figura 1.5. Lacunas existentes no MDE SRTM para a área da BHRR.	10
Figura 2.1. Localização das cicatrizes de movimentos de massa, Bacia Hidrográfica do Rio Rolante - RS.	15
Figura 2.2. Fluxograma da metodologia de modelagem a partir de RNA.....	16
Figura 2.3. Áreas amostrais e variação da elevação: A) Área amostral 1; B) Área amostral 2; C) Variação do valor da elevação, de acordo com a área amostral.	18
Figura 2.4. Inventário de Cicatrizes de Movimentos de Massa.....	19
Figura 2.5. Gráficos da distribuição dos valores mínimos, médios e máximos dos atributos do terreno em áreas de ocorrência e de não ocorrência de movimentos de massa.	20
Figura 2.6. Mapas de suscetibilidade a movimentos de massa de acordo com o modelo de maior AUC de cada RNA: (A) Área amostral 1, reamostragem 50; (B) Área amostral 1, reamostragem 80; (C) Área amostral 2, reamostragem 50; (D) Área amostral 2, reamostragem 80.....	23
Figura 3.1. Localização das cicatrizes de movimentos de massa, Bacia Hidrográfica do Rio Rolante, RS.....	30
Figura 3.2 Cicatrizes de movimentos de massa resultantes do evento extremo de precipitação de 05 de janeiro de 2017. Fonte: 2º Pelotão de Polícia Ambiental de Canela (PATRAM) de Canela, RS, 13 de janeiro de 2017.....	31
Figura 3.3. Fluxograma da metodologia utilizando dois modelos digitais de elevação.....	32
Figura 3.4. Área amostral restritiva (área 1) e abrangente (área 2) para coleta de amostras de não ocorrência de movimentos de massa.	35
Figura 3.5. Níveis de suscetibilidade conforme o melhor modelo para cada conjunto de dados.	42

Figura 3.6. Mapa de diferença entre os dois mapas de maior acurácia.	44
Figura 3.7. Média das imagens resultantes para cada configuração de modelo e método de amostragem.....	45
Figura 3.8. Desvio padrão das imagens resultantes para cada configuração de modelo e método de amostragem.....	46

LISTA DE TABELAS

Tabela 2.1. Importância dos atributos de entrada em cada modelo de RNA selecionado. Importância dos atributos de entrada em cada modelo de RNA selecionado.	21
Tabela 2.2. Acurácia dos modelos para os diferentes conjuntos amostrais.....	22
Tabela 3.1. Importância dos atributos de terreno nos modelos de RNA.	38
Tabela 3.2. Importância dos atributos do terreno no modelo RF.	38
Tabela 3.3. Acurácia (AUC) dos modelos RNA e RF.....	40

LISTA DE ABREVIATURAS E SIGLAS

- AD – Árvore de Decisão
- AP – ALOS-PALSAR
- AG – ASTER GDEM
- AUC – *Area Under the ROC Curve*
- BHRR – Bacia Hidrográfica do Rio Rolante
- CVH – Curvatura Horizontal
- CVV – Curvatura Vertical
- DCL – Declividade
- ELV – Elevação
- FLS – Fator LS
- LS – *Slope Length and Steepness Factor*
- MDE – Modelo Digital de Elevação
- PRV – Profundidade do Vale
- RF – *Random Forest*
- RLS – Relevo Sombreado
- ROC – *Receiver Operating Characteristics*
- RNA – Redes Neurais Artificiais
- SIG – Sistemas de Informações Geográficas
- SR – Sensoriamento Remoto

SUMÁRIO

CAPÍTULO 1 : INTRODUÇÃO	1
1.1 APRESENTAÇÃO.....	1
1.2 CONTEXTO CIENTÍFICO E JUSTIFICATIVA.....	1
1.3 TÉCNICAS DE APRENDIZADO DE MÁQUINA	4
1.4 SELEÇÃO DE DADOS	8
1.5 OBJETIVOS E HIPÓTESE	11
CAPÍTULO 2 : MODELAGEM DE ÁREAS SUSCETÍVEIS A MOVIMENTOS DE MASSA A PARTIR DE REDES NEURAS ARTIFICIAIS.....	12
2.1 INTRODUÇÃO.....	13
2.2 MATERIAIS E MÉTODOS.....	14
2.2.1 Área de Estudo.....	14
2.2.2 Métodos	15
2.3 RESULTADOS E DISCUSSÕES.....	19
2.4 CONCLUSÕES	24
2.5 AGRADECIMENTOS	24
2.6 REFERÊNCIAS	24
CAPÍTULO 3 : MODELAGEM DE ÁREAS SUSCETÍVEIS A MOVIMENTOS DE MASSA: AVALIAÇÃO COMPARATIVA DE TÉCNICAS DE AMOSTRAGEM, APRENDIZADO DE MÁQUINA E MODELOS DIGITAIS DE ELEVAÇÃO.....	27
3.1 INTRODUÇÃO.....	28
3.2 MATERIAIS E MÉTODOS.....	30
3.2.1 Área de Estudo.....	30
3.2.2 Métodos	32
3.2.3 Atributos Morfométricos	32
3.2.4 Inventário de Cicatrizes de Movimentos de Massa	33
3.2.5 Amostragem	34
3.2.6 Modelagem e Análise Espacial das Áreas Suscetíveis	36
3.3 RESULTADOS E DISCUSSÕES.....	37
3.4 CONCLUSÕES	47
3.5 AGRADECIMENTOS	48
3.6 REFERÊNCIAS	48

CAPÍTULO 4 : CONSIDERAÇÕES FINAIS51
REFERÊNCIAS53

CAPÍTULO 1 : INTRODUÇÃO

1.1 APRESENTAÇÃO

Esta dissertação desenvolveu e analisou mapeamentos de suscetibilidade, a partir do uso de modelo digital de elevação (MDE) e técnicas de aprendizagem de máquina, com o objetivo de verificar a influência da resolução espacial dos MDEs, dos métodos de modelagem, das áreas amostrais e dos processos de reamostragem quanto à acurácia do mapa final.

Este estudo está dividido em quatro capítulos. O primeiro capítulo é introdutório, no qual são apresentados o contexto científico, uma breve revisão teórica dos métodos de aprendizagem de máquinas utilizados: *random forest* (RF) e redes neurais artificiais (RNA), e dos objetivos propostos. O segundo capítulo trata sobre modelagem de áreas suscetíveis a movimentos de massa a partir de redes neurais artificiais, considerando sete atributos topográficos extraídos a partir do MDE ALOS-PALSAR (elevação, declividade, curvaturas longitudinal e vertical, relevo sombreado, fator LS – *Slope Length and Steepness Factor*, e profundidade de vale), selecionados a partir de duas áreas amostrais e dois métodos de reamostragem.

O terceiro capítulo apresenta a modelagem de suscetibilidade a partir do RF e RNA. Para tal, foram extraídos os sete atributos do terreno supracitados dos MDEs ALOS-PALSAR (AP) e ASTER GDEM (AG). Foram comparados os dois métodos, a influência da resolução espacial dos dois MDEs, além do uso de conjuntos de amostras que consideram duas áreas amostrais e dois processos de reamostragem. Finalmente, no quarto capítulo foram feitas as considerações finais e conclusões acerca dos resultados apresentados e discutidos neste estudo.

1.2 CONTEXTO CIENTÍFICO E JUSTIFICATIVA

Os desastres relacionados a fenômenos naturais extremos são responsáveis por perdas e danos nos âmbitos social, econômico e ambiental, em escala global (TOMINAGA; SANTORO; AMARAL, 2009). Nos últimos anos, houve um aumento quanto à recorrência e à intensidade dos impactos gerados por desastres, afetando cada vez mais pessoas ao redor do mundo (CEPED UFSC, 2013; IFRC, 2016).

Uma das principais causas é o rápido processo de urbanização, a partir do qual houve um crescimento de 30% para 54% da população mundial residindo em áreas urbanas entre os anos de 1950 e 2014 (UNITED NATIONS, 2014). Este processo de urbanização acelerado resultou na expansão das cidades, muitas vezes de forma desordenada, com aumento de moradias em áreas impróprias para ocupação, como planícies de inundação e terrenos com altos declives, sem a infraestrutura adequada, causando maior pressão ao meio ambiente (TOMINAGA; SANTORO; AMARAL, 2009).

As intervenções antrópicas resultantes desse processo, como modificações de vertentes, impermeabilização do solo, remoção de vegetação, entre outros, podem ser consideradas como fatores condicionantes para a ocorrência de eventos adversos (ZÊZERE, 2005). Com isso, os eventos que ocorrem devido às características do terreno e o seu grau de suscetibilidade, como inundações e movimentos de massa, resultam em desastres, causando danos materiais e perdas de vidas humanas (CEPED/RS - UFRGS, 2016; QUEVEDO, 2016).

Os movimentos de massa são deslocamentos descendentes de material de encostas, associados à dinâmica de vertentes, que estão sob influência direta da gravidade, podendo ter como agente catalisador as intervenções antrópicas, e como agentes detonadores terremotos, degelo ou chuvas intensas (ZÊZERE, 2005; CHRISTOPHERSON, 2012).

No Brasil, entre os anos de 1991 e 2012, a ocorrência de movimentos de massa resultou no segundo maior número de óbitos em desastres decorrentes de fenômenos naturais (CEPED UFSC, 2013). Na Bacia Hidrográfica do Rio Rolante (BHRR), localizada na porção nordeste do estado do Rio Grande do Sul, após um evento de precipitação torrencial concentrada, houve uma série de movimentos de massa. Neste evento, foram deslocadas grandes quantidades de materiais para o fundo do vale do rio Mascarada (Figura 1.1), afluente do rio Rolante, em janeiro de 2017 (SEMA, 2017). Por se tratar de um vale encaixado, houve o barramento das águas do rio, com posterior rompimento e ocorrência de enxurrada. Dessa forma, as águas escoaram de maneira rápida, atingindo e causando estragos no município de Rolante.

Para minimizar esses impactos, faz-se necessário que a gestão de risco de desastres envolva o conhecimento das ameaças às quais determinada população está exposta (CEPED/RS - UFRGS, 2016). Nesse sentido, o mapeamento de suscetibilidade apresenta-se como uma ferramenta importante para o planejamento urbano, a fim de minimizar os possíveis impactos negativos advindos da ocorrência de fenômenos naturais. Desta forma, as tecnologias de Sistema de Informações Geográficas (SIG) e de Sensoriamento Remoto (SR) são identificadas como ferramentas favoráveis para a otimização de análises referentes aos movimentos de

massa, em todas as etapas de estudo, incluindo a detecção, o monitoramento e a espacialização das áreas suscetíveis, até mesmo a previsão de recorrência da ameaça (TRALLI *et al.*, 2005).



Figura 1.1. Material remanescente dentre aqueles deslocados desde a encosta. Fonte: PATRAM Canela, 13 de janeiro de 2017.

Os mapeamentos de suscetibilidade têm sido realizados considerando os dados morfométricos de cicatrizes de deslizamentos já ocorridos (RIFFEL; RUIZ; GUASSELLI, 2016). O desenvolvimento dos métodos quantitativos fez com que os modelos estatísticos, bivariados e multivariados, fossem os mais utilizados para esse tipo de análise, uma vez que possibilitam determinar o peso das variáveis e, também, a validação dos resultados (MELO; ZÊZERE, 2017). Nesse sentido, diversos estudos utilizam métodos de aprendizado de máquina, com o intuito de aprimorar o mapeamento (KAWABATA; BANDIBAS, 2009; SAITO; NAKAYAMA; MATSUYAMA, 2009; PRADHAN; LEE, 2010; DOU *et al.*, 2015; ADITIAN; KUBOTA; SHINOHARA, 2018; CHU *et al.*, 2018). A seguir, serão explicados os métodos de aprendizado de máquinas, com foco para aqueles utilizados nessa dissertação: *Random Forest* e Redes Neurais Artificiais.

1.3 TÉCNICAS DE APRENDIZADO DE MÁQUINA

O aprendizado de máquina é um conjunto de métodos de análise de dados, da inteligência artificial, focado em processos de aprendizagens computacionais, que utilizam dados e teorias pré-estabelecidos para obter conhecimento de forma induzida (JENSEN, 2005). Abrange diferentes técnicas inspiradas no modo de aprendizagem humana, utilizadas para descobrir e descrever padrões, a fim de aplicá-los em previsões para futuras ocorrências (WITTEN; FRANK, 2005). São utilizadas sempre duas amostras baseadas no conhecimento real (*e.g.* locais com e sem cicatrizes de movimentos de massa, previamente conhecidos), com seleção de um percentual para o treinamento e aprendizagem do modelo, enquanto que o restante é aplicado na validação, o que permite avaliar o desempenho do modelo utilizado (WITTEN; FRANK, 2005).

As técnicas de aprendizado de máquina podem ser divididas em dois grandes grupos: caixa transparente (*transparent box*) e caixa preta (*black box*), sendo que a primeira possui uma estrutura visível do padrão, enquanto a segunda possui uma estrutura oculta (WITTEN; FRANK, 2005). Como exemplos, o método de Árvore de Decisão (*Decision Tree*) possui uma estrutura conhecida, enquanto as Redes Neurais Artificiais (*Artificial Neural Network*) possuem camadas escondidas.

Árvore de decisão (AD) é um método baseado em regras e condições, em que se é dada uma hipótese e existem duas possibilidades para respondê-la. Possui uma estrutura hierárquica, em que a hipótese representa o tronco de uma árvore, cada regra simboliza um galho e cada condição, uma folha. Pode ser utilizado com dados discretos (também chamada de aprendizagem de classificação) ou contínuos (regressão) (JENSEN, 2005). Diversos estudos vêm utilizando o método de AD para aplicações com referência espacial, como a geração de mapas de suscetibilidade a movimentos de massa (SAITO; NAKAYAMA; MATSUYAMA, 2009; YEON; HAN; RYU, 2010; PRADHAN, 2013).

Saito, Nakayama e Matsuyama (2009) utilizaram AD para mapeamento de suscetibilidade, a partir de imagens Landsat-5 TM e ASTER GDEM, em que a declividade e a geologia foram os parâmetros que apresentaram maior importância. O índice Kappa (K) resultou em 0,61, com um intervalo de confiança de 99%, demonstrando que se trata de um modelo apropriado para o estudo. Yeon, Han e Ryu (2010) utilizaram mapas topográfico, pedológico e de cobertura florestal, além de um MDE de 5x5m. Foram considerados 590 movimentos de massa, sendo metade utilizada para treinamento e a outra metade para

validação, com um intervalo de confiança de 95%. Verificou-se forte relação, entre os movimentos de massa e declividades entre 20 e 39°. Pradhan (2013) comparou três métodos: AD, *Support Vector Machine* (SVM) e *Adaptive Neuro-Fuzzy Inference System* (ANFIS). O inventário de deslizamentos continha 113 cicatrizes, 8.403 píxeis, e foi dividido em dois de forma aleatória para treinamento e validação. Foram adquiridos os parâmetros morfométricos a partir de um MDE de 20 m. A AD apresentou acurácia sutilmente maior: 83,07 (82,80 para ANFIS e 81,46 para SVM). Os resultados demonstram que esse método é adequado para a temática, além de desempenhar um importante papel no planejamento ambiental, uma vez que o custo é reduzido.

O presente estudo utilizou o algoritmo *Random Forest* (RF) para o desenvolvimento de mapas de suscetibilidade a partir de AD. O RF é um método *ensemble-learning*, proposto por Breiman (2001), que seleciona de maneira aleatória percentuais dos dados de entrada e os separa em subconjuntos (nós), escolhendo o valor do atributo preditivo que minimiza o nó da impureza. Forma uma coleção de classificadores estruturados em árvore, e para cada sub-amostra é construída uma AD diferente e o valor estimado é referente à média da predição das árvores (TAALAB; CHENG; ZHANG, 2018). Como critério de partição dos nós das árvores, foi selecionado o índice de *Gini Impurity* (ΔI), descrito na equação (1).

$$\Delta I = G(\tau) - p_l G(\tau_l) - p_r G(\tau_r) \quad (1)$$

Em que: ΔI quantifica a diminuição do índice de Gini, após a divisão (impureza Gini); $G(\tau)$ é o valor do índice de Gini para o nó, antes da divisão; p_r é a probabilidade das predições das classes para o nó direito (τ_r), p_l é a probabilidade das predições das classes para o nó esquerdo τ_l ; $G(\tau_r)$ é o valor do índice de Gini para o nó direito (τ_r) e $G(\tau_l)$ é o valor do índice de Gini para o nó esquerdo.

O RF tem sido utilizado por diversos autores (CHEN *et al.*, 2017; CHU *et al.*, 2018; KIM *et al.*, 2018; TAALAB; CHENG; ZHANG, 2018). Chu *et al.* (2018) utilizaram os modelos RF e *multivariate adaptive regression spline* (MARSpline) para mapear as áreas suscetíveis na cidade de Gero, no Japão. Como resultado, RF apresentou maior acurácia. Kim *et al.* (2018) utilizaram RF para mapeamento de suscetibilidade em Pyeong-Chang, na Coreia do Sul, o qual

apresentou AUC acima de 79%, demonstrando-se como método adequado para este tipo de estudo. Taalab, Cheng e Zhang (2018) mapearam as áreas suscetíveis em larga escala, na região do Piemonte, na Itália, uma região heterogênea e com diferentes tipos de movimentos de massa a partir de RF e obtiveram uma acurácia de 85%. Chen *et al.* (2017) comparam o desempenho de três modelos: *logistic model tree*, RF e *classification and regression tree*. A acurácia do RF foi maior que dos demais modelos (0,84%), demonstrando, assim, bom desempenho para este tipo de mapeamento.

