

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

Samuel Gameiro

**CAPACIDADE DE GENERALIZAÇÃO E EXTRAPOLAÇÃO ESPACIAL DE REDES
NEURAS ARTIFICIAIS NO MAPEAMENTO A SUSCETIBILIDADE DE
DESLIZAMENTOS**

Porto Alegre
2020

Samuel Gameiro

**CAPACIDADE DE GENERALIZAÇÃO E EXTRAPOLAÇÃO DE REDES NEURAIIS
ARTIFICIAIS NO MAPEAMENTO A SUSCETIBILIDADE DE DESLIZAMENTOS**

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 Mestre em Sensoriamento Remoto e Geoprocessamento.

Área de Concentração: Sensoriamento Remoto aplicado.

Orientador: Laurindo Antonio Guasselli

Co-Orientador: Guilherme Garcia de Oliveira

Porto Alegre

2020

CIP - Catalogação na Publicação

Gameiro, Samuel
Capacidade de generalização e extrapolação espacial
de redes neurais artificiais no mapeamento a
susceptibilidade de deslizamentos / Samuel Gameiro. --
2020.
99 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, 2020.

1. Desastres Naturais. 2. Modelagem. 3. Aprendizado
de Máquina. 4. Sensoriamento Remoto. 5. Amostragem. I.
Guasselli, Laurindo Antonio, orient. II. de Oliveira,
Guilherme Garcia, coorient. III. Título.

Samuel Gameiro

**CAPACIDADE DE GENERALIZAÇÃO E EXTRAPOLAÇÃO DE REDES NEURAIS
ARTIFICIAIS NO MAPEAMENTO A SUSCETIBILIDADE DE DESLIZAMENTOS**

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 Mestre em Sensoriamento Remoto e Geoprocessamento.

Aprovada em 23 de março de 2020

Prof. Dr. Laurindo Antonio Guasselli – Orientador

Prof. Dr. Guilherme Garcia de Oliveira – Co-orientador

Prof. Dr. José Mario Vicente Grzybowski – UFFS

Prof. Dr. Clódis de Oliveira Andrades Filho – UFRGS

Prof. Dr. Cristiano Lima Hackmann - PPGSR

*Dedico esta dissertação a Vida, a minha família
e a todos que emanaram energias boas nessa
minha jornada!!!*

AGRADECIMENTOS

Ao Ser Superior e a toda energia positiva que rege a tudo e todas.

Ao meu pai (Fernando), mãe (Nanci) e irmã (Samira) que sempre estão comigo independente do que aconteça e são a minha força.

A Ana Flávia Correia de Oliveira por todo o tempo juntos. Pelo incentivo, carinho, compreensão, amizade, respeito, confiança, segredos, saudades, presentes, momentos, viagens, paixão e amor! Agradeço por ter despertado sempre o melhor de mim. Este trabalho também é seu pois uma gigantesca parte disso tudo foi por você e graças a você! Te amo!

As famílias que me fazem sentir parte delas e que sempre me apoiaram em tudo:

- Claudinho, Cristiane, Cláudio, Cris e todos tios e primos (Gameiro)
- A todos tios e primos (Bronzato)
- Elaine, Amanda e Maria (Camargo).
- Lares (*in memorian*), Silvia, Lilian, Sabrina, Rodrigo e John (Correia).
- Du, Tati, Dudinha, Vó Dirce, Vô Flávio, Mara, Beto, Flavinho, Ju e todos tios e primos (Oliveira).

Ao Marco por ter sido mais do que um melhor amigo, ter sido um irmão. E por ter ajudado a suportar todos os momentos difíceis, fosse com conversas ou enchendo a cara.

Aos amigos, professores, funcionários e a banca examinadora do PPGSR/CEPSRM/LAGAM/UFFS – Alan, Grazi, Nunes, Giu, Felipe, Fabi, Manu, João(s), Tássia, Douglas, Luiz, Victor, Bruno, Guilherme, Laurindo, Clódis, Cristiano, José e todos que conviviam diariamente comigo por lá e me ajudaram e aconselharam em tudo.