Além de RF, este estudo utilizou as Redes Neurais Artificiais (RNA) para modelagem das áreas suscetíveis. As RNA são um método utilizado para classificação, modelagem matemática, simulação e predição, em que existem duas ou mais camadas de processamento, sendo uma de entrada, uma de saída e, as demais, caso existam, são camadas ocultas (IYODA, 2000). É inspirado no processo de pensamento humano, em que os neurônios são pontos de conexão para processar a informação recebida, formando estruturas próprias (JENSEN, 2005).

O propósito das RNA é construir um modelo de atribuição de pesos para os dados de entrada, a fim de prever dados de saída. A etapa de treinamento foi realizada pelo método retropropagativo (*back-propagation*) com múltiplas camadas (RUMELHART; HILTON; WILLIAMS, 1986), usando a Regra Delta (WIDROW; HOFF, 1960) para atualização dos pesos sinápticos da rede. Os valores dos atributos de entrada foram escalonados a partir de transformações lineares (Equação 2). A função de ativação utilizada nas camadas do modelo foi a sigmoideal. Uma série de validação cruzada paralela à série de treinamento foi utilizada para evitar o superajustamento do modelo. As variáveis de entrada foram excluídas de maneira sistemática, a partir do *Relative Contribution Index* (OLIVEIRA; PEDROLLO; CASTRO, 2015). A saída do modelo foi calculada a partir da função apresentada na Equação 3.

$$\frac{(y_t - p_o)}{s_o} = RNA\left(\frac{(x_t - p_i)}{s_i}\right) \quad (2)$$

Em que: x_t e y_t são variáveis de entrada e saída, respectivamente; s_o e p_o são parâmetros de escala e posição das saídas do modelo; s_i e p_i são os parâmetros de escala e posição das entradas do modelo.

$$RNA\left(\frac{(x_t - p_i)}{s_i}\right) = f_o\left(\sum_h w_o f_h\left(\sum_i w_h\left(\frac{x_t - p_i}{s_i}\right) + b_h\right) + b_o\right) + e_o \quad (3)$$

Em que: w_h , b_h , f_h , w_o , b_o e f_o são pesos sinápticos (w), tendências (b) e funções de ativação (f), respectivamente, das camadas ocultas (h) e de saída (o), enquanto e_o é o erro esperado na camada de saída.

Diversos estudos utilizaram as RNA para mapeamento de áreas suscetíveis (ERMINI; CATANI; CASAGLI, 2005; KAWABATA; BANDIBAS, 2009; PRADHAN; LEE, 2010). Ermini, Catani e Casagli (2005) modelaram a suscetibilidade a movimentos de massa na bacia de Riomaggiore (Itália), a partir da aplicação de RNA. Foram utilizados 19 variáveis de entrada e um inventário com 34 cicatrizes. Os resultados apresentaram acurácia entre 68% e 73%. Pradhan e Lee (2010) aplicaram o estudo em Cameron Highlands, na Malásia. Utilizaram um inventário de movimentos de massa com 324 cicatrizes, numa área de 293 km². Os atributos de terreno foram obtidos a partir de um MDE de 10 m. Foi utilizado o algoritmo retropropagação e uma estrutura de 10 entradas, 22 neurônios e 2 saídas. O parâmetro mais importante foi a declividade e a acurácia foi de 83%. Kawabata e Bandibas (2009) utilizaram o algoritmo retropropagação para modelar áreas de suscetíveis em Chuetsu, município de Niigata, no Japão, a partir das RNA. Foram considerados dois MDEs, seis dados de entrada, 40 neurônios na camada oculta e dois dados de saída. A acurácia de ambos os mapas foi superior a 90%.

Verifica-se que os modelos RF e RNA são utilizados por diversos autores para modelagem e mapeamento de áreas suscetíveis a movimentos de massa e que ambos apresentam resultados satisfatórios, resultando em mapas com elevada acurácia. Entretanto, a qualidade dos resultados está diretamente relacionada à qualidade dos dados de entrada. Assim, os conjuntos amostrais a serem modelados devem passar por um processo de análise prévio, para garantir a precisão do mapeamento. Como as amostras para modelar áreas suscetíveis devem conter pontos de ocorrência e de não ocorrência de movimentos de massa, as áreas de coleta de amostras de não ocorrência devem ser analisadas, pois podem influenciar significativamente no processo de modelagem, algo pouco discutido na literatura.

Além disso, com a crescente disponibilização de MDEs gratuitos, deve-se verificar como a resolução espacial se comporta quanto ao tipo de estudo. MDEs com melhores resoluções acarretam num aumento do número de píxeis nas amostras, o que pode acabar exigindo maior capacidade de processamento computacional na etapa de treinamento dos modelos. Nesse

sentido, métodos de reamostragem que diminuam o conjunto amostral de treinamento de modo a manter a representatividade podem vir a ser uma alternativa.

1.4 SELEÇÃO DE DADOS

O evento de precipitação extrema ocorrido em 05 de janeiro de 2017 na BHRR resultou em ao menos 297 cicatrizes de movimentos de massa. Como característica principal, as cicatrizes apresentaram em sua maioria largura reduzida, inferior a 15 metros, e comprimento da ordem de 200 metros (Figura 1.2). A seleção de dados para este estudo esteve diretamente relacionada com a distribuição gratuita de imagens que se adequassem às características da área de estudo.

Primeiramente foi realizado o mapeamento das cicatrizes de movimentos de massa a partir de imagens *Sentinel 2*, porém, devido à resolução das imagens (10 m), somente as maiores cicatrizes, semelhantes às da Figura 1.3, eram passíveis de serem identificadas (Figura 1.4). Assim, com a atualização das imagens disponibilizadas pelo programa *Google Earth*, dos dias 06 e 24 de agosto de 2017 – imagens mais próximas temporalmente do evento, passou-se a utilizar tais imagens para confeccionar o inventário de cicatrizes necessário para a modelagem de áreas suscetíveis.



Figura 1.2. Cicatrizes de movimentos de massa na BHRR. Fonte: 2º Pelotão de Polícia Ambiental de Canela (PATRAM) de Canela, RS, 13 de janeiro de 2017.



Figura 1.3. Cicatrizes de movimentos de massa na BHRR, evento de 05/01/2017. Fonte: 2º Pelotão de Polícia Ambiental de Canela (PATRAM) de Canela, RS, 13 de janeiro de 2017.

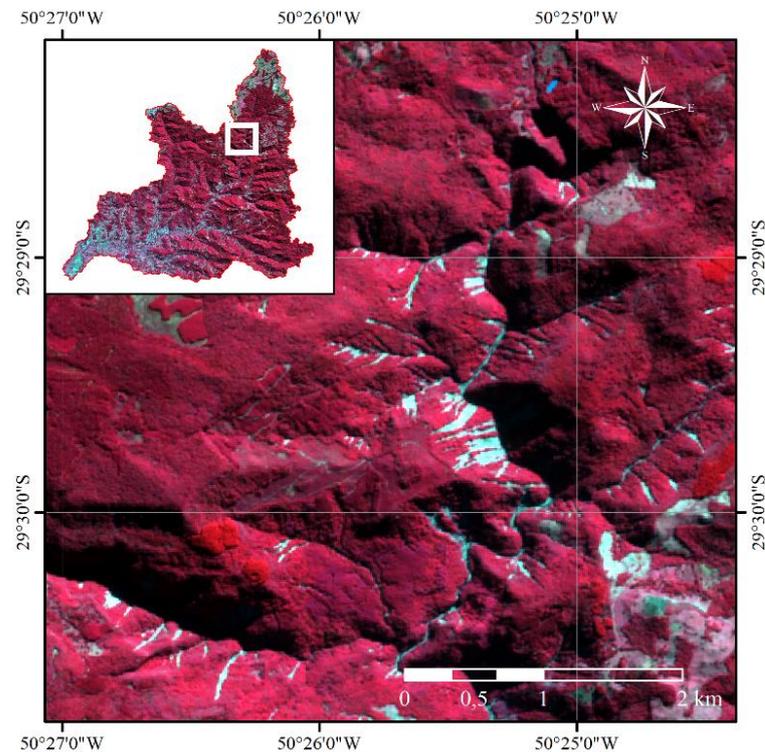


Figura 1.4. Identificação de cicatrizes a partir de imagem *Sentinel 2* (R1G2B3).

Quanto à comparação entre MDEs, optou-se pelo MDE AP devido à resolução espacial (12,5 m), que é uma das maiores dentre aquelas disponibilizadas de forma gratuita. Foi

selecionado o MDE SRTM, com resolução de 30 m para realizar este paralelo, todavia percebeu-se que para a área da BHRR este MDE apresenta lacunas, ou seja, áreas sem informações (Figura 1.5). Desta forma, optou-se por utilizar o MDE AG, o qual apresenta resolução espacial igual ao MDE SRTM, 30 metros.

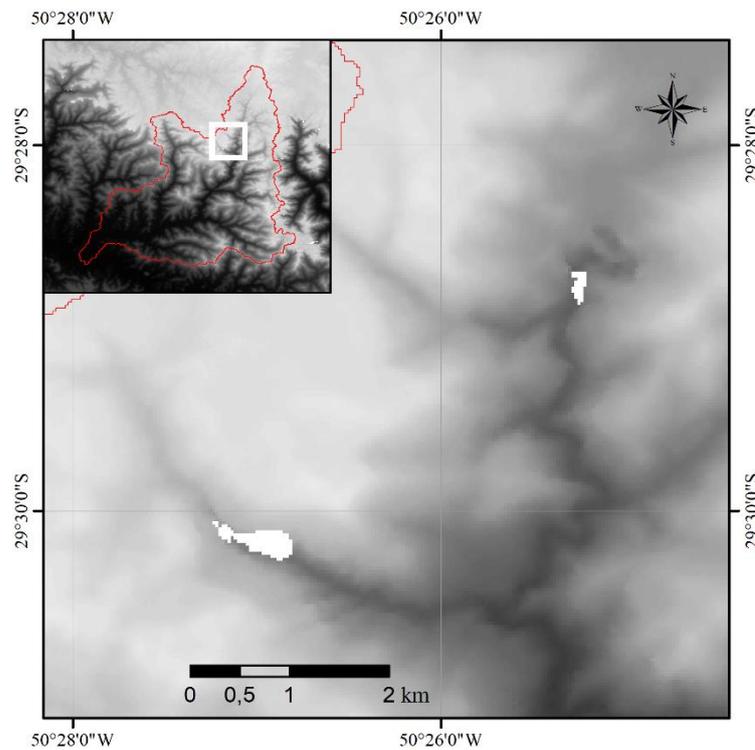


Figura 1.5. Lacunas existentes no MDE SRTM para a área da BHRR.

Definidas as imagens a serem utilizadas no trabalho, passou-se a elaborar os conjuntos amostrais, os quais são os dados de entrada das modelagens. Estes processos estão detalhados nas metodologias encontradas nos capítulos 2 e 3.

1.5 OBJETIVOS E HIPÓTESE

O presente estudo partiu da hipótese de que diferentes métodos de modelagem, bem como o uso de diversos conjuntos amostrais, produzem distintos resultados quanto ao mapeamento de suscetibilidade a movimentos de massa.

Nesse sentido, o objetivo geral foi comparar o desempenho de duas técnicas de aprendizado de máquinas para o mapeamento de suscetibilidade a movimento de massa na BHRR. Para que o objetivo geral fosse alcançado, os seguintes objetivos específicos foram realizados:

- Elaboração de um inventário de cicatrizes de movimentos de massa;
- Definição dos principais parâmetros morfométricos que influenciam na ocorrência de movimentos de massa;
- Análise do uso de diferentes áreas amostrais e métodos de reamostragem;
- Mapeamento das áreas suscetíveis a movimentos de massa a partir do modelo que apresentou melhor desempenho nas Redes Neurais Artificiais e no *Random Forest*;
- Análise a influência de MDEs com diferentes resoluções espaciais;
- Comparação entre as acurácias dos mapas gerados.

CAPÍTULO 2 : MODELAGEM DE ÁREAS SUSCETÍVEIS A MOVIMENTOS DE MASSA A PARTIR DE REDES NEURAS ARTIFICIAIS

RESUMO

Os movimentos de massa são um dos principais fenômenos responsáveis por desastres naturais no Brasil. O mapeamento pode auxiliar no ordenamento territorial das áreas suscetíveis. As redes neurais artificiais (RNAs) se destacam na modelagem e mapeamento de suscetibilidade por sua elevada acurácia, capacidade de aprendizagem e generalização dos resultados. Assim, este estudo teve como objetivo mapear áreas suscetíveis a movimentos de massa, considerando quatro conjuntos amostrais, a partir de um modelo de RNA. Para tal, foi elaborado um inventário de cicatrizes, extraídos parâmetros morfométricos, organizados os conjuntos amostrais, realizados treinamentos, validação dos modelos e reclassificação e espacialização das áreas suscetíveis. As RNAs se mostraram capazes de modelar as áreas suscetíveis, com acurácia no mapeamento próximas ou superiores a 0,9. Foi observado que a restrição de área para coleta de amostras aleatórias de não ocorrência pode afetar a capacidade de generalização do modelo. Pode-se concluir que a redução do conjunto amostral de treinamento diminui o tempo de processamento, sem interferir significativamente na acurácia.

Palavras-chave: Parâmetros Morfométricos; Modelo Digital de Elevação; ALOS-PALSAR; Bacia Hidrográfica do Rio Rolante

ABSTRACT

Landslides are one of the main phenomena responsible for natural disasters in Brazil. The existence of mappings can help in the territorial ordering of these areas. Mapping can help in the spatial planning of susceptible areas. Artificial neural networks (RNAs) stand out in the modeling and mapping of susceptibility by their high accuracy, learning and generalization capacity of the results. Thus, this study aimed to map susceptible areas to landslides, considering four different sample sets, from an RNA model. For this, an inventory of scars was drawn, morphometric parameters were extract, the sample sets were organized, training and validation of the models and reclassification and spatialization of the susceptible areas. The

RNAs were able to model the susceptible areas, with accuracy in the mapping near or greater than 0.9. It was observed that the restriction of area for random sampling of non-occurrence may affect the generalization capacity of the model. It can be concluded that the reduction of the training sample set decreases the processing time, without significantly interfering with the accuracy.

Key words: Morphometric Parameters; Digital Elevation Model; ALOS-PALSAR; Rolante River Basin

2.1 INTRODUÇÃO

Os movimentos de massa foram os desastres naturais responsáveis pelo segundo maior número de óbitos no Brasil, entre 1991 e 2012, atrás somente das enchurradas (CEPED UFSC, 2013). A fim de reduzir os impactos oriundos da ocorrência de eventos extremos, a redução de risco de desastres busca compreender a dinâmica do fenômeno e identificar as áreas suscetíveis à sua ocorrência (ZÊZERE *et al.*, 2017).

O mapeamento de áreas suscetíveis a movimentos de massa é uma ferramenta auxiliar no ordenamento do território (RIFFEL; RUIZ; GUASSELLI, 2016). Entre os modelos utilizados para simular e mapear a suscetibilidade a movimentos de massa, o uso de Redes Neurais Artificiais (RNA) se destaca devido à acurácia e à capacidade de aprendizagem e de generalização dos resultados (KAWABATA; BANDIBAS, 2009; PRADHAN; LEE, 2010).

As RNA são utilizadas para classificação, modelagem matemática, simulação e predição. São formadas por duas ou mais camadas de processamento, sendo uma de entrada, uma de saída e, as demais, caso existam, são camadas escondidas. É inspirado no processo de pensamento humano, em que os neurônios são pontos de conexão para processar a informação recebida, formando estruturas próprias (JENSEN, 2005). Um dos algoritmos mais utilizados é o *Error Back-Propagation*, em que os neurônios são organizados em camadas (entrada, intermediária e saída), com pesos distintos para cada relação entre neurônios. Para que esses pesos sejam ajustados da maneira mais adequada, a modelagem conta com etapa de treinamento antes da validação cruzada e do teste (KAWABATA; BANDIBAS, 2009).

Diversos estudos utilizam a modelagem por RNAs para mapeamento de suscetibilidade a movimentos de massa e os dados de entrada mais comuns são mapas geológicos e geomorfológicos e atributos do terreno, extraídos de Modelo Digital de Elevação (MDE), como elevação, declividade, curvaturas vertical e longitudinal, entre outros (KAWABATA; BANDIBAS, 2009; PRADHAN; LEE, 2010; DOU *et al.*, 2015; ADITIAN; KUBOTA; SHINOHARA, 2018).

Nesse contexto, propõem-se como objetivo mapear áreas suscetíveis a movimentos de massa a partir de um modelo de RNAs, tendo como variáveis de entrada atributos do terreno extraídos do MDE ALOS-PALSAR e quatro conjuntos amostrais diferentes de movimentos de massa, obtidos em função de duas áreas de referência e dois processos de reamostragem.

2.2 MATERIAIS E MÉTODOS

2.2.1 Área de Estudo

A Bacia Hidrográfica do Rio Rolante (BHRR) localiza-se no nordeste do Rio Grande do Sul (Figura 2.1). A área de drenagem da BHRR é de 828 km², e compreende um relevo escarpado, com altitudes entre 152 e 992 m, e declividade média de 13°. O clima é o subtropical muito úmido, com precipitação entre 1700 e 2000 mm anuais, influenciado pelo ingresso de sistemas frontais de precipitação e pela orografia (ROSSATO, 2011).

Um evento extremo de precipitação que ocorreu em 05/01/2017 se caracterizou pela alta concentração temporal e espacial, nas cabeceiras da BHRR, na sub-bacia do rio Mascarada. Dados oficiais subestimaram a precipitação (50 mm), devido à falta de medidores nos locais de ocorrência. De acordo com pluviômetros particulares, instalados em áreas rurais, a precipitação variou entre 90 e 272 mm em diferentes pontos (SEMA, 2017).

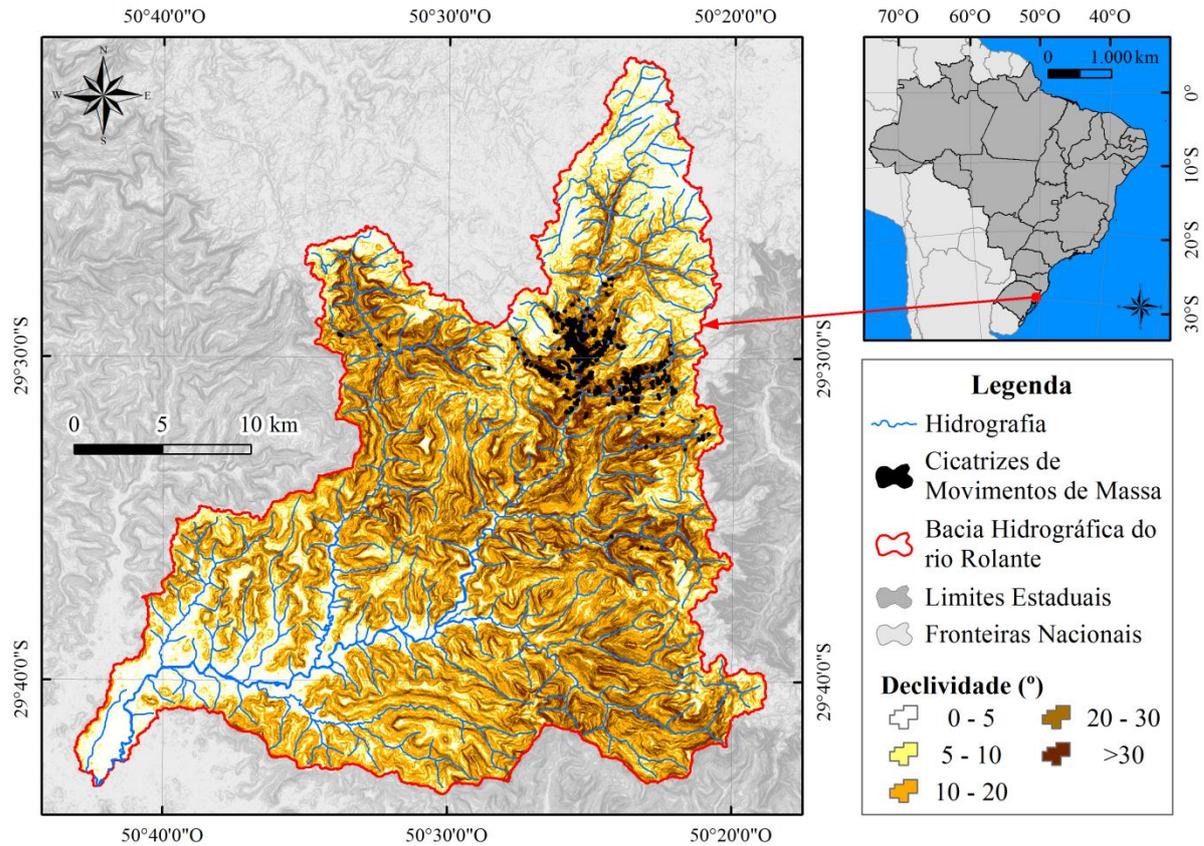


Figura 2.1. Localização das cicatrizes de movimentos de massa, Bacia Hidrográfica do Rio Rolante - RS.

Esse evento extremo durou quatro horas, e resultou em centenas cicatrizes de movimentos de massa. O material deslocado para o fundo do vale formou barreiras naturais no leito dos rios. Inicialmente, houve o bloqueio do fluxo e consequente aumento do nível da água a montante. Em seguida, com o rompimento das barreiras, ocorreu uma enxurrada de grandes proporções a jusante, atingindo o município de Rolante (SEMA, 2017).

2.2.2 Métodos

O estudo envolveu seis etapas (Figura 2.2): i) extração de parâmetros morfométricos do MDE ALOS-PALSAR; ii) pré-processamento e organização do conjunto amostral; iii) treinamento e calibração das RNAs; iv) validação cruzada; v) teste dos modelos, e; vi) reclassificação dos níveis de suscetibilidade em SIG.

Foi utilizado o MDE do satélite ALOS (*Advanced Land Observing Satellite*), sensor PALSAR (*Phased Array type L-band Synthetic Aperture Radar*), disponibilizado a partir do *website Alaska Satellite Facility*, com resolução espacial de 12,5 m. Foram utilizados os softwares *SAGA-GIS* e *ArcGIS*, para extrair sete atributos morfométricos do MDE: elevação (ELV); declividade (DCL); curvatura horizontal (CVH); curvatura vertical (CVV); relevo sombreado (RLS); fator LS (LS); profundidade do vale (PRV).

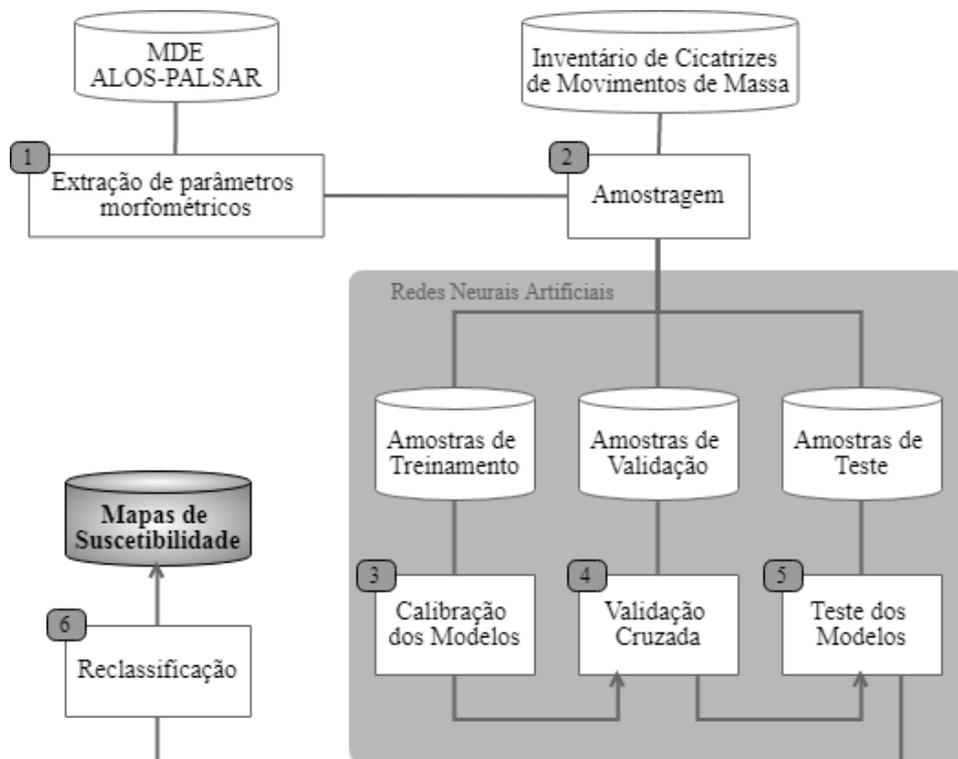


Figura 2.2. Fluxograma da metodologia de modelagem a partir de RNA.

O inventário de cicatrizes é a principal informação de entrada na modelagem de movimentos de massa. Registros de eventos passados têm sido obtidos por análise de imagens aéreas ou orbitais e auxílio de coleta de campo (GARCIA-URQUIA; YAMAGISHI, 2017; LE *et al.*, 2017; ADITIAN *et al.*, 2018). Neste estudo, o inventário foi elaborado a partir da interpretação e delimitação das cicatrizes, em imagens de agosto de 2017 disponíveis no *Google Earth*, e de trabalhos de campo para validação. Os polígonos foram convertidos para pontos, de modo que cada píxel do MDE represente uma amostra do modelo.

Foram consideradas duas áreas para coleta de amostras de não ocorrência a partir de pontos aleatórios externos às cicatrizes (Figura 2.3). O conjunto amostral foi estabelecido por pontos localizados no interior das cicatrizes (Classe=1) e igual número de pontos externos aleatórios (Classe=0). A área 1 é restritiva (125 km²) e concentrada na área com cicatrizes, enquanto a área 2 é mais abrangente (386 km²).

Visando reduzir o tempo de processamento, foi testada uma técnica de reamostragem dos dados para obter conjuntos representativos de menor tamanho para calibração. Esse processo foi realizado a partir de cada parâmetro, sendo reordenados os valores dos atributos e coletada uma amostra de treinamento a cada x amostras, seguindo uma distribuição uniforme. Foram testados dois valores de x (50 e 80), resultando em dois conjuntos para cada área amostral.

O modelo RNA foi escrito no *software MATLAB*, e o treinamento realizado pelo método retropropagativo com múltiplas camadas (RUMELHART; HINTON; WILLIAMS, 1986), com a Regra Delta para atualizar os pesos sinápticos da rede (WIDROW; HOFF, 1960). Os valores dos atributos de entrada foram escalonados a partir de transformações lineares. A função de ativação nas camadas foi a sigmoideal. Uma série de validação cruzada paralela à série de treinamento foi utilizada para evitar o superajustamento do modelo.

Foram testadas dezoito configurações de modelo para cada área e conjunto amostral, variando o conjunto de atributos de entrada. Do modelo mais completo, com todos os atributos, até os mais compactos, em que atributos menos importantes eram excluídos de forma automática. Para cada configuração, foram testados diferentes números de neurônios na camada oculta da rede. Como a inicialização dos pesos é aleatória, foram executadas 15 iterações em cada configuração. Em cada iteração, foi definido um número máximo de 20.000 ciclos de aprendizagem para interrupção, caso o modelo não atinja o limiar de convergência.

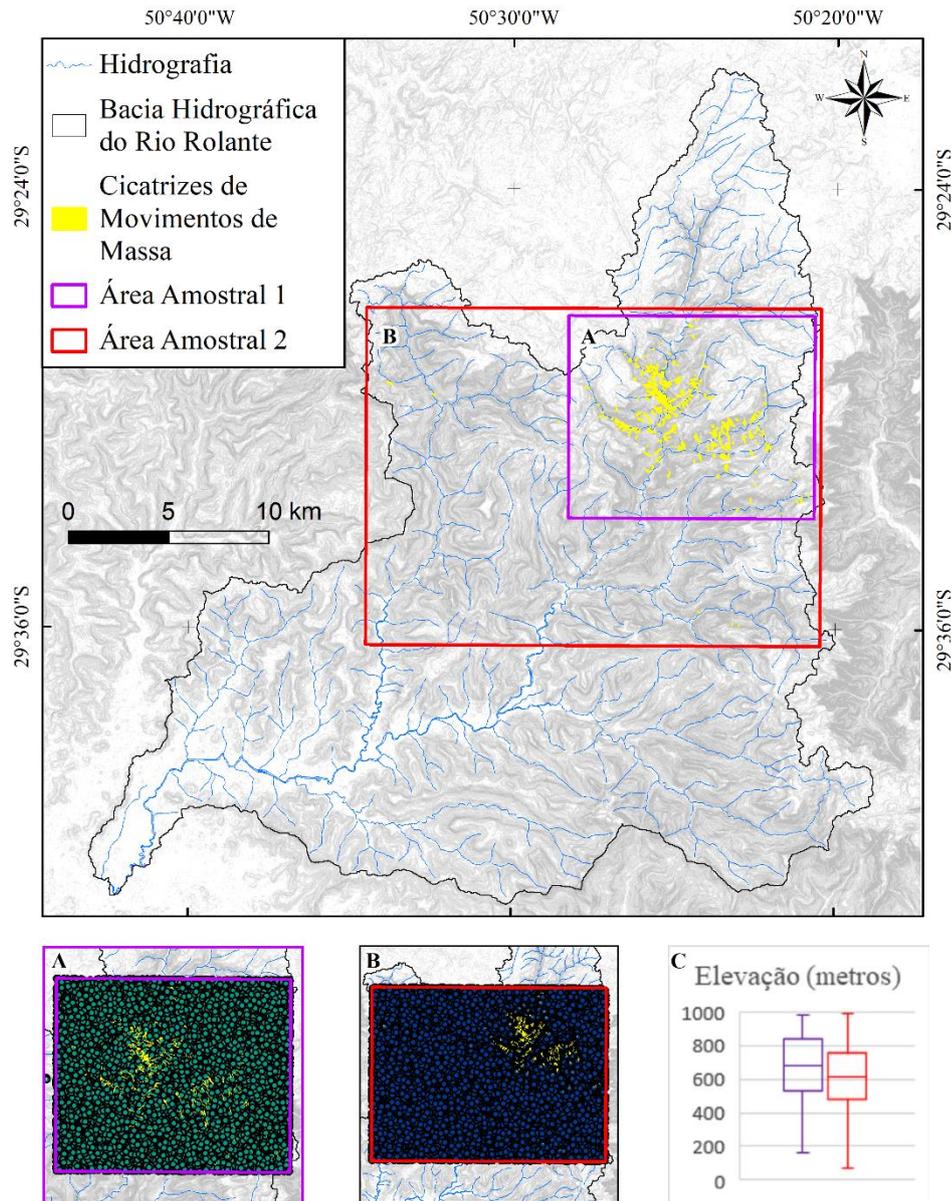


Figura 2.3. Áreas amostrais e variação da elevação: A) Área amostral 1; B) Área amostral 2; C) Variação do valor da elevação, de acordo com a área amostral.

Para validação, o índice de acurácia foi calculado por meio da *Area Under the Receiver Operating Characteristic Curve* (ROC). A curva ROC é uma representação bidimensional do desempenho do modelo. Uma vez que o modelo esteve ajustado, as amostras de teste foram preditas no intervalo entre 0 e 1. As amostras foram divididas em vários grupos de duas classes por meio de limiares, para cada divisão foram calculadas a Taxa de Verdadeiros Positivos e a Taxa de Falsos Positivos, formando a curva ROC (FAWCETT, 2006).

2.3 RESULTADOS E DISCUSSÕES

No inventário, foram identificadas 297 cicatrizes relativas ao evento de 05/01/2017, com área total de 1,06 km² (Figura 2.4). A maior cicatriz apresentou 4,0 ha, e a maior parte das cicatrizes apresentou largura inferior a 15 m e comprimento na ordem de 200 m. O conjunto amostral resultou em 39.318 amostras correspondentes a píxels do MDE, 50% representando áreas de ocorrência (1), e 50% locais de não ocorrência (0). Na modelagem, as amostras foram divididas em três séries, sendo a quantidade relativa ao método de reamostragem: treinamento (21% dos dados para $x=50$, e 14% para $x=80$), validação cruzada (44% e 46%, respectivamente) e teste (35% e 40%, respectivamente).

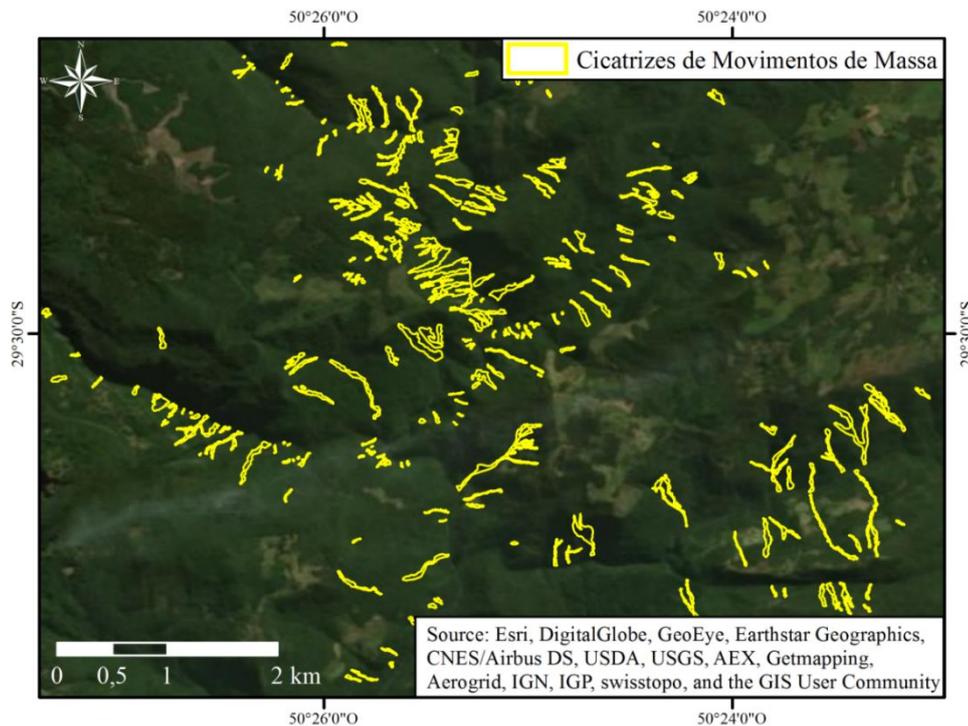


Figura 2.4. Inventário de Cicatrizes de Movimentos de Massa.