A Dona Olinda, Seu Pontes, Jeff, Lurdes e todos os vizinhos com quem passei esses 2 anos.

Ao Sidir, Suzana e Jadson por todo incentivo, conversas, apoio e por terem apostado em mim.

A Alice, Aline, Amanda, Ana Lu, Andréa, Brenin, Ingrid, Liziane, Pauly, Raíssa, Rayane, Vitin e outros que me ajudaram nos momentos finais, seja me distraíndo, me aconselhando, me fazendo rir ou mesmo dando uma palavra de conforto e incentivo. Cada palavra e cada conversa foi de extrema importância para mim!

E a todos que de alguma forma contribuíram ou mandaram boas energias no decorrer dessa jornada e para a conclusão dessa dissertação.

GRATIDÃO!

“Tenho para mim que o amor é o que há de mais importante no mundo. Analisar o mundo, explicá-lo, menosprezá-lo, talvez caiba aos grandes pensadores. Mas a mim interessa exclusivamente que eu seja capaz de amar o mundo, de não sentir desprezo por ele, de não odiar nem a ele nem a mim mesmo, de contemplar a ele, a mim, a todas as criaturas com amor, admiração e reverência.”

Sidarta

RESUMO

Os deslizamentos de terra podem provocar sérias consequências ambientais, econômicas e sociais. O primeiro passo para a mitigação desses deslizamentos é o mapeamento da suscetibilidade. Entre as diversas técnicas que são atualmente utilizadas para esse mapeamento, destacam-se as redes neurais artificiais (RNA), sendo considerado um dos métodos mais precisos para essa finalidade. Como um dos principais fatores que podem influenciar no desempenho do modelo de RNA e de seus mapas resultantes, afetando ainda sua capacidade de generalização e extrapolação, destaca-se a amostragem. Pensando nisto, esta dissertação tem o intuito de avaliar a influência do processo de amostragem no resultado final dos modelos de suscetibilidade a deslizamentos e como esse processo pode interferir na generalização e extrapolação dos modelos de RNA. Para tal, utilizou-se 16 atributos morfométricos de terreno e confeccionou-se um inventário de cicatrizes de deslizamento utilizando 5 diferentes áreas presentes na Serra Geral do Sul do Brasil. A partir desse inventário, foram criadas as amostras de ocorrência de deslizamento. O mesmo número de amostras foi criado aleatoriamente para serem amostras de não ocorrência, porém, para essas amostras, utilizou-se limitações de distâncias (*buffers*) a fim de se analisar a influência exercida por essa distância nos modelos. No treinamento da rede, o qual utilizou-se do algoritmo retropropagativo, foram usadas, primeiramente, amostras de cada área e, posteriormente, um conjunto de todas as áreas juntas, para se analisar como o modelo se comportaria com o uso de amostras de múltiplos eventos de deslizamentos em comparação com eventos separados. Finalmente, ainda se avaliou a importância de cada variável independente (atributos morfométricos) com relação aos melhores modelos de suscetibilidade gerados. Na análise da distância, os modelos que utilizaram amostras de não ocorrência mais distantes (40 km) foram os que obtiveram maiores acurácias. Isso aconteceu pelo fato de que quanto maiores as distâncias, mais compartimentos do relevo são utilizados para amostras de não ocorrência, o que facilita o processo de modelagem pela RNA. Analisando as amostras de eventos separados e múltiplos eventos, notou-se a maior acurácia e capacidade de extrapolação para os múltiplos eventos, atingindo acurácias de 0,95, contra 0,93 dos que utilizaram apenas um único evento. Essa capacidade se dá pela maior variação e amplitude dos dados utilizados para treinar o modelo, fazendo com que as amostras de treinamento atinjam valores mais significativos de toda a Serra Geral e não de apenas uma área específica dentro da Serra. Na relação dos atributos, percebeu-se que os mais importantes foram a declividade, o fator LS, a profundidade de vale com índices de relação linear maiores do que 0,36. Entretanto a elevação ainda demonstrou ser um fator de relevância para a modelagem de suscetibilidade a deslizamentos, mesmo com valores mais baixos de relação, principalmente quando utilizado somente um evento de deslizamento para o treinamento da rede.