A Figura 2.5 apresenta gráficos do tipo *box-plot* com a distribuição dos valores de cada atributo do terreno, para os pontos de ocorrência e de não ocorrência de movimentos de massa. Os valores das curvaturas vertical e longitudinal e relevo sombreado se sobrepõem quanto à ocorrência e não ocorrência, e demonstra que esses parâmetros não são bons para identificar áreas suscetíveis. A elevação foi um atributo adequado para a área amostral 1, uma vez que os

dados não se sobrepõem e a identificação de áreas suscetíveis é facilitada. Contudo, os dados de elevação da área 2 se sobrepõem, dificultando, assim, a identificação de um limiar entre ocorrência e não ocorrência de movimentos de massa. Os demais atributos, declividade, fator LS e profundidade de vale, realçam essa diferença entre os dados.

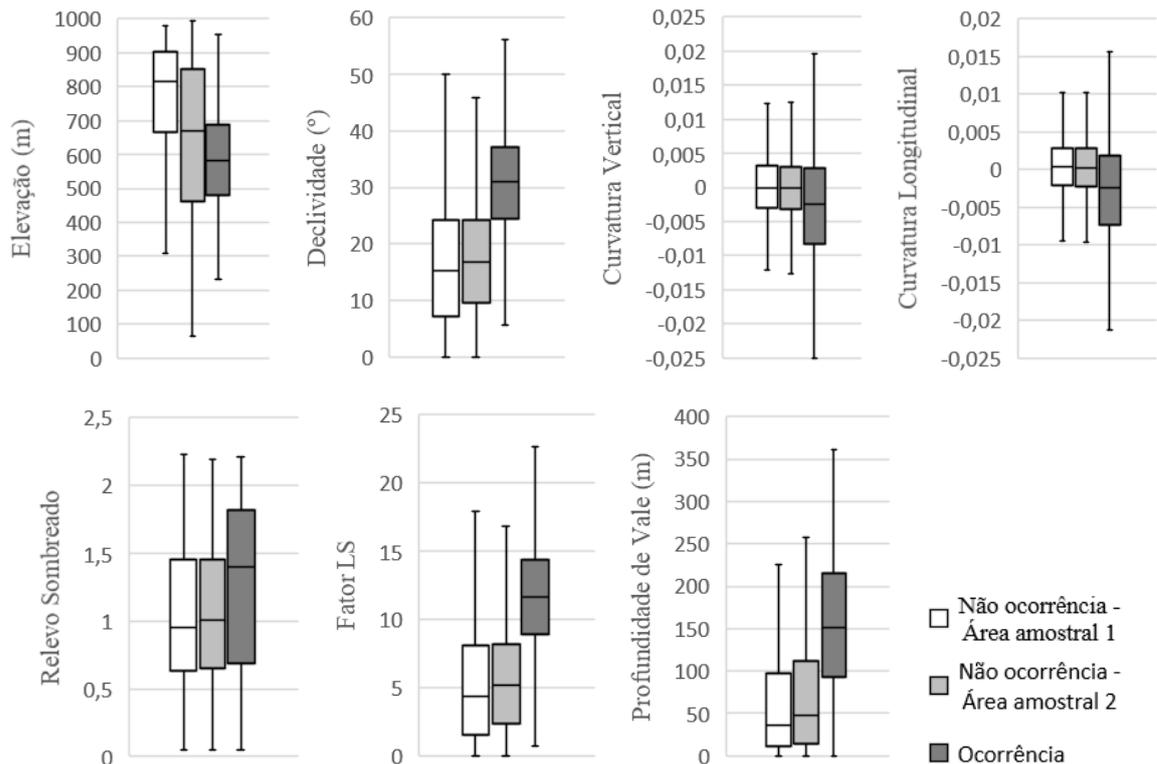


Figura 2.5. Gráficos da distribuição dos valores mínimos, médios e máximos dos atributos do terreno em áreas de ocorrência e de não ocorrência de movimentos de massa.

A declividade é um dos fatores mais utilizados em RNAs (PRADHAN; LEE, 2010), devido à sua relação direta com a ocorrência de movimentos de massa. Quanto maior a declividade, maior a suscetibilidade de movimentação do material na vertente (METEN; PRAKASHBHANDARY; YATABE, 2015). Entretanto, segundo Chen *et al.* (2016), quando a declividade atinge um ângulo máximo maior que 65°, a suscetibilidade diminui, uma vez que o aporte de material diminui.

O grau de importância das variáveis (Tabela 2.1) mostra que a elevação apresentou maior peso nos 4 modelos selecionados para análise (a melhor configuração de RNA para cada área amostral e intervalo de reamostragem, conforme índice AUC). Conforme Kawabata e Bandibas

(2009), a retirada deste parâmetro reduz a acurácia de 94% para 72%. O segundo atributo mais importante variou entre declividade, fator LS e profundidade de vale. As curvaturas vertical e longitudinal e o relevo sombreado apresentaram pouca importância, devido à sobreposição de valores quanto a ocorrência e não ocorrência (PRADHAN; LEE, 2010).

Tabela 2.1. Importância dos atributos de entrada em cada modelo de RNA selecionado.
Importância dos atributos de entrada em cada modelo de RNA selecionado.

	Área 1 $x=50$	Área 1 $x=80$	Área 2 $x=50$	Área 2 $x=80$
Elevação	37%	34%	32%	38%
Declividade	17%	24%	18%	19%
Curvatura Vertical	02%	02%	03%	02%
Curvatura Longitudinal	04%	06%	05%	06%
Relevo Sombreado	06%	02%	10%	08%
Fator LS	20%	21%	13%	05%
Profundidade de Vale	14%	11%	19%	21%

Os índices de acurácia (AUC) para as séries de treinamento, validação cruzada e teste das RNAs (Tabela 2.2), consideram a melhor configuração obtida para cada combinação de dados de entrada. Estudos consideram uma boa acurácia em torno de 80% (PRADHAN; LEE, 2010; DOU *et al.*, 2015), os valores obtidos neste estudo, independente do conjunto de atributos usados e da complexidade da RNA, as AUC foram próximas a 0,9, indicando um excelente grau de acerto no mapeamento da suscetibilidade a movimentos de massa.

A área 1 apresentou acurácia menor quando comparada à área 2, que é mais abrangente. Quanto ao método de reamostragem, não há um padrão, uma vez que para a área 1, o método que apresentou o melhor desempenho foi aquele que considerava $x=50$, enquanto que na área 2 foi a reamostragem mais restritiva ($x=80$) que obteve o melhor resultado. Isso permite destacar que a redução do conjunto amostral de treinamento não afeta a acurácia, desde que o processo de reamostragem resulte em amostras de treinamento representativas do conjunto inicial.

Tabela 2.2. Acurácia dos modelos para os diferentes conjuntos amostrais.

Área	Amostragem	Modelo	AUC			Nº. Neurônios
			Calibração	Teste	Geral	
1	50	5	0,898	0,890	0,890	11
	80	2	0,879	0,880	0,879	05
2	50	6	0,926	0,919	0,920	13
	80	5	0,937	0,920	0,920	11

As quatro melhores configurações de RNA foram espacializadas para análise visual dos mapas de suscetibilidade (Figura 2.6). Os mapas resultantes foram classificados em cinco graus de suscetibilidade a movimentos de massa, considerando o intervalo de 0 a 1. O melhor modelo referente à área 1, reamostragem 80 (Figura 2.6B), não representou de maneira adequada as áreas suscetíveis a movimentos de massa, apesar de apresentar uma AUC de 0,88. O mapa apresenta áreas suscetíveis em fundo de vale, incoerente com o tipo de fenômeno estudado, demonstrando, assim, uma baixa capacidade de generalização. É possível que isto tenha acontecido porque a área amostral considerada foi mais restrita e a amostra de treinamento do modelo foi menor, devido à reamostragem a cada 80 amostras.

Os mapas da área 2, mais abrangente, apresentaram AUC acima de 0,9 e elevada capacidade de generalização. Não mapearam fundos de vale e abrangeram de maneira adequada as cicatrizes. O modelo com melhor AUC foi o da reamostragem 80 (Figura 2.6D), porém, este se aproxima do modelo de reamostragem 50 tanto em AUC quanto aos elementos visuais do mapa de suscetibilidade. Nem sempre os modelos que apresentam melhor AUC são os modelos mais adequados. É necessária uma análise visual dos resultados para selecionar aqueles que resultam no melhor mapeamento de áreas suscetíveis a movimentos de massa.

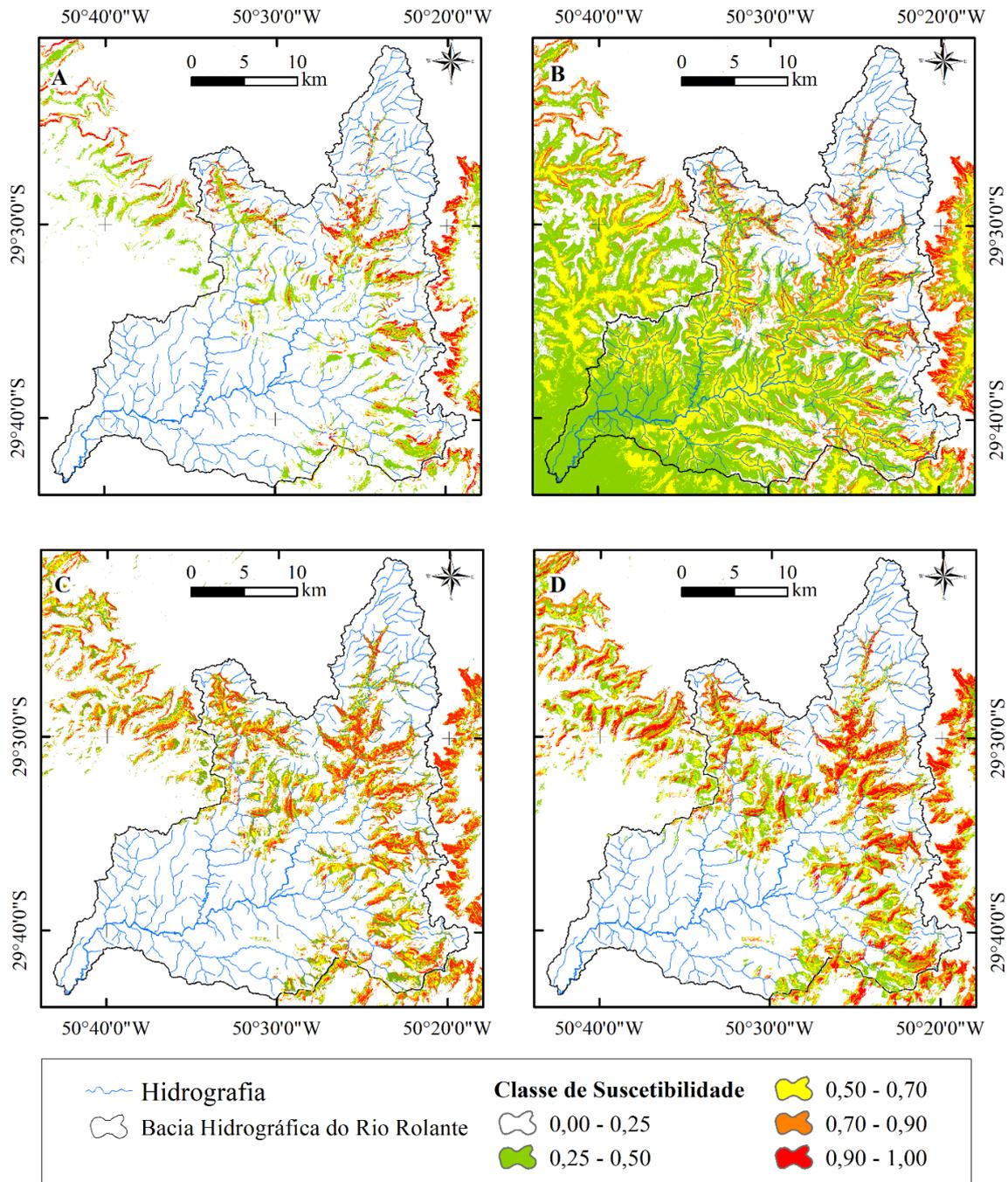


Figura 2.6. Mapas de suscetibilidade a movimentos de massa de acordo com o modelo de maior AUC de cada RNA: (A) Área amostral 1, reamostragem 50; (B) Área amostral 1, reamostragem 80; (C) Área amostral 2, reamostragem 50; (D) Área amostral 2, reamostragem 80.

Os resultados reforçam as análises de Dou *et al.* (2015) e Oliveira, Pedrollo e Castro (2015) sobre a importância da pré-seleção dos atributos de entrada nos modelos de RNAs, para eliminar variáveis redundantes e de baixa correlação, e identificar os principais fatores condicionantes.

2.4 CONCLUSÕES

Os movimentos de massa são um dos fenômenos que mais causam desastres no Brasil. O uso de RNAs para mapeamento de suscetibilidade a movimentos de massa têm se apresentado adequado devido à elevada acurácia dos modelos. Neste estudo, as RNAs se mostraram capazes de modelar as áreas suscetíveis, com acurácia no mapeamento próximas ou superiores a 0,9.

Foram utilizados sete atributos do terreno como dados de entrada para modelar a suscetibilidade a partir de RNAs. Verificou-se que parâmetros morfométricos como elevação, declividade, fator LS e profundidade de vale são variáveis com maior importância quando da modelagem de áreas suscetíveis a movimentos de massa.

As amostras foram obtidas a partir de duas áreas amostrais e dois métodos de reamostragem. Conclui-se que a restrição de área para coleta de amostras aleatórias de não ocorrência pode afetar a capacidade de generalização do modelo, e que a redução do conjunto amostral de treinamento não afeta significativamente na acurácia dos resultados.

2.5 AGRADECIMENTOS

O presente trabalho foi realizado com apoio: i) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001; ii) da Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul – FAPERGS, com apoio financeiro via Edital 01/2017 – ARD, processo 17/2551-0000894-4.

2.6 REFERÊNCIAS

- ADITIAN, A.; KUBOTA, T.; SHINOHARA, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. **Geomorphology**, v. 318, p. 101 – 111, 2018.
- CEPED UFSC. **Atlas Brasileiro de Desastres Naturais – 1991 a 2012**. Volume Brasil. 2ed. rev. amp. Florianópolis: CEPED UFSC, 2013. 127 p.
- CHEN, X.; LIU, C.; CHANG, Z.; ZHOU, Q. The relationship between the slope angle and the landslide size derived from limit equilibrium simulations. **Geomorphology**, v. 253, p. 547 – 550, jan. 2016.

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

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861 – 874. 2006

GARCIA-URQUIA, E.; YAMAGISHI, H. Landslide Susceptibility Mapping Based on Aerial Photograph Interpretation Inventory for Tegucigalpa, Honduras: An Application of the Matrix Method. *In*: YAMAGISHI, H.; BHANDARY, N. P. (Eds). **GIS Landslide**. Tokyo: Springer, 2017, p. 163 - 181.

JENSEN, J. R. **Introductory digital image processing: a remote sensing perspective**. 3 ed., New Jersey: Prentice-Hall, 2005. 526 p.

KAWABATA, D.; BANDIBAS, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). **Geomorphology**, v. 113, p. 97 – 109, 2009.

LE, Q. H.; NGUYEN, T. H. V.; DO, M. D.; LE, T. C. H.; PHAM, V. S.; NGUYEN, H. K.; LUU, T. B. TXT-tool 1.084-3.1: Landslide Susceptibility Mapping at a Regional Scale in Vietnam. *In*: SASSA, K. *et al.* (Eds). **Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools**. Volume 1: Fundamentals, Mapping and Monitoring. Cham: Springer, 2017. p. 161 - 174.

METEN, M.; PRAKASHBHANDARY, N.; YATABE, R. Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia. **Geoenvironmental Disasters**, v. 2: 9, p. 1 – 17, dec. 2015.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection. **Engineering Applications of Artificial Intelligence**, v. 40, p. 47 – 61, 2015.

PRADHAN, B.; LEE, S. Regional landslide susceptibility analysis using backpropagation neural network model at Cameron Highland, Malaysia. **Landslides**, v. 7, p. 13 – 30, 2010.

RIFFEL, E. S.; RUIZ, L. F. C.; GUASSELLI, L. A. Mapeamento de suscetibilidade a deslizamentos a partir de Mineração de Dados e do modelo SHALSTAB. **Revista Brasileira de Cartografia**, Edição Especial Movimentos de Massa e Processos Erosivos, nº. 68/9, p. 1805 – 1818, 2016.

ROSSATO, M. S. **Os Climas do Rio Grande do Sul: variabilidade, tendências e tipologia**. Porto Alegre: UFRGS, 2011. 253 p. Tese (Doutorado em Geografia) – Programa de Pós-Graduação em Geografia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, p. 533- 536, 1986.

SEMA. **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, 26 p. 2017.

WIDROW, B.; HOFF, M. E. Adaptive Switching Circuits. *In:* 1960 IRE WESCON CONVENTION RECORD, New York: IRE Part, 1960. p. 96– 104.

ZÊZERE, J. L. *et al.* Mapping landslide susceptibility using data-driven methods. **Science of the Total Environment**, v. 589, p. 250-267, july 2017.

CAPÍTULO 3 : MODELAGEM DE ÁREAS SUSCETÍVEIS A MOVIMENTOS DE MASSA: AVALIAÇÃO COMPARATIVA DE TÉCNICAS DE AMOSTRAGEM, APRENDIZADO DE MÁQUINA E MODELOS DIGITAIS DE ELEVAÇÃO.