Palavras-chave: amostragem, aprendizagem de máquina, acurácia, desastres naturais

ABSTRACT

Landslides can have serious environmental, economic and social consequences. The first step in mitigating these landslides is the mapping of susceptibility. Among the various techniques that are currently used for this mapping, artificial neural networks (ANN) stand out, being considered one of the most accurate methods for this purpose. As one of the main factors that can influence the performance of the RNA model and its resulting maps, also affecting its ability to generalize and extrapolate, sampling stands out. With this in mind, this dissertation aims to evaluate the influence of the sampling process on the final result of the slide susceptibility models and how this process can interfere in the generalization and extrapolation of the RNA models. To this end, 16 terrain morphometric attributes were used and an inventory of sliding scars was made using 5 different areas present in the Serra Geral do Sul do Brasil. From this inventory, samples of slip occurrence were created. The same number of samples was created randomly to be non-occurrence samples, however, for these samples, limitations of distances (buffers) were used in order to analyze the influence exerted by this distance on the models. In the training of the network, which used the retropropagative algorithm, samples of each area were used first and, subsequently, a set of all areas together, to analyze how the model would behave with the use of multiple samples. slip events compared to separate events. Finally, the importance of each independent variable (morphometric attributes) was also evaluated in relation to the best susceptibility models generated. In the analysis of the distance, the models that used more distant samples (40 km) were the ones that obtained greater accuracy. This was due to the fact that the greater the distances, the more relief compartments are used for non-occurrence samples, which facilitates the process of differentiating between occurrence and non-occurrence by ANN. Analyzing the samples of separate events and multiple events, it was noted the highest accuracy and extrapolation capacity for multiple events, reaching accuracy of 0,95, against 0,93 of those who used only a single event. This capacity is due to the greater variance and amplitude of the data used to train the model, making the training samples reach more significant values from the entire Serra Geral and not from just a specific area within the Serra. In the relation of the attributes, it was noticed that the most important ones were the slope, the LS factor, the valley depth with linear relation indexes greater than 0,36. However, the elevation still proved to be a relevant factor for the modeling of susceptibility to landslides, even with lower values of ratio, especially when using only one landslide event for training the network.

Keywords: sampling, machine learning, accuracy, natural disaster

LISTA DE FIGURAS

Figura 2.1. Classificação e regressão em aprendizado supervisionado. Fonte: Sousa, 2019.....	07
Figura 2.2. Clusterização em aprendizado não-supervisionado.....	07
Figura 2.3. Modelo artificial de neurônio biológico Fonte: Neto & Bonini, 2010.....	10
Figura 2.4. Rede unidirecional com única camada de neurônios Fonte: Haykin, 2007.....	11
Figura 2.5. Rede unidirecional com multicamadas de neurônios. Fonte: Haykin, 2007.....	12
Figura 2.6. Rede recorrente com neurônios ocultos. Fonte: Haykin, 2007.....	13
Figura 3.1. Fluxograma das etapas do trabalho.....	16
Figura 4.1. Localização da área de estudo e das amostras de cicatrizes.....	23
Figura 4.2. Fluxograma da metodologia utilizada.....	24
Figura 4.3. Exemplo de buffers utilizados para amostragem de não ocorrência.....	26
Figura 4.4. Atributos morfométricos de terreno utilizados como entrada na modelagem.....	29
Figura 4.5. Gráficos de boxplots com os parâmetros mais importantes para o <i>buffer</i> de 40 km.....	32
Figura 4.6. Comparação visual dos melhores modelos com amostras de treinamento de cada área: (A) Maquiné; (B) Rolante; (C) Forromeco; (D) Fão; (E) Timbé do Sul.....	33
Figura 4.7. Mapa de suscetibilidade a deslizamentos, na Serra Geral, sul do Brasil.....	37
Figura 4.8. Suscetibilidade do melhor modelo para as áreas amostrais dos deslizamentos.....	37
Figura 5.1 Área de estudo e localização das áreas amostrais de cicatrizes a deslizamentos.....	51
Figura 5.2 Exemplo de distâncias de <i>buffers</i> utilizados para amostragem de não ocorrência...	53
Figura 5.3 Atributos morfométricos de terreno utilizados como entrada na modelagem.....	54
Figura 5.4 Comparação entre os modelos que utilizaram amostras totais.....	64
Figura 5.5 Comparação das áreas de cicatrizes entre os melhores modelos: A) melhor modelo com amostras separadas, <i>single model</i> ; B) melhor modelo com amostras conjuntas, <i>full model</i>	65
Figura 5.6 Subtração entre o <i>full model</i> e o <i>single model</i>	66
Figura 5.7. Gráficos da relação entre o índice de suscetibilidade a deslizamentos dos melhores modelos e os principais atributos morfométricos.....	69
Figura 5.8 Relação entre o índice de suscetibilidade e a elevação para os melhores modelos. A) Uso de todas as amostras para o <i>single model</i> ; B) Uso de todas as amostras para o <i>full model</i> ; C) Uso das amostras da área 0 para o <i>single model</i> ; D) Uso das amostras da área 0 para o <i>full model</i>	70