RESUMO

Métodos de redes neurais artificiais (RNA) e *random forest* (RF) apresentam bom desempenho para mapear áreas suscetíveis a movimentos de massa. Entretanto, a modelagem a movimentos de massa por meio dessas técnicas de aprendizado de máquina é sensível à amostragem, à escala do modelo digital de elevação (MDE), ao conjunto de atributos do terreno e aos seus parâmetros de ajuste, influenciando o mapa final e o poder de generalização dos modelos. O objetivo deste artigo foi avaliar e comparar técnicas de amostragem e MDE para a modelagem de áreas suscetíveis a movimentos de massa, utilizando RNA e RF. O estudo foi aplicado à Bacia Hidrográfica do Rio Rolante. Foram extraídos sete atributos do terreno, a partir dos MDEs ALOS-PALSAR (AP) e ASTER GDEM (AG). As amostras foram selecionadas considerando-se duas áreas amostrais e comparando dois métodos de reamostragem para redução do conjunto de treinamento dos modelos. Todas as combinações de modelos resultaram em acurácias entre 0,88 e 0,94, demonstrando que os modelos RNA e RF, combinados com os MDEs AP e AG, possibilitam identificar áreas suscetíveis a movimentos de massa. Na modelagem se destacaram os seguintes atributos: elevação, declividade, fator LS e profundidade do vale. A definição de uma área amostral mais abrangente para coleta de amostras de não ocorrência aumenta a acurácia e a capacidade de generalização dos modelos. A redução do conjunto amostral de treinamento diminuiu o tempo de processamento, sem interferir significativamente na acurácia do mapa.

Palavras-chave: redes neurais artificiais, *random forest*, ALOS-PALSAR, ASTER GDEM.

ABSTRACT

Artificial neural network (ANN) and random forest (RF) methods present good performance to map landslides susceptible areas. However, the landslides modeling through these machine learning techniques is sensitive to the sampling, the scale of the digital elevation models (DEM),

to the set of terrain attributes and their adjustment parameters, influencing on the final map and the models generalization power. This article aimed to evaluate and compare sampling techniques and DEM for modeling landslides susceptible areas, using ANN and RF. The study area corresponds to the Rolante River Basin. Seven attributes of the terrain were extracted from DEM ALOS-PALSAR (AP) and DEM ASTER GDEM (AG). These samples were selected considering two areas and comparing two resampling methods to reduce the training set of the models. All combinations resulted in values for accuracy between 0.88 and 0.94. This demonstrated that ANN and RF models, combined with AP and AG DEMs, serve the purpose of identifying landslides susceptible areas. In general, the terrain attributes that stood out in the modeling were: elevation, slope, LS factor, and valley depth. Concluding that the most comprehensive sampling areas for non-occurrence samples increased the accuracy and the generalization capacity. In addition, the training sample reduction decreased the processing time, without significantly increasing the maps accuracy.

Keywords: artificial neural network, random forest, ALOS-PALSAR, ASTER GDEM.

3.1 INTRODUÇÃO

Os movimentos de massa podem ser desencadeados por terremotos, precipitações extremas ou atividades antrópicas e estão entre os desastres que causam maior número de vítimas fatais. No período de 2004 e 2016, o número de mortes devido à ocorrência de movimentos de massa ao redor do mundo foi de 4.862 (FROUDE; PETLEY, 2018). No Brasil, os eventos de Itajaí, em Santa Catarina, em 2008, e da região serrana do Rio de Janeiro, em 2011, foram responsáveis por um elevado número de vítimas, 135 e 917 mortes, respectivamente.

Devido a este cenário, foi elaborado o Plano Nacional de Gestão de Riscos e Resposta a Desastres Naturais no ano de 2012, com o objetivo de mitigar os riscos das áreas suscetíveis a desastres ocasionados por fenômenos naturais (BRASIL, 2012; CEPED UFSC, 2013).

O mapeamento de suscetibilidade é uma importante ferramenta de auxílio aos gestores locais para o planejamento urbano, de forma a evitar a ocupação de áreas suscetíveis a movimentos de massa ou inundações, além de possibilitar a reorganização do espaço urbano e identificar áreas prioritárias para instalação de sistemas de alerta (ADITIAN; KUBOTA; SHINOHARA, 2018; OLIVEIRA; PEDROLLO; CASTRO, 2018). O aprendizado de

máquinas, como redes neurais artificiais (RNA) e *Random Forest* (RF), se apresenta como uma ferramenta importante para esse tipo de mapeamento (KAWABATA; BANDIBAS, 2009; PRADHAN; LEE, 2010; DOU *et al.*, 2015; CHEN *et al.*, 2017; ADA; SAN, 2018; ADITIAN; KUBOTA; SHINOHARA, 2018).

Modelagens por RNA e/ou por RF para mapeamento de áreas suscetíveis a movimentos e massa foram realizados por Kawabata e Bandibas (2009), a partir do uso de dois MDEs, que demonstraram resultados com acurácia superior a 0,9, sendo 0,937 para o MDE de 15 m e 0,944 para o MDE de 55 m. Pradhan e Lee (2010) obtiveram acurácia de 0,83 e o atributo do terreno que mais contribuiu na modelagem foi a declividade. Dou *et al.* (2015) selecionaram seis fatores condicionantes dentre 14 disponíveis e perceberam que a acurácia com seis atributos foi de 0,82, enquanto que com 14 parâmetros foi de 0,73. Chen *et al.* (2017) compararam os métodos *maximum entropy*, *support vector machine* (SVM) e RNA, em que a RNA apresentou maior acurácia (0,824) e a elevação foi um dos fatores mais importantes. Aditian, Kubota e Shinohara (2018) compararam três métodos diferentes para modelagem e mapeamento de suscetibilidade na Indonésia: *bivariate frequency ratio*, *multivariate logistic regression* e RNA. Dentre eles, o modelo que apresentou maior taxa de sucesso foi a RNA, com acurácia de 0,734. Ada e San (2018) compararam as técnicas SVM e RF, e obtiveram acurácias entre 0,82 e 0,87, respectivamente.

Os métodos RF e RNA apresentam um desempenho satisfatório para mapear áreas suscetíveis a movimentos de massa. Entretanto, por serem métodos empíricos, a modelagem de áreas de movimentos de massa por meio das técnicas de aprendizado de máquina é sensível à amostragem, à escala do MDE, ao conjunto de atributos do terreno e aos seus parâmetros de ajuste, influenciando o mapa final e o poder de generalização dos modelos. Logo, é importante avaliar o desempenho desses modelos em função de diferentes conjuntos amostrais e fontes de dados de entrada, de modo a analisar sua capacidade de generalização para áreas maiores, bem como a sensibilidade dos modelos em relação aos atributos de entrada.

Assim, o objetivo deste artigo foi avaliar o desempenho e a sensibilidade da modelagem de áreas suscetíveis a movimento de massa considerando: (i) os métodos de aprendizagem de máquina RNA e RF; (ii) duas áreas para coleta de amostras de não ocorrência, combinadas com dois métodos de reamostragem dos conjuntos de treinamento; e (iii) usando os MDEs ALOS-PALSAR e o ASTER GDEM para obtenção dos atributos do terreno.

3.2 MATERIAIS E MÉTODOS

3.2.1 Área de Estudo

Este estudo foi aplicado à área da Bacia Hidrográfica do Rio Rolante (BHRR), Região Hidrográfica do Guaíba. Encontra-se na porção nordeste do estado do Rio Grande do Sul, entre as latitudes $29^{\circ}20'43''\text{S}$ e $29^{\circ}43'33''\text{S}$ e as longitudes $50^{\circ}18'48''\text{O}$ e $50^{\circ}43'06''\text{O}$ (**Erro! Fonte de referência não encontrada.**). A área de drenagem da BHRR é de aproximadamente 830 km², com altitudes entre 152 e 992 m. De acordo com Rossato (2011), a região apresenta um clima subtropical muito úmido, com precipitação entre 1700 e 2000 mm anuais, com ingresso de sistemas frontais de precipitação e influência da orografia.

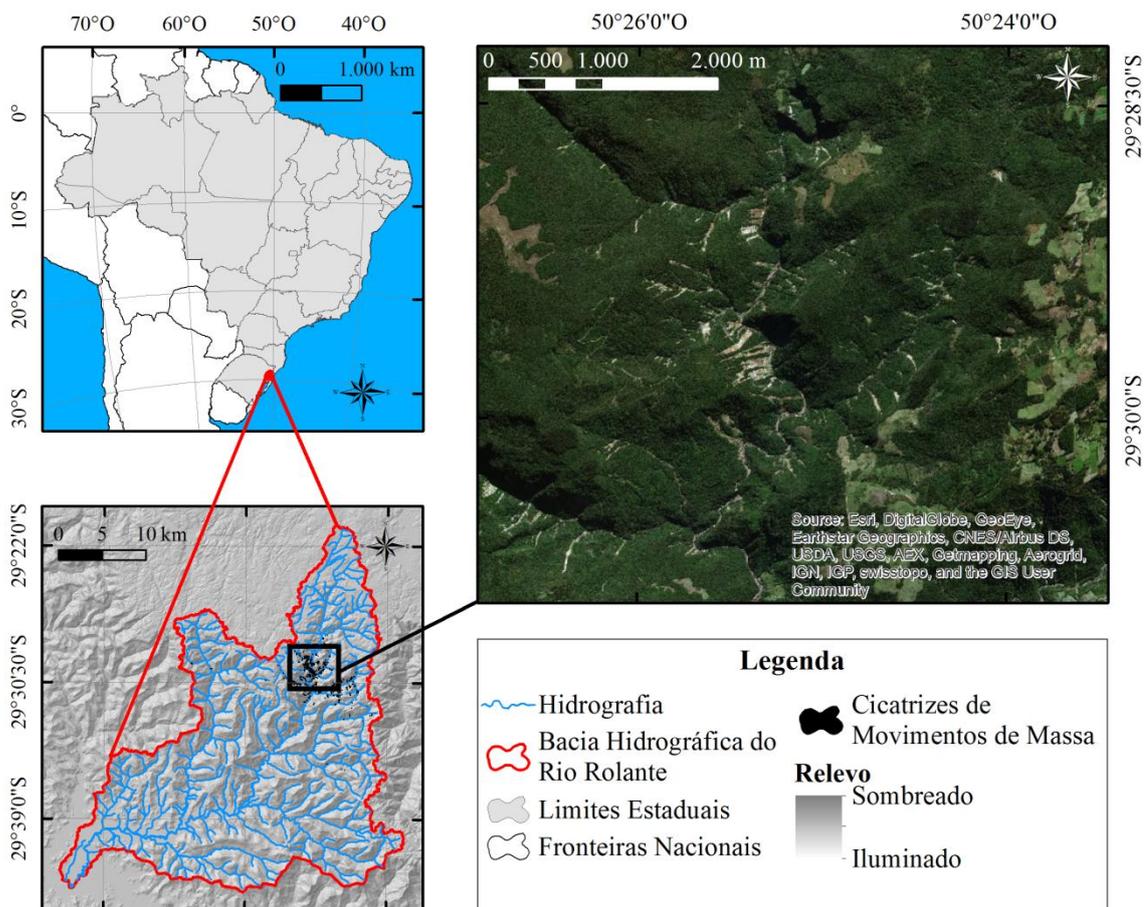


Figura 3.1. Localização das cicatrizes de movimentos de massa, Bacia Hidrográfica do Rio Rolante, RS.

Na BHRR, ocorreu um evento extremo de precipitação, em 05 de janeiro de 2017, que desencadeou mais de duzentas cicatrizes de movimentos de massa (Figura 3.2). As chuvas apresentaram significativa concentração espaço-temporal, na região das cabeceiras da BHRR, na sub-bacia do rio Mascarada. Devido à falta de pluviômetros oficiais na região, foram consultados medidores particulares, os quais apresentaram valores entre 90 e 272 mm, diferentemente da estimativa oficial que foi de 50 mm (SEMA, 2017).



Figura 3.2 Cicatrizes de movimentos de massa resultantes do evento extremo de precipitação de 05 de janeiro de 2017. Fonte: 2º Pelotão de Polícia Ambiental de Canela (PATRAM) de Canela, RS, 13 de janeiro de 2017.

As chuvas tiveram duração de quatro horas, causando o deslocamento do material da encosta para o fundo do vale. Consequentemente, formaram-se barreiras no leito dos rios, houve aumento do nível das águas e a carga de energia retidos. Quando houve o rompimento dessas barreiras, uma consequente enxurrada atingiu o município de Rolante, causando diversas perdas e danos materiais (SEMA, 2017).

3.2.2 Métodos

Os métodos foram estruturados em cinco etapas de processamento, conforme (Figura 3.3): i) extração de atributos do terreno dos MDEs AP e AG; ii) pré-processamento e organização dos conjuntos amostrais; iii) treinamento e calibração dos modelos; iv) teste dos modelos; v) reclassificação dos modelos com melhor acurácia, para níveis de suscetibilidade em ambiente de Sistema de Informações Geográficas.

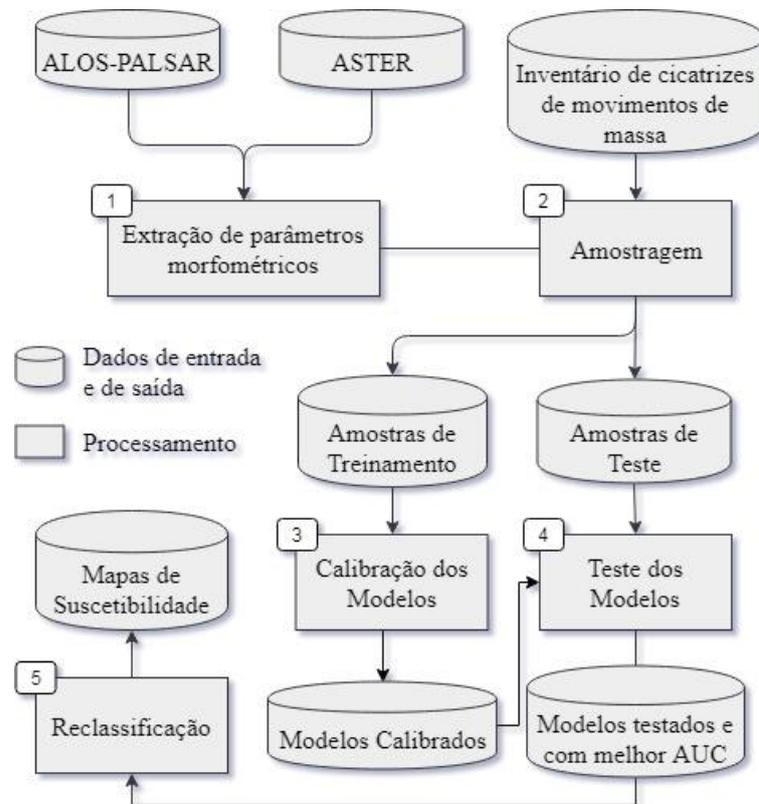


Figura 3.3. Fluxograma da metodologia utilizando dois modelos digitais de elevação.

3.2.3 Atributos Morfométricos

Foram utilizados os MDEs AP e AG para a extração dos atributos do terreno, com resolução espacial de 12,5 m e 30 m, respectivamente. Foram selecionados sete com maior correlação com ocorrência de movimentos de massa: elevação (ELV), declividade (DCV), curvaturas vertical (CVV) e longitudinal (CVL), relevo sombreado (RLS), fator LS (FLS) e profundidade de vale (PRV).

A ELV se refere aos valores de altitude representada em metros. A DCV representa a inclinação da vertente em graus, com influência direta na espessura do solo, ou seja, do material passível de movimentação, e é considerado um dos fatores topográficos mais importantes quanto aos movimentos de vertente (METEN; PRAKASHBHANDARY; YATABE, 2015). Quanto maior a declividade, maior é a suscetibilidade, apresentando um máximo entre 35 e 40°. Segundo Dai e Lee, 2002, a suscetibilidade diminui em declives maiores do que 40°, no entanto para Chen *et al.* (2016) a diminuição ocorre a partir de um ângulo máximo maior que 65°.

As CVV e CVL descrevem o formato da encosta, influenciando diretamente nas tensões de condução e resistência, e no grau de convergência ou divergência do escoamento da água e do material transportado. A curvatura vertical apresenta-se ao longo de uma linha formada pela intersecção de um plano vertical imaginário e a superfície do solo, valores negativos representam formas côncavas e valores positivos, formas convexas. A curvatura longitudinal está relacionada à intersecção de um plano horizontal imaginário, sendo os valores negativos referentes a áreas convergentes e valores positivos, formas divergentes de escoamento (METEN; PRAKASHBHANDARY; YATABE 2015).

O mapa de RLS demonstra o padrão geomorfológico, os lineamentos estruturais, a localização e o formato dos vales e encostas. O FLS (*Slope Length and Steepness Factor*) apresenta o efeito da topografia na erosão do solo. Combina os fatores S, que mede o efeito da inclinação da encosta, e L, que descreve o impacto do comprimento da vertente (PANAGOS; BORRELLI; MEUSBURGUER, 2015).

O atributo PRV é baseado na distância vertical entre os cumes das áreas mais elevadas e as linhas de contorno de profundidade. Segundo Meinhardt, Fink e Tünschel (2015), a profundidade do vale é diretamente proporcional à suscetibilidade, sendo considerada alta para vales com profundidades superiores a 100 m.

3.2.4 Inventário de Cicatrizes de Movimentos de Massa

Estes são os dados de entrada que auxiliam a entender o padrão dos atributos do terreno quando da ocorrência dos movimentos de massa. A construção do inventário de cicatrizes do evento ocorrido em 05 de janeiro de 2017 foi realizado por meio de identificação visual a partir de imagens disponíveis no programa *Google Earth*, as quais apresentavam resolução espacial

mais adequada para este tipo de mapeamento, dentre aquelas disponibilizadas de maneira gratuita. As imagens utilizadas correspondem aos dias 06 e 24 de agosto de 2017, e são as mais próximas temporalmente do evento ocorrido.

Foram mapeadas 297 cicatrizes, distribuídas entre os municípios de São Francisco de Paula e Riozinho, as quais abrangem aproximadamente 1 km². A maioria das cicatrizes apresentou largura menor do que 15 m e comprimento próximo a 200 m, sendo a maior representada por um complexo de cicatrizes interligadas, com área de 4 ha. Os vetores correspondentes às cicatrizes foram transformados para o formato *raster* e, assim, extraído um ponto por píxel. Os pontos obtidos foram utilizados como amostras de áreas com ocorrência de movimentos de massa.

3.2.5 Amostragem

O conjunto amostral foi formado por pontos com e sem ocorrência de movimentos de massa. Os pontos relativos à presença de cicatrizes foram extraídos a partir do inventário, resultando em 19.659 píxeis, para o MDE AP, e 3.407 píxeis, para o MDE AG.

Para a seleção das amostras de não ocorrência, foram considerados duas áreas distintas, as quais abarcam o inventário (**Erro! Fonte de referência não encontrada.**4). Uma das áreas é mais restritiva, abrange 125 km², enquanto a outra possui uma área de 386 km². Foram desconsideradas as áreas do inventário e obtidos pontos aleatórios de modo a representar as áreas sem movimentos de massa. A quantidade de pontos correspondentes às áreas de não ocorrência é igual ao número de pontos com cicatrizes, de acordo com cada MDE, visando aumentar o comprometimento dos modelos em buscar limiares para a separação das classes de suscetibilidade.

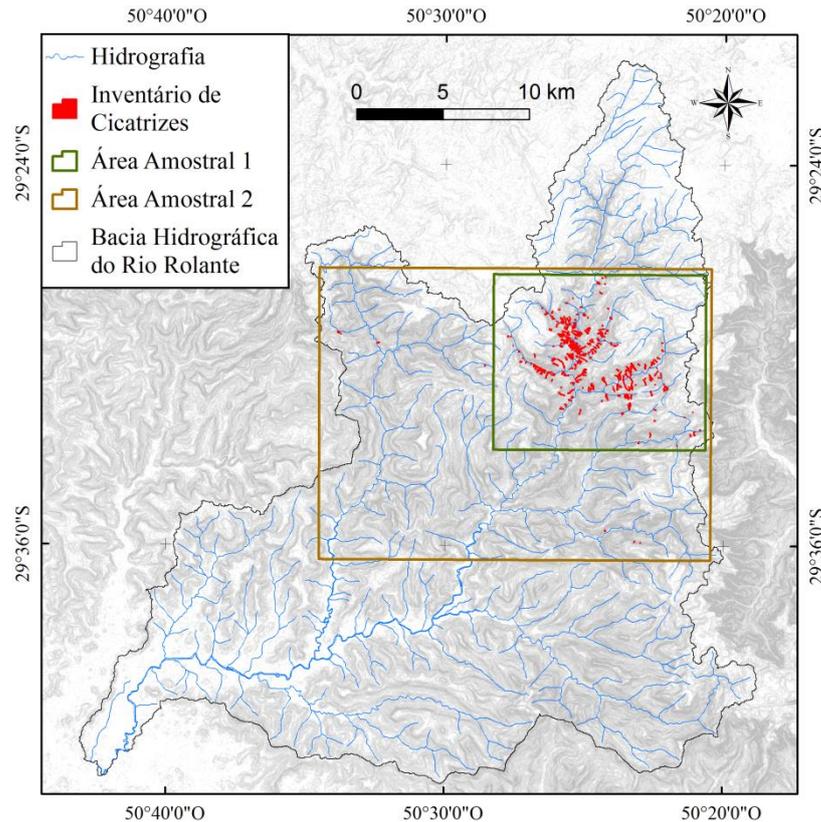


Figura 3.4. Área amostral restritiva (área 1) e abrangente (área 2) para coleta de amostras de não ocorrência de movimentos de massa.

Obtidos os dois conjuntos amostrais, e considerando a grande quantidade de píxeis do MDE AP, foram aplicados dois métodos de redução de amostras, a fim obter conjuntos menores para a etapa de treinamento dos modelos (diminuindo o custo computacional). As reamostragens foram realizadas a partir do reordenamento dos valores de cada atributo e seleção de um valor a cada x itens. Foram utilizados dois valores para x , 50 e 80.

Os conjuntos amostrais foram subdivididos em amostras de treinamento (21%, para reamostragem $x=50$ e 14%, para $x=80$), de validação cruzada (44% e 46%) e de teste (35% e 40%). A validação cruzada foi realizada apenas para o modelo RNA, a fim de evitar o superajustamento do modelo, enquanto que para RF foram utilizadas somente as amostras correspondentes a treinamento e teste.

Foram obtidos oito conjuntos amostrais, dos quais quatro conjuntos de dados foram extraídos do MDE AP e quatro do MDE AG. As amostras serviram como dados de entrada para a modelagem realizada a partir de RNA e RF.

3.2.6 Modelagem e Análise Espacial das Áreas Suscetíveis

Foi realizada a modelagem utilizando o algoritmo RF, desenvolvido com a linguagem de programação *Python*, e a RNA com o algoritmo retropropagativo, desenvolvido no programa *MATLAB*.

O RF é um método *ensemble-learning*, que seleciona de maneira aleatória percentuais dos dados de entrada e os separa em subconjuntos. Para cada sub-amostra é construída uma árvore de decisão e o valor estimado é referente à média da predição das árvores. RF gera um modelo de regressão e estima a relevância dos atributos do terreno (BREIMAN, 2001). O número de árvores e a suas profundidades foram avaliados com valores de 5 até 50, variando 5, para ambos os parâmetros.

Quanto às RNA, a etapa de treinamento foi realizada pelo método retropropagativo com múltiplas camadas (RUMELHART; HINTON; WILLIAMS, 1986), usando a Regra Delta (WIDROW; HOFF, 1960) para atualização dos pesos sinápticos da rede e a função de ativação sigmoideal. Após o treinamento, foi realizada uma validação cruzada, a fim de evitar o superajustamento do modelo. Foram testadas diversas configurações de modelo, com variação do conjunto de atributos de entrada e do número de neurônios na camada oculta. As variáveis de entrada foram excluídas de maneira sistemática, a partir do *Relative Contribution Index* (OLIVEIRA; PEDROLLO; CASTRO, 2015). Foram consideradas 15 iterações e 15.000 ciclos de aprendizagem, caso o modelo não atinja o limiar de convergência, para cada configuração do modelo.

A análise do desempenho dos modelos foi realizada a partir da acurácia (*Area Under the ROC Curve* - AUC) e foram selecionados os melhores resultados. Os dados de saída foram trabalhados no programa *ArcGIS*, de modo a preparar os dezesseis mapas finais. Foi realizada a classificação das imagens em intervalos de suscetibilidade.

Logo foram realizados testes de proporção com nível de significância de 5%. Este teste está baseado na existência de duas populações que representam determinada característica e uma hipótese que afirma que as proporções são iguais, ou seja, $H_0 : p_1 = p_2$. Assim, foi realizado o teste de proporção considerando-se a quantidade de amostras (n) e a AUC (p), conforme a Eq. 4 (MORETTIN; BUSSAB, 2010) :

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_1(1-p_1)}{n_2}}} \quad (4)$$

Em que: p representa o valor da AUC; n representa a quantidade de amostras; Z representa o valor da região crítica. Como este teste foi bilateral, se Z apresenta valores entre -1,96 e 1,96, rejeitamos a H_0 , caso contrário, não a rejeitamos.

Assim, os testes de proporção foram realizados comparando-se os pares: RF e RNA, MDE AP e MDE AG, áreas amostrais entre si e processos de reamostragem. A cada comparação realizada, foram obtidos oito cenários, os quais demonstraram similaridade ou não entre as proporções.

3.3 RESULTADOS E DISCUSSÕES

Os sete atributos do terreno selecionados foram extraídos dos MDEs e utilizados como dados de entrada para os modelos de RNA e RF. Foram calculados os percentuais de importância das variáveis de acordo com cada método (RNA e RF), cada MDE (AP e AG) e cada área amostral, considerando-se o método de reamostragem utilizado ($x=50$ e $x=80$), conforme Tabelas 3.1 e 3.2.

Nos modelos de RNA, o atributo que apresentou maior importância, quanto ao MDE AP, foi a ELV (Tabela 3.1). Em diversos estudos, a ELV é apresentada como um dos fatores mais eficazes para modelar a ocorrência de movimentos de massa (KAWABATA; BANDIBAS, 2009; CHEN *et al.*, 2017). Chen *et al.* (2017) utilizaram as RNA, entre outros métodos, para modelar as áreas suscetíveis, considerando 11 parâmetros, com destaque para a ELV. Kawabata e Bandibas (2009) perceberam que a remoção da ELV causa uma redução em mais de 20% no valor da AUC. O segundo fator com maior influência foi alternado entre FLS (área 1 com reamostragem $x=50$), DCV (área 1, $x=80$) e PRV (área 2, $x=50$ e $x=80$).

Tabela 3.1. Importância dos atributos de terreno nos modelos de RNA.

	ALOS-PALSAR				ASTER GDEM			
	Área 1		Área 2		Área 1		Área 2	
	$x=50$	$x=80$	$x=50$	$x=80$	$x=50$	$x=80$	$x=50$	$x=80$
Elevação	37%	34%	32%	38%	35%	38%	41%	43%
Declividade	17%	24%	18%	19%	19%	15%	10%	10%
Curvatura Longitudinal	04%	06%	05%	06%	00%	01%	01%	00%
Curvatura Vertical	02%	02%	03%	02%	01%	00%	01%	00%
Relevo Sombreado	06%	02%	10%	08%	04%	04%	04%	01%
Fator LS	20%	21%	13%	05%	01%	02%	02%	02%
Profundidade de Vale	14%	11%	19%	21%	40%	40%	40%	44%

Tabela 3.2. Importância dos atributos do terreno no modelo RF.

	ALOS-PALSAR				ASTER GDEM			
	Área 1		Área 2		Área 1		Área 2	
	$x=50$	$x=80$	$x=50$	$x=80$	$x=50$	$x=80$	$x=50$	$x=80$
Elevação	17%	17%	17%	16%	09%	05%	21%	21%
Declividade	00%	06%	06%	06%	12%	09%	13%	12%
Curvatura Longitudinal	09%	08%	07%	06%	09%	04%	04%	09%
Curvatura Vertical	09%	09%	07%	08%	08%	05%	07%	08%
Relevo Sombreado	09%	08%	07%	08%	08%	03%	02%	00%
Fator LS	43%	42%	39%	40%	12%	11%	34%	27%
Profundidade de Vale	13%	10%	17%	16%	42%	61%	19%	23%

Na modelagem com o MDE AG, a PRV foi o atributo mais relevante e a ELV apresentou segunda maior importância, havendo alternância quanto à área 2 com reamostragem $x=50$. Vales com profundidades entre 100 e 200 metros apresentaram maior número de ocorrência de movimentos de massa, o mesmo foi demonstrado em Meinhardt, Fink e Tünschel (2015). O somatório das importâncias das variáveis PRV, ELV e DCV apresentou valores entre 91% e

97%, o que demonstra que esses três atributos foram os que mais influenciaram na modelagem por RNA com MDE AG.

As curvaturas foram os atributos menos importantes, as quais somadas representaram menos de 2% de importância na modelagem com o MDE AG, e entre 6% e 8% com MDE AP. As curvaturas tendem a ser os parâmetros com menor peso nas modelagens, como verificado em Chen *et al.* (2017), que utilizaram o índice Kappa para avaliar a importância dos atributos, e obtiveram valores menores do que 0,09 para ambas as curvaturas. Pourghasemi & Rahmati (2018) compararam 10 métodos de aprendizado de máquinas utilizando 11 parâmetros, e ambas as curvaturas demonstraram menor importância. Meinhardt, Fink e Tünschel (2015) perceberam que a exclusão das curvaturas gerava um acréscimo na acurácia dos seus modelos.

A modelagem utilizando o método RF apresentou o FLS como o atributo mais importante quanto ao MDE AP, sendo o único a apresentar valores acima de 18% (entre 39% e 43%) nos quatro modelos considerados (Tabela 3.2). Juntamente com os fatores ELV e PRV, concentraram entre 69% e 73% da importância. Na modelagem com MDE AG, os fatores mais influentes foram PRV e FLS, os quais apresentaram juntos 54% e 72% da importância quanto à área 1, $x=50$ e $x=80$, respectivamente. A área 2 destacou, além dos fatores PRV e FLS, o atributo ELV, que somados apresentam valores de importância entre 71% e 74%.

A DCV é considerada como um dos fatores mais fortemente relacionados com a ocorrência de movimentos de massa (PRADHAN; LEE, 2010; OLIVEIRA *et al.*, 2018; POURGHASEMI *et al.*, 2018). Entretanto, a modelagem do MDE AP por RF apresentou a DCV como atributo menos influente no modelo. Para Kawabata e Bandibas (2009), a DCV apresentou pouca relevância quando comparado com os demais atributos utilizados, pois sua exclusão da modelagem modificou levemente a AUC (de 93,68% para 92,71%, ou seja, em menos de 1%).

Ao se considerar dois fatores com informações semelhantes, tende-se a considerar um com maior peso, ao passo que o outro recebe um peso menor. Assim, nesta modelagem, o atributo com maior importância foi o FLS, enquanto que a DCV foi considerada com menor importância. O FLS demonstra o efeito da topografia na erosão do solo, combinando o efeito da inclinação da encosta (DCV) e o comprimento da vertente. Ou seja, há informação comum entre esses atributos, sendo essa a provável causa do resultado obtido.

De modo geral, a CVV, a CVL e o RLS foram os atributos menos importantes após a DCV. Os percentuais variaram de 0% (variável excluída do modelo) a 9%. Os valores de DCV, com modelagem a partir de RF, não superaram os 13%. Com isso, notou-se que as variáveis que apresentaram grau de importância significativo (acima de 20%) em ao menos dois modelos foram ELV, FLS e PRV com o MDE AG.

As acurácias (AUC) obtidas nos modelos de RNA e RF consideram a melhor configuração para cada combinação de dados de entrada (considerando os dois MDEs, as duas áreas amostrais e as duas técnicas de reamostragem, totalizando 16 mapas de suscetibilidade) apresentados na Tabela 3.3. Estudos consideram uma boa AUC valores superiores a 0,8 (PRADHAN; LEE, 2010; Dou *et al.*, 2015), neste estudo, independentemente do conjunto de amostras utilizados e da complexidade das RNA ou do modelo RF, a acurácia AUC apresentou valores entre 0,88 e 0,94, indicando um elevado grau de acerto no mapeamento de suscetibilidade a movimentos de massa.

Tabela 3.3. Acurácia (AUC) dos modelos RNA e RF.

	Área	Amostragem	RNA		RF	
			AUC	Nº. Neurônios	AUC	Nº. Árvores / Profundidade
ALOS- PALSAR	1	50	0,8902	11	0,9124	45 / 35
		80	0,8797	05	0,9001	45 / 50
	2	50	0,9194	13	0,9356	50 / 25
		80	0,9199	11	0,9296	50 / 50
ASTER GDEM	1	50	0,8888	09	0,8807	40 / 30
		80	0,8857	05	0,8762	45 / 05
	2	50	0,9319	05	0,9224	45 / 50
		80	0,9229	13	0,9156	35 / 10

De maneira geral, o desempenho do modelo RF foi melhor do que as RNA, apresentando somente dois valores abaixo de 0,9, semelhante ao observado em Pourghasemi e Rahmati (2018). Ao analisar de maneira individual cada resultado, verifica-se que o modelo RF

apresentou AUC maior do que as RNA com o uso do MDE AP, em contrapartida, com o MDE AG, quase todos os conjuntos de dados das RNA apresentam AUC levemente maior, à exceção da área 2, com reamostragem por $x=80$. A realização do teste de proporção, com nível de significância de 5%, demonstrou que, em três dos oito cenários, a hipótese de que os índices de acurácia dos dois modelos (RF e RNA) são semelhantes deve ser rejeitada, indicando que o modelo RF apresentou desempenho superior à RNA no nível de significância do teste.

Entre os MDE, o AP resultou em valores de AUC maiores quando comparado ao uso do MDE AG em todos os resultados referentes ao RF. Quanto à modelagem por RNA, a área 1, reamostragem por $x=50$, e a área 2, $x=80$, apresentaram valores muito próximos entre os modelos, 0,89 e 0,92, respectivamente. Quanto às outras duas configurações, o MDE AG apresentou AUC levemente maior, semelhante ao apresentado por Arnone *et al.* (2016) em que foram testados MDE com resoluções de 10 m e de 30 m e percebeu-se uma diferença de 0,08 entre os índices AUC, 0,87 para o MDE com melhor resolução e 0,79 para o MDE de 30 metros.

Ao verificar-se que a acurácia dos mapas resultantes da modelagem usando os MDEs AP e AG não possui uma diferença maior do que 0,03, percebe-se que o MDE AG apresenta um bom desempenho com o uso de um conjunto amostral menos numeroso do que o MDE AP. O teste de proporção realizado demonstrou que não se pode rejeitar a hipótese de superioridade do MDE AP somente em duas das oito comparações, enquanto que nas demais essa hipótese deve ser rejeitada, indicando que as acurácias podem ser equivalentes para o uso dos dois MDEs. Cabe ressaltar, no entanto, que a resolução espacial do mapa final é inferior no caso do MDE AG, não se adequando para escalas superiores a 1:50.000.