LISTA DE TABELAS

Tabela 2.1. Elementos de um neurônio artificial Fonte: Nola, 2015.....	09
Tabela 3.1. Datas das imagens utilizadas para mapeamento de cicatrizes.....	17
Tabela 4.1. Correlação linear entre atributos morfométricos e a variável dependente do modelo.....	30
Tabela 4.2. Valores mínimos médios e máximos das variáveis de entrada, separados por <i>buffer</i> de amostras de não ocorrência.....	30
Tabela 4.3. Importância relativa de cada atributo no melhor modelo.....	32
Tabela 4.4. Acurácia do melhor modelo, considerando apenas as amostras de teste, separados por área de treinamento e <i>buffer</i> de amostras de não ocorrência.....	33
Tabela 5.1 Localização das áreas amostrais de deslizamento na Serra Geral.....	52
Tabela 5.2 Valores estatísticos dos atributos morfométricos de terreno.....	55
Tabela 5.3 Descrição das áreas e atributos morfométricos mais importantes nas amostras de cicatrizes.....	57
Tabela 5.4 Médias dos atributos em cada um dos melhores modelos para as amostras de ocorrência.....	58
Tabela 5.5 Coeficientes de correlação linear das variáveis independentes com a variável dependente do modelo (ocorrência ou não ocorrência de deslizamentos), por <i>buffer</i>	59
Tabela 5.6 Indicadores de acurácia global e erros de omissão e comissão para os modelos com amostras totais. VP = verdadeiro positivo; FP = falso positivo; FN = falso negativo; VN = verdadeiro negativo.....	61
Tabela 5.7 Indicadores de acurácia global e erros de omissão e comissão para os melhores modelos com amostras de cara área. VP = verdadeiro positivo; FP = falso positivo; FN = falso negativo; VN = verdadeiro negativo.....	61
Tabela 5.8 Acurácias médias dos modelos, considerando apenas as amostras de teste, separados por área de treinamento e <i>buffer</i> de amostras de não ocorrência.....	63
Tabela 5.9 Área com elevada suscetibilidade a deslizamento para cada um dos melhores modelos.....	64
Tabela 5.10 Valores de importância atribuídos pela RNA.....	67