Houve distinção quanto às áreas amostrais. A área 2 apresentou maiores AUC e a área 1 representou cerca de um terço da área amostral 2, o que resulta em uma maior restrição quanto à representatividade dos atributos do terreno, uma vez que abrange menos áreas diferentes, como fundos de vales, por exemplo. A maior abrangência da área 2 fez com que houvessem pontos amostrais referentes a áreas sem cicatrizes com maior variação dos valores e representatividade. Em todos os cenários realizados no teste de proporção, se aceita a hipótese de que a área 2 demonstra melhor desempenho do que a área 1.

A área 2 tende a superestimar áreas de alta suscetibilidade a movimentos de massa (0,90 – 1,00) quando comparada à área 1 (3.5), principalmente para o RF (Figuras 4B.3a, 4B.3b, 4B.4a, 4B.4b). A área 1 apresentou as classes de maneira contínua, aumentando as áreas de

suscetibilidade em fundos de vale. A área 2, quanto ao MDE AP apresentou os resultados mais coerentes para as áreas de fundo de vale, tanto para RNA, quanto para RF, com valores baixos de suscetibilidade.

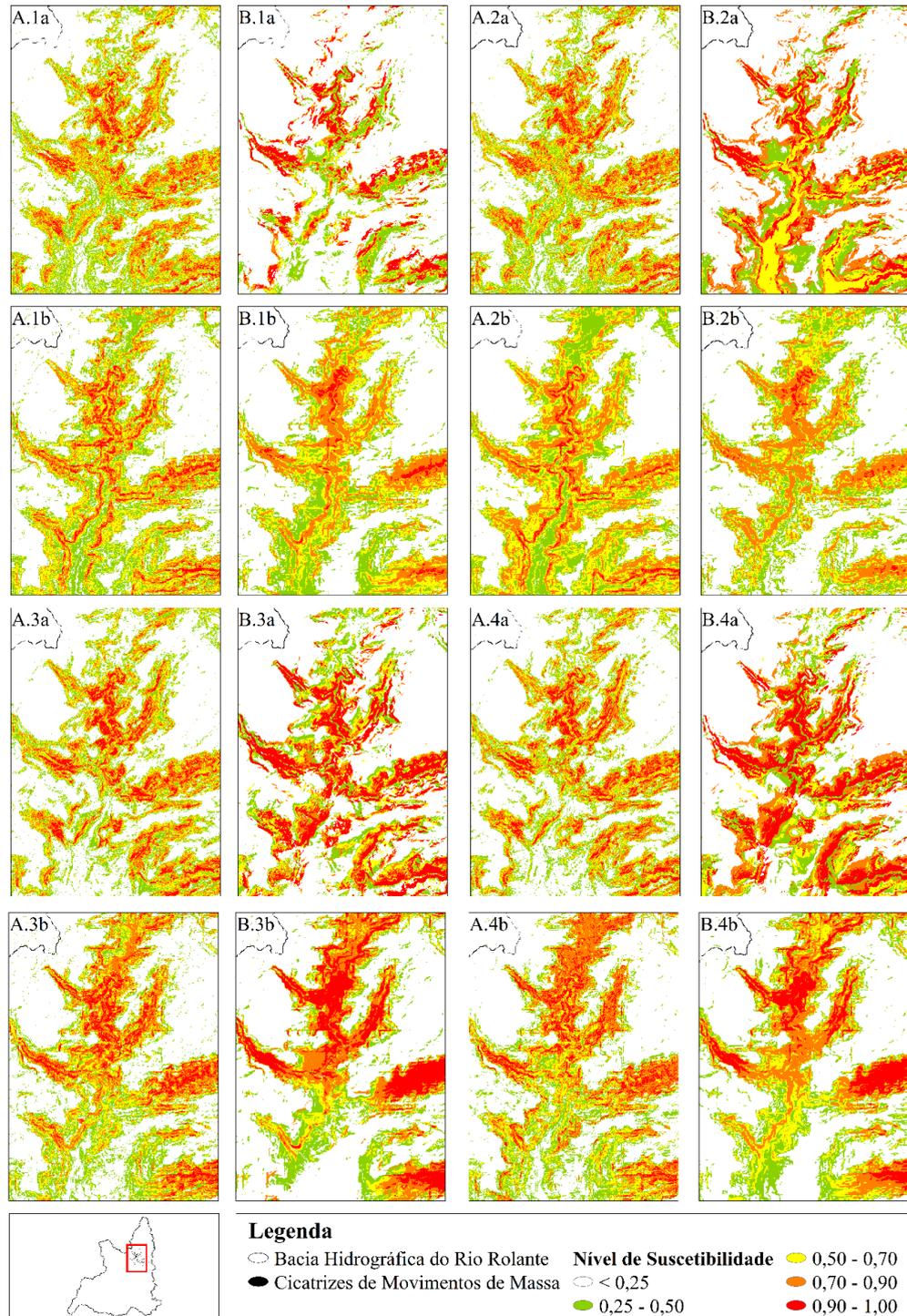


Figura 3.5. Níveis de suscetibilidade conforme o melhor modelo para cada conjunto de dados.

Em que A representa o método RNA; B representa RF; 1 representa a área 1 com reamostragem $x=50$; 2 representa a área 1 com reamostragem $x=80$; 3 representa a área 2 com reamostragem $x=50$; 4 representa a área 2 com reamostragem $x=80$; a representa AP; b representa AS.

Não foram observadas diferenças significativas quanto ao uso das reamostragem $x=50$ ou $x=80$, as quais apresentaram valores similares de acurácia. O teste de proporção demonstrou que em sete de oito cenários não se pode rejeitar a hipótese de que são equivalentes, com nível de significância de 5%. Realizar este tipo de reamostragem, de modo a exigir uma menor capacidade de processamento computacional, se demonstrou adequada e representativa, pois não houve perda de acurácia com a seleção mais rigorosa de amostras para o treinamento. Considerando o MDE AP, houve uma redução entre 7% e 8% no tempo de processamento dos dados, enquanto que para modelagem com MDE AG, a redução foi de 3% a 4%.

Os resultados da Figura 3.5 mostram mapas distintos entre si, ainda que todos os 16 resultados apresentem valores semelhantes de acurácia. Essa diferença revela que, de acordo com o conjunto amostral e o modelo utilizado, o relacionamento entre os atributos do terreno e os níveis de suscetibilidade resultantes do processo de modelagem podem ser bem distintos. Isso é observado em destaque no mapa de diferenças (Figura 5), obtido pela subtração dos dois mapas de maior acurácia, correspondentes à área 2, por reamostragem de $x=50$, com o MDE AG para as RNA e AP para o RF.

O mapa da Figura 3.6 mostra que as áreas com maiores variações quanto aos valores de suscetibilidade estão nas encostas e nas linhas de escarpa erosiva. A figura 3.6a mostra um zoom na área com concentração de cicatrizes e sua distribuição sobre o mapa de diferenças e a figura 3.6b sem considerar a sobreposição das cicatrizes. A área destacada apresenta diferenças elevadas em muitas áreas com cicatrizes. As áreas de vale a jusante não sofreram diferenças significativas, a grande maioria apresentou valores de suscetibilidade abaixo de 0,25.

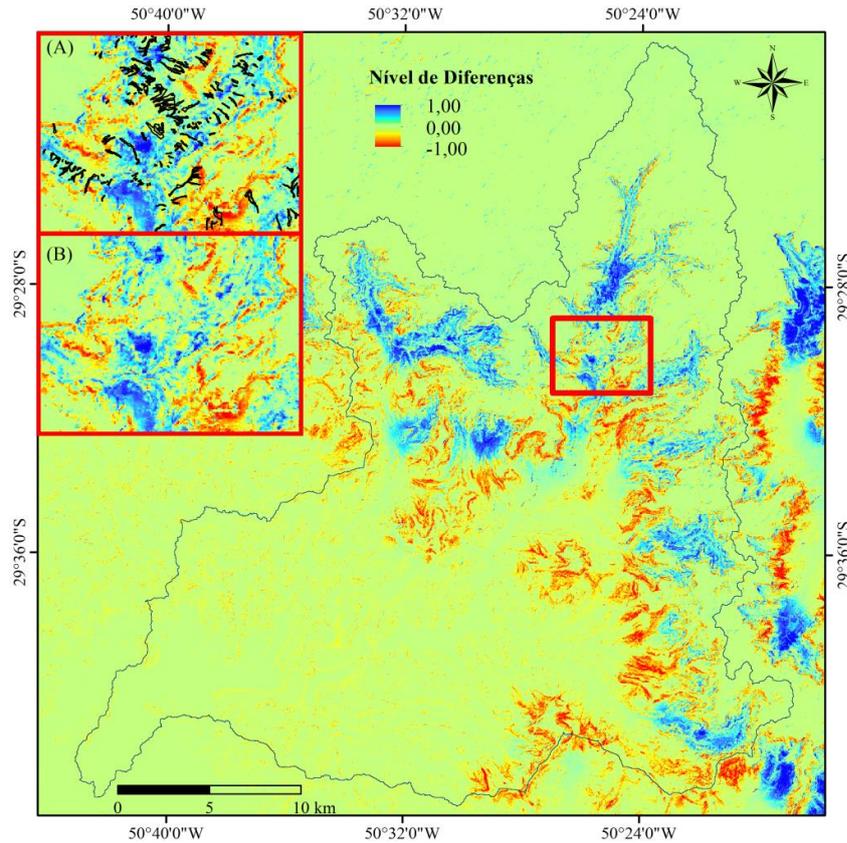


Figura 3.6. Mapa de diferença entre os dois mapas de maior acurácia.

Em função das sensíveis diferenças verificadas nos níveis de suscetibilidade, a escolha do mapa resultante considerando-se apenas aquele que obteve a maior acurácia pode não ser a estratégia mais adequada, ainda que normalmente seja a decisão mais utilizada. Uma alternativa poderia ser a extração de uma média dos modelos que tenham obtido acurácia satisfatória e equivalente, na tentativa de suavizar tais diferenças e identificar um padrão espacial nas manchas de suscetibilidade.

Os mapas de média, Figura 3.7, ilustram os modelos (RNA e RF), os MDEs (AP e AS), as áreas de coleta de amostras e os métodos de reamostragem. Foi extraída a média geral dos 16 mapas, combinando todas as técnicas testadas. A análise dos resultados mostra um padrão das áreas suscetíveis, com maior similaridade entre os resultados. Essa concordância pode representar uma segurança a mais na escolha do mapa considerado ideal para representar a suscetibilidade.

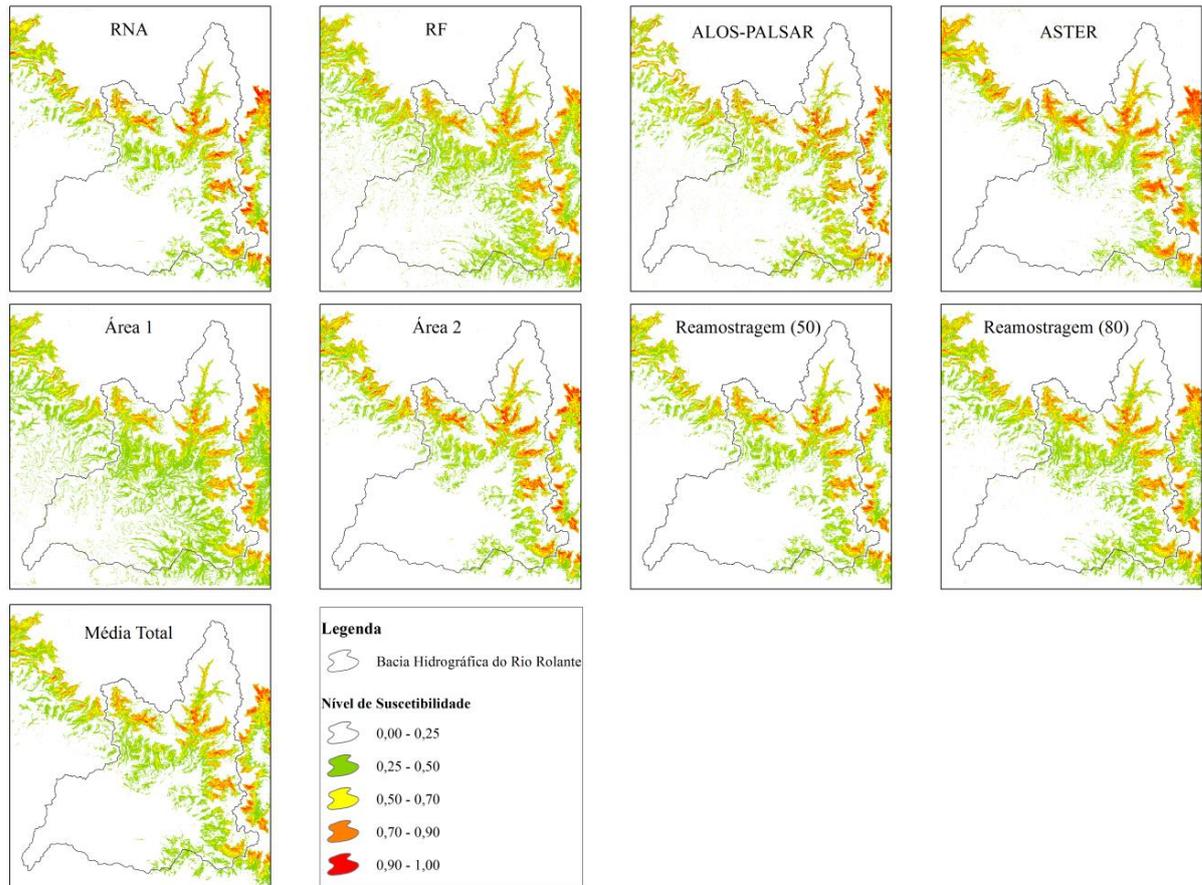


Figura 3.7. Média das imagens resultantes para cada configuração de modelo e método de amostragem.

A comparação entre a média dos métodos mostra que as RNAs apresentam maiores áreas de alta suscetibilidade (0,90 – 1,00) do que o RF. O mesmo acontece quanto ao MDE AS, em relação ao MDE AP, e quanto a área 2, se comparada à área 1. Entretanto, as diferenças verificadas não são tão acentuadas quanto aquelas apresentadas pelos mapas de suscetibilidade da Figura 4.

A análise do desvio padrão do conjunto de imagens resultantes (Figura 3.8) mostra que as maiores variações estão concentradas nas encostas e áreas de maior declividade, justamente as áreas mais suscetíveis. Ainda que todos os modelos indiquem corretamente essas áreas como suscetíveis, os níveis de suscetibilidade variam bastante de um mapa para outro.

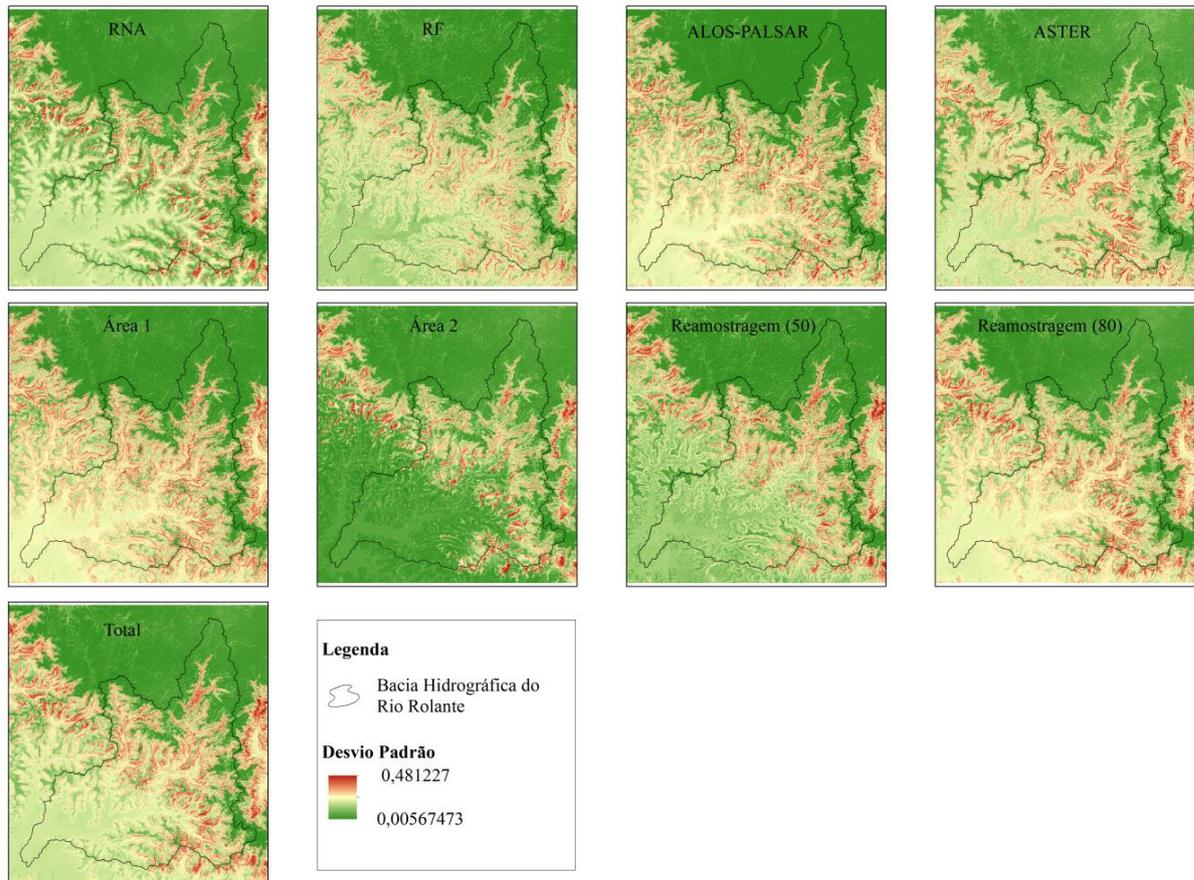


Figura 3.8. Desvio padrão das imagens resultantes para cada configuração de modelo e método de amostragem.