LISTA DE ABREVIATURAS E SIGLAS

AD – Árvore de decisão

AHS - *Analytical Hillshading*

AM – Aprendizado de máquina

ASP - *Aspect*

BSC - *buffer controlled sampling*

CIN - *Convergence Index*

CNBL - *Channel Network Base Level*

CSC - *Cross-section curvature*

ELV - *Elevation*

FA – Floresta aleatória

FAC - *Flow Accumulation*

FN - Falsos negativos

FP - Falsos positivos

IC – Índice de contribuição

LOC - *Longitudinal Curvature*

LSF - *LS Factor*

MDE – Modelo Digital de Elevação

MIB - *Mass Balance Index*

MVS – Maquinas de vetores de suporte

RCI – *Relative contribution index*

RNA – Redes neurais artificiais

ROC - *Area Under the Receiver Operating Characteristic Curve*

RSP - *Relative Slope Position*

SIG – Sistemas de informações geográficas

SLP - *Slope*

SR – Sensoriamento remoto

SRTM - *Shuttle Radar Topography Mission*

TWI - *Topographic Wetness Index*

TPI - *Topographic Position Index*

USGS - *United States Geological Survey*

VDCN - *Vertical Distance to Channel Network*

VDP - *Valley Depth*

VN - Verdadeiros negativos

VP - Verdadeiros positivos

SUMÁRIO

CAPÍTULO 1: INTRODUÇÃO	1
1.1 CONTEXTUALIZAÇÃO	1
1.2 PROBLEMAS DE PESQUISA	3
1.3 HIPÓTESES E OBJETIVOS	4
1.4 ESTRUTURA DA DISSERTAÇÃO	5
CAPÍTULO 2: REVISÃO TEÓRICA	6
2.1 APRENDIZAGEM DE MÁQUINA	6
CAPÍTULO 3: MATERIAIS E MÉTODOS	16
3.1 METODOLOGIA	16
CAPÍTULO 4: REDES NEURAIS ARTIFICIAIS APLICADAS À SUSCETIBILIDADE A DESLIZAMENTOS: O EFEITO DAS ÁREAS AMOSTRAIS NA CAPACIDADE DE GENERALIZAÇÃO E EXTRAPOLAÇÃO DO MODELO	19
4.1 INTRODUÇÃO	21
4.2 METODOLOGIA	22
4.2.1 Área de estudo	22
4.2.2 Preparação dos dados	24
4.2.3 Inventário de cicatrizes	25
4.2.4 Fatores condicionantes de deslizamentos	26
4.3 MODELAGEM E VALIDAÇÃO	28
4.4 RESULTADOS	29
4.4.1 Análise de correlação e importância das variáveis	29
4.4.2. Análise de acurácia	33
4.5 CONCLUSÕES	38
4.6 AGRADECIMENTOS	39
4.7 REFERÊNCIAS	39
CAPÍTULO 5: COMPARAÇÃO ENTRE DISTINTAS ÁREAS DE AMOSTRAGEM NO MAPEAMENTO DE SUSCETIBILIDADE A DESLIZAMENTOS COM REDES NEURAIS ARTIFICIAIS	47
5.1 INTRODUÇÃO	49
5.2 METODOLOGIA	50
5.2.1 Área de estudo	50
5.2.2 Conjunto de dados	52

5.2.3 Inventário de cicatrizes	52
5.2.4 Fatores condicionantes de deslizamento.....	53
5.2.5 Modelagem e validação	55
5.3. RESULTADOS E DISCUSSÕES	56
5.3.1 Inventário de cicatrizes e morfometria das áreas amostrais	56
5.3.2 Seleção dos atributos	59
5.3.4 Análise e comparação dos modelos.....	61
5.3.5 Importância dos atributos morfométricos.....	67
5.4 CONCLUSÕES	71
5.5 AGRADECIMENTOS	72
5.6 REFERÊNCIAS	72
CAPÍTULO 6: CONSIDERAÇÕES FINAIS	77
CAPÍTULO 7: REFERÊNCIAS	79

CAPÍTULO 1: INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Os desastres naturais sempre estiveram presentes na vida das populações em todas as regiões do mundo, mas devido ao aumento de suas ocorrências (Tominaga *et al.*, 2009; Paixão *et al.*, 2014) e da intensidade dos danos e prejuízos causados, vem recebendo crescente atenção de órgãos públicos e de instituições de pesquisa. Para Marcelino (2008) os desastres naturais podem ser definidos como o resultado do impacto de fenômenos naturais extremos ou intensos sobre um sistema social, causando sérios danos e prejuízos. Os desastres naturais que ocorrem no Brasil são, na sua maioria, de origem atmosférica, com a precipitação, seja em excesso ou em escassez, sendo o elemento atmosférico que mais contribui para a ocorrência de desastres na região Sul/Sudeste do Brasil (Barbieri *et al.*, 2009).

Entre os principais desastres naturais no Brasil, destacam-se os deslizamentos, definidos por Tominaga *et al.* (2009) como movimentos de solo, rocha e/ou vegetação ao longo da vertente sob a ação direta da gravidade. A ocorrência desses deslizamentos passou a interessar mais a comunidade científica e a sociedade de forma geral, à medida que um contingente cada vez maior da população passou a ser afetada por tais processos, de forma direta ou indireta (Pinto *et al.*, 2012). Nos últimos 60 anos vários episódios de deslizamentos podem ser destacados no Brasil, entre eles estão os eventos em Minas Gerais (1948, 2017, 2018, 2019), São Paulo (1956, 1967, 2010, 2018, 2019), Rio de Janeiro (1966, 1967, 2010, 2011, 2018), Ceará (1974), Santa Catarina (1974, 1995, 2008, 2017) e no Rio Grande do Sul (2011, 2017, 2018) (modificado de Guidicine & Nieble, 1984).

Para a prevenção e análise desses deslizamentos, ressalta-se a utilização de sensoriamento remoto (SR) e os sistemas de informação geográfica (SIG). Segundo Blaschke & Kux (2005), o SR é uma ferramenta importante para a obtenção de dados básicos e espaciais, para realização de tarefas de proteção do ambiente, bem como para decisões de planejamento, fornecendo características básicas para a observação do meio. Enquanto que os SIGs permitem agrupar diversas camadas de informações sobre a área afetada por desastres naturais, cruzar múltiplos dados e criar mapas do terreno que possibilitem uma rápida análise da situação.

De acordo com IPT (2007), Vanacôr (2012), Riffel (2017) e Quevedo *et al.* (2019), o mapeamento de suscetibilidade é o primeiro passo no sentido de mitigar efetivamente áreas suscetíveis a deslizamentos e minimizar as consequências negativas destes eventos na

sociedade. A utilização de diferentes métodos para mapear e modelar áreas suscetíveis a deslizamentos têm se tornado mais frequente. Um modelo pode ser descrito como uma representação simplificada e que auxilia no processo de entendimento da realidade através de objetos e processos, permitindo a realização de simulações (Lopes, 2006).

Na modelagem de suscetibilidade a deslizamentos, os modelos empíricos são amplamente utilizados. Esses modelos levam em conta a relação entre as variáveis utilizadas fazendo uso de relações baseadas em observações e possuem a vantagem de poderem ser adaptados a cada região (Rennó, 2003). Nesses processos de modelagem, destaca-se a utilização da aprendizagem de máquina (AM), que são métodos que extraem conhecimento através de amostras de dados e podem adaptar-se e aprender a partir da experiência. Nesta técnica, os modelos passam por uma fase de treinamento através de amostras de ocorrência e não ocorrência, uma fase de validação e uma fase final, na qual o modelo é testado com amostras diferentes das usadas no treinamento. Como modelos que utilizam essa técnica e estão em destaque atualmente, pode-se citar as redes neurais artificiais (RNA) (Yilmaz, 2009; Park, 2013; Aditian, 2018), Máquinas de Vetores de Suporte (MVS) (Huang, 2017; Pourghasemi, 2017; Kalantar, 2018) e Florestas Aleatórias (FA) (Pourghasemi, 2017; Sahin *et al.*, 2018; Ada & San, 2018), os quais tem apresentado resultados satisfatórios (precisão maiores de 85%) para identificar áreas mais suscetíveis a esses eventos.