Pode-se destacar que a dispersão dos valores de suscetibilidade foi inferior quando utilizada uma área mais abrangente de amostragem (Área 2), indicando que os mapas modelados com base nesse conjunto amostral apresentam maior similaridade entre si, divergindo em níveis de suscetibilidade apenas na linha da escarpa erosiva. Esses resultados reforçam que a abrangência da área de coleta de amostras influência de forma positiva na generalização dos modelos, contribuindo para uma melhor representação da realidade.

O uso de RNA e de RF demonstrou ser eficaz para o mapeamento de áreas suscetíveis a movimentos de massa. O MDE AP demonstrou AUC levemente maior, porém, ambos os MDEs apresentaram acurácia elevada. Definir uma área amostral de maneira mais abrangente mostrou resultados melhores do que áreas mais restritivas. Enquanto isso, as reamostragens não apresentaram padrão nos resultados, ambos apresentaram AUC elevada. Utilizar a reamostragem de $x=80$ apresenta-se como uma boa estratégia para diminuir a exigência de

processamento quando da calibração dos modelos. Assim, os resultados se demonstraram satisfatórios, com mapeamentos adequados.

3.4 CONCLUSÕES

A comparação entre diferentes métodos de modelagem de áreas suscetíveis a movimentos de massa demonstrou que o atributo do terreno mais importante nas RNA foi a ELV, com valores acima de 30%. Enquanto que no RF foi o FLS que, somado aos parâmetros ELV e PRV apresentaram valores de importância entre 69% e 73%. Quando o FLS apresenta alto valor de importância, a DCV demonstra-se pouco importante, devido presença de informações repetidas nestes fatores. As curvaturas demonstraram pouca importância nas modelagens.

A realização do teste de proporção demonstrou que não se pode rejeitar a hipótese de que os métodos RF e RNA são equivalentes estatisticamente em cinco cenários entre oito. O MDE AP demonstrou AUC levemente maior e se aceita a hipótese de que em dois cenários dentre oito o MDE AP apresentou melhor desempenho, enquanto que nos demais, não se rejeita a hipótese de os MDEs serem equivalentes.

Definir a área amostral de maneira mais abrangente mostrou resultados melhores do que áreas mais restritivas quanto a acurácia e à capacidade de generalização dos modelos, representando de maneira mais adequada a suscetibilidade em áreas de fundo de vales, por exemplo. Com nível de significância de 5%, o teste de proporção demonstrou em todos os cenários que não se rejeita a hipótese de a área 2 apresentar desempenho melhor que a área 1.

As reamostragens não apresentaram padrão nos resultados. A partir do teste de proporção, em sete de oito cenários não se pode rejeitar a hipótese de que são equivalentes. A redução do conjunto amostral de treinamento diminuiu o tempo de processamento entre 3% e 4% para o MDE AG, e entre 7% e 8% para o MDE AP, sem interferir de modo significativo na acurácia dos mapas. Assim, estes processos de amostragem demonstram-se como uma alternativa para grandes conjuntos amostrais. Por fim, conclui-se que o uso de RNA e de RF se demonstraram eficazes para o mapeamento de áreas suscetíveis a movimentos de massa, apresentando acurácias próximas ou superiores a 90%.

3.5 AGRADECIMENTOS

O presente trabalho foi realizado com apoio: i) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001; ii) da Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul – FAPERGS, com apoio financeiro via Edital 01/2017 – ARD, processo 17/2551-0000894-4.

3.6 REFERÊNCIAS

- ADA, M.; SAN, T. Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. **Nat Hazards**, v. 90, p. 237–263, sept. 2018.
- ADITIAN, A.; KUBOTA, T.; SHINOHARA, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. **Geomorphology**, v. 318, p. 101–111, 2018.
- ARNONE, E.; FRANCIPIANE, A.; SCARBACI, A.; PUGLISI, C.; NOTO, L. V. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. **Environmental Modelling and Software**, v. 84, p. 467–481, 2016.
- BRASIL. Lei 12.608, de 10 de Abril de 2012. Institui a Política Nacional de Proteção e Defesa Civil PNPDEC. 2012.
- BREIMAN, L. Random Forests. **Machine Learning**, The Netherlands, v. 45, p. 5–32, 2001.
- CEPED UFSC. **Atlas Brasileiro de Desastres Naturais – 1991 a 2012**. Volume Brasil. 2. ed. rev. amp. Florianópolis: CEPED UFSC, 2013. 127 p.
- CHEN, X.; LIU, C.; CHANG, Z.; ZHOU, Q. The relationship between the slope angle and the landslide size derived from limit equilibrium simulations. **Geomorphology**. v. 253, p. 547-550, feb. 2016.
- CHEN, W.; POURGHASEMI, H. R.; KORNEJADY, A.; ZHANG, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. **Geoderma**, v. 305, p. 314–327, june 2017.
- DAI, F. C.; LEE, C. F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. **Geomorphology**, v. 42, p. 213–228, 2002.
- DOU, J.; YAMAGISHI, H.; POURGHASEMI, H. R.; YUNUS, A. P.; SONG, X.; XU, Y.; ZHU, Z. An integrated artificial neural network model for landslide susceptibility assessment of Osado Island, Japan. **Natural Hazards**, v. 78, p. 1749 - 1776. 2015.

FROUDE, M. J.; PETLEY, D. N. Global fatal landslide occurrence from 2004 to 2016. **Natural Hazards and Earth System Sciences**, v. 18, p. 2161–2181, 2018.

KAWABATA, D.; BANDIBAS, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). **Geomorphology**, v. 113, p. 97–109, 2009.

MEINHARDT, M.; FINK, M.; TÜNSCHEL, H. Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: Comparison of a new method to calculate weighting factors by means of bivariate statistics. **Geomorphology**, v. 234, p. 80–97, jan. 2015.

METEN, M.; PRAKASHBHANDARY, N.; YATABE, R. Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia. **Geoenvironmental Disasters**, v. 2: 9, p. 1 – 17, dec. 2015.

OLIVEIRA, G. G.; GUASSELLI, L. A.; QUEVEDO, R. P.; RUIZ, L. F. C.; BRESSANI, L. A.; RIFFEL, E. S. Identificação e análise de áreas suscetíveis a fluxos de detritos na bacia hidrográfica do Rio Taquari-Antas, RS. **Pesquisas em Geociências**, Porto Alegre, v. 45, n. 2, p. 1–25, 2018.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection. **Engineering Applications of Artificial Intelligence**, v. 40, p. 47–61, 2015.

PANAGOS, P.; BORRELLI, P.; MEUSBURGER, K. A New European Slope Length and Steepness Factor (LS-Factor) for Modeling Soil Erosion by Water. **Geosciences**, v. 5, p. 117–126, 2015. DOI: 10.3390/geosciences5020117

POURGHASEMI, H. R.; GAYEN, A.; PARK, S.; LEE, C. W., LEE, S. Assessment of landslide-prone areas and their zonation using logistic regression, LogitBoost, and naïvebayes machine-learning algorithms. **Sustainability**, Switzerland, v. 10, p. 1-23, oct. 2018.

POURGHASEMI, H. R.; RAHMATI, O. Prediction of the landslide susceptibility: Which algorithm, which precision? **Catena**, v. 162, p. 177–192, mar. 2018.

PRADHAN, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models landslide susceptibility mapping using GIS. **Computer & Geosciences**, v. 51, p. 350-365, 2013.

PRADHAN, B.; LEE, S. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. **Landslides**, v. 7, n. 1, p. 13–30, 2010.

ROSSATO, M. S. **Os Climas do Rio Grande do Sul: variabilidade, tendências e tipologia**. Porto Alegre: UFRGS, 2011. 253 p. Tese (Doutorado em Geografia) – Programa de Pós-Graduação em Geografia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, p. 533- 536, 1986.

SEMA. **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, 26 p. 2017.

WIDROW, B.; HOFF, M. E. Adaptive Switching Circuits. *In:* 1960 IRE WESCON CONVENTION RECORD, New York: IRE Part, 1960. p. 96– 104.

CAPÍTULO 4 : CONSIDERAÇÕES FINAIS

O mapeamento de cicatrizes de movimentos de massa ocorrido na Bacia Hidrográfica do Rio Rolante no evento de precipitação extrema de 05 de janeiro de 2017 resultou em 297 cicatrizes, com área total de 1,06 km². Ao compor as amostras de ocorrência, foram considerados 19.659 píxeis para MDE AP e 3.407 para o MDE AG, os mesmos valores foram considerados para amostras de não ocorrência, compondo a outra metade dos conjuntos amostrais.

Os gráficos de *boxplot* dos atributos de terreno demonstraram que os parâmetros que diferenciam mais facilmente áreas de ocorrência de movimentos de massa são elevação, declividade, fator LS e profundidade de vale. Da mesma forma, os atributos supracitados foram os que apresentaram maior importância nas modelagens por RF e RNA. A declividade, comumente considerada importante, apresentou maior relevância nas RNA e menor no RF. Ressalta-se que quando há dois atributos do terreno com informações semelhantes, a tendência é que um dos fatores receba uma importância elevada, enquanto o outro apresente importância menor, o que ocorreu com a presença do fator LS e da declividade. As curvaturas longitudinal e vertical apresentaram os menores percentuais de importância em todas as modelagens.

As acurácias obtidas variaram entre 0,8762 e 0,9356, demonstrando alta capacidade de modelagem de áreas suscetíveis em ambos os métodos. A modelagem a partir de RF com o MDE AP resultou em valores de acurácia maiores do que com o MDE AS. Quanto às RNA, o MDE AS resultou em maiores AUC, na maioria das configurações. A diferença entre as acurácias dos MDEs AP e AS não foi superior a 0,03, o que demonstra que um conjunto amostral menos numeroso não causou prejuízo nos resultados.

De modo geral, o desempenho do modelo RF apresentou resultados melhores do que as RNA com a utilização do MDE AP, enquanto que as RNA foram melhores com o MDE AS. Ao aplicar-se o teste de proporção, com nível de significância de 5%, foi verificado que em 3 de oito cenários, a hipótese de que os índices de acurácia dos dois modelos são semelhantes deve ser rejeitada, indicando que o modelo RF apresentou desempenho superior à RNA.

Quanto aos MDEs, verificou-se que o MDE AP apresentou AUC maiores e o teste de proporção demonstrou que não se pode rejeitar a hipótese de superioridade deste MDE (AP) quanto ao MDE AG. Ainda assim, deve-se considerar os valores de AUC do MDE AG foram

elevados e sua modelagem se demonstra eficaz, devido ao menor conjunto de dados amostrais, podendo ser utilizado para escalas inferiores a 1:50.000.

As diferentes áreas amostrais demonstraram que a restrição de área para coleta de amostras aleatórias de não ocorrência pode afetar a capacidade de generalização do modelo, pois a área amostral abrangente apresentou resultados melhores quanto acurácia e capacidade de generalização, além de o teste de proporção ter demonstrado que a área 2 é superior à área 1 em todos os oito cenários.

Os resultados das reamostragens não demonstraram padrão, quase todos os testes de proporção resultaram em equivalência. Ao apresentarem elevada acurácia e não interferirem na capacidade de generalização dos modelos, estes métodos podem ser considerados como alternativa para grandes conjuntos amostrais, já que ao serem utilizados diminuem o tempo de processamento (neste estudo a redução esteve entre 3% e 4% para o MDE AG e entre 7% e 8% para o MDE AP).

Conclui-se que o uso de RNA e de RF foram eficazes para o mapeamento de áreas suscetíveis a movimentos de massa, apresentando acurácias próximas ou superiores a 0,09. Este estudo contribuiu para entender a importância da escolha de áreas amostrais que sejam representativas, uma vez que as áreas restritivas podem influenciar de maneira negativa na modelagem, além de demonstrar que a redução no conjunto amostral, se realizada de modo representativo, não afeta necessariamente a AUC dos modelos.

REFERÊNCIAS

- ADITIAN, A.; KUBOTA, T.; SHINOHARA, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. **Geomorphology**, v. 318, p. 101–111, 2018.
- ARNONE, E.; FRANCIPIANE, A.; SCARBACI, A.; PUGLISI, C.; NOTO, L. V. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. **Environmental Modelling and Software**, v. 84, p. 467–481, 2016.
- BREIMAN, L. Random Forests. **Machine Learning**, The Netherlands, v. 45, p. 5–32, 2001.
- CEPED/RS - UFRGS. **Capacitação em gestão de riscos**. 2. ed. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2016. 270 p.
- CEPED UFSC. **Atlas Brasileiro de Desastres Naturais – 1991 a 2012**. Volume Brasil. 2. ed. rev. amp. Florianópolis: CEPED UFSC, 2013. 127 p.
- CHEN, W.; LIU, C.; CHANG, Z.; ZHOU, Q. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. **Catena**, v. 151, p. 147–160, 2017.
- CHRISTOPHERSON, R. W. **Geossistemas: uma introdução à geografia física**. Tradução: AQUINO, F. E. *et al.* 7. ed. Porto Alegre: Bookman, 2012. 752 p.
- CHU, L.; WANG, L.; JIANG, J.; LIU, X.; SAWADA, K.; ZHANG, J. Comparison of landslide susceptibility maps using random forest and multivariate adaptive regression spline models in combination with catchment map units. **Geosciences Journal**, v. 23, n. 2, p. 341–355. 2019.
- DOU, J.; YAMAGISHI, H.; POURGHASEMI, H. R.; YUNUS, A.; SONG, X.; XU, Y.; ZHU, Z. An integrated artificial neural network model for landslide susceptibility assessment of Osado Island, Japan. **Natural Hazards**, v. 78, p. 1749 - 1776. 2015.
- ERMINI, L.; CATANI, F.; CASAGLI, N. Artificial Neural Networks applied to landslide susceptibility assessment. **Geomorphology**, v. 66, p. 327–343, 2005.
- INTERNATIONAL FEDERATION OF RED CROSS AND RED CRESCENT SOCIETIES - IFRC. **World Disasters Report 2016: Resilience : saving lives today, investing for tomorrow**. 2016. 282 p.
- IYODA, E. M. **Inteligência computacional no projeto automático de redes neurais híbridas e redes neurofuzzy heterogêneas**. Campinas: UNICAMP, 2000. 173 p. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2000.
- JENSEN, J. R. **Introductory digital image processing: a remote sensing perspective**. 3. ed., New Jersey: Prentice-Hall, 2005. 526 p.

KAWABATA, D.; BANDIBAS, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). **Geomorphology**, v. 113, p. 97–109, 2009.

KIM, J.; LEE, S.; JUNG, H.; LEE, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. **Geocarto International**, v. 33, n. 9, p. 1000–1015, 2018.

MELO, R.; ZÊZERE, J. L. Avaliação da suscetibilidade à rutura e propagação de fluxos de detritos na bacia hidrográfica do Rio Zêzere (Serra Da Estrela, Portugal). **Revista Brasileira de Geomorfologia**, São Paulo, v. 18, n. 1, p. 81–106, jan./mar. 2017.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010, 557 p.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection. **Engineering Applications of Artificial Intelligence**, v. 40, p. 47–61, 2015.

PRADHAN, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. **Computers and Geosciences**, v. 51, p. 350–365, 2013.

PRADHAN, B.; LEE, S. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. **Landslides**, v. 7, n. 1, p. 13–30, 2010.

QUEVEDO, R. P. **Gestão de risco de desastres no município de Encantado - RS**. Porto Alegre: UFRGS, 2016. 80 p. Trabalho de Conclusão de Curso (Bacharelado em Geografia) - Departamento de Geografia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

RIFFEL, E. S.; GUASSELLI, L. A. Ocorrência de movimentos de massa na bacia hidrográfica do rio paranhana - RS. **Boletim gaúcho de geografia**, Porto Alegre, v. 39, p. 193–209, 2012.

RUMELHART, D. E. ; HILTON, G. E. ; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, p. 533–536, 1986.

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. 3–4, p. 108–121, 2009.

SEMA. **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, 26 p. 2017.

TAALAB, K.; CHENG, T.; ZHANG, Y. Mapping landslide susceptibility and types using Random Forest. **Big Earth Data**, v. 2, n. 2, p. 159–178, 2018. DOI: 10.1080/20964471.2018.1472392

TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. **Desastres naturais**: conhecer para prevenir. 1ª ed. São Paulo: Instituto Geológico, 2009. 197 p.

TRALLI, D. M.; BLOM, R. G.; ZLOTNICKI, V.; DONNELLAN, A.; EVANS, D. L. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 59, p. 185–198, apr. 2005.

UNITED NATIONS. Department of Economic and Social Affairs. Population Division. **World Urbanization Prospects: The 2014 Revision, Highlights**. New York: United Nations, 2014. 32 p.

WITTEN, I. H. .; FRANK, E. **Data Mining**: Practical Machine Learning Tools and Technicals. 2. ed. San Francisco: Elsevier, 2005. 558 p.

YEON, Y. K.; HAN, J. G.; RYU, K. H. Landslide susceptibility mapping in Injae, Korea, using a decision tree. **Engineering Geology**, v. 116, p. 274–283, 2010.
DOI:10.1016/j.enggeo.2010.09.009

ZÊZERE, J. L. **Dinâmica de Vertentes e Riscos Geomorfológicos – Programa**. Relatório n.º14, Centro de Estudos Geográfico, Lisboa, 2005. 128 p. ISBN: 972-636-157-5