Os modelos de suscetibilidade levam em consideração análises de parâmetros morfométricos e/ou estatísticos da área em estudo. A definição desses parâmetros é de extrema importância, visto que um elevado número de parâmetros tende a elevar o custo computacional do processo (Al-Najjar *et al.*, 2019; Wang *et al.*, 2019) e ainda podem interferir drasticamente no resultado final da modelagem. Por esse motivo, vários são os métodos utilizados para a melhor escolha desses parâmetros, tais como correlação de Pearson (Wang *et al.*, 2020; Lucchese *et al.*, 2020), taxa de ganho de informações (Bui *et al.*, 2016; Zhou *et al.*, 2018) e o fator de inflação de variância (Chen *et al.*, 2017a; Zhao *et al.*, 2019).

Como são empíricos, esses modelos podem apresentar problemas quanto à generalização e extrapolação dos resultados se não forem adequadamente treinados. Lorena e Carvalho (2007) definem generalização como a capacidade de um classificador prever corretamente a classe de novos dados do mesmo domínio em que o treinamento ocorreu, enquanto que a extrapolação pode ser definida como uma tentativa de previsão através do incremento nos valores de entrada (Araújo, 2015), ou seja, a utilização de valores de teste não contidos dentro dos intervalos máximos e mínimos das amostras de treinamento. Ao simular vazões hidrológicas Oliveira *et al.* (2014) utilizaram, para a validação do modelo, amostras com valores que excediam os

limites utilizados para treinar o modelo e constataram uma razoável capacidade de extrapolação das RNA. Höfig *et al.* (2014) utilizaram Árvores de Decisão (AD) para mapeamento digital de solo com acurácias de até 80% na extrapolação do mapa de solo para áreas fisiograficamente semelhantes, ressaltando que as variáveis de entrada e o MDE utilizado influenciam diretamente no resultado da extrapolação devido principalmente a distorções em áreas de declividade.

Além dos parâmetros morfométricos, outro fator importante na modelagem de suscetibilidade a deslizamentos é a amostragem. Diferentes técnicas são utilizadas para o processo de amostragem. Para amostras de ocorrência existem as técnicas que utilizam apenas o centróide (Chen *et al.*, 2017b), o local de ruptura (Hong *et al.*, 2019) e as que utilizam todos os pixels presentes nas cicatrizes (Quevedo *et al.*, 2019). Para as amostras de não ocorrência, de acordo com Xiao *et al.* (2010) existem 2 técnicas altamente empregadas, sendo elas a aleatoriedade em toda a área de estudo e o método *buffer controlled sampling* (BCS), o qual delimita uma distância mínima/máxima para a amostragem.

Apesar disso, para estudos de suscetibilidade a deslizamento, há uma carência de estudos que avaliem a influência das áreas amostrais na capacidade de generalização e extrapolação dos resultados e, por esse fator, o presente trabalho analisa a capacidade que esses modelos possuem na generalização dos dados e a sua eficiência na extrapolação desses mapeamentos de suscetibilidade para grandes áreas e com distintas áreas amostrais.

1.2 PROBLEMAS DE PESQUISA

Diante do que foi exposto na contextualização e visto que há uma carência nos estudos que avaliem como as áreas amostrais influenciam no resultado final de suscetibilidade com o uso de RNA, tem-se as seguintes questões:

- i) No processo de amostragem para não ocorrência, quão significativa é a influência da distância nos resultados finais da modelagem de áreas suscetíveis a deslizamentos com RNA?
- ii) O quanto as áreas amostrais influenciam na capacidade de extrapolação das RNA?
- iii) Qual a variação no resultado final quando se utilizam simples ou múltiplos eventos no treinamento das RNA?

- iv) Como avaliar a influência dos atributos morfométricos do terreno (variáveis independentes) na composição final dos mapas de suscetibilidade?

1.3 HIPÓTESES E OBJETIVOS

Este estudo parte da premissa que modelos de suscetibilidade a deslizamentos feitos a partir de RNA, por serem empíricos, possuem restrita capacidade de generalização e extrapolação do domínio numérico ao qual o modelo foi treinado, tendo seu desempenho comprometido à medida que há a extrapolação desse domínio com relação às amostras de treinamento. Entretanto, a extrapolação espacial dentro de uma área geomorfologicamente mais homogênea deve apresentar bons resultados, principalmente se os valores referentes aos atributos do terreno forem semelhantes entre as áreas de treinamento e de teste.

Assume-se ainda que a distância no processo de amostragem para não ocorrência é um fator determinante para boas acurácias e representações da suscetibilidade, uma vez que com amostras coletadas mais distantes dos deslizamentos e conseqüentemente com valores numéricos mais diferenciados, a RNA terá maior capacidade de separar as áreas suscetíveis das não suscetíveis.

Assim, o presente estudo parte da hipótese de que as RNA são capazes de realizar a extrapolação espacial entre áreas de estudos que sejam semelhantes e, que elas também sofrem influência direta da distância as quais são realizadas as amostragens.

A fim de verificar essa afirmação, os seguintes objetivos foram executados:

- Analisar a influência da distância do processo de amostras de não ocorrência nas acurácias e nos mapas finais;
- Apresentar como diferentes áreas amostrais podem influenciar na capacidade de extrapolação das RNA;
- Explicar como, quando usados para o treinamento da RNA, o uso de um único evento e de múltiplos eventos de deslizamentos podem alterar os resultados finais;
- Analisar a influência dos atributos morfométricos de terreno na composição do mapa final de suscetibilidade.

1.4 ESTRUTURA DA DISSERTAÇÃO

Nesta dissertação foram elaborados mapeamentos de suscetibilidade, a partir do uso de um modelo digital de elevação (MDE), de técnicas de aprendizagem de máquina (AM) e de diferentes conjuntos amostrais com o objetivo de verificar a influência da amostragem na capacidade de generalização, extrapolação e acurácia do mapa final.

Este estudo está estruturado em sete capítulos:

- i) O primeiro capítulo é introdutório e apresenta uma contextualização acerca de deslizamentos e modelos de suscetibilidade, bem como os problemas de pesquisa e as hipóteses e objetivos;
- ii) O segundo capítulo traz uma revisão teórica a respeito de AM, com ênfase na técnica de RNA, a qual foi utilizada na dissertação;
- iii) O terceiro capítulo mostra os materiais utilizados como base para o processo de modelagem e descreve a metodologia utilizada;
- iv) O capítulo quatro refere-se ao primeiro artigo da dissertação, o qual trata sobre a influência de diferentes áreas amostrais na capacidade de generalização e extrapolação dos modelos de redes neurais artificiais (RNA), considerando distintos atributos morfométricos extraídos a partir do MDE *Shuttle Radar Topography Mission* (SRTM - 90 m) e distintas áreas de amostragem;
- v) O capítulo cinco consta do segundo artigo da dissertação e apresenta a comparação entre os modelos gerados levando em consideração as distintas áreas de amostragem separadamente e os modelos confeccionados a partir da mistura de todas as amostras das diferentes regiões, bem como a influência dos atributos morfométricos no processo de modelagem com RNA;
- vi) Finalmente, o sexto capítulo aborda as considerações finais e conclusões acerca dos resultados e discussões apresentados neste estudo.

CAPÍTULO 2: REVISÃO TEÓRICA

2.1 APRENDIZAGEM DE MÁQUINA

A aprendizagem de máquina é um campo de pesquisa da inteligência computacional que estuda o desenvolvimento de métodos capazes de extrair conceitos (conhecimento) a partir de amostras de dados, que tem como objetivo construir modelos computacionais que podem adaptar-se e aprender a partir da experiência (Mitchell, 1997).

Em geral, os diversos algoritmos de AM são utilizados de forma a gerar classificações/reconhecimento de padrões, modelagens, simulações e previsões numéricas para um conjunto de amostras (Russel & Norvig, 2010). Portanto, as técnicas de AM são empregadas na indução (a partir de um conjunto de treinamento) de um classificador, que deve ser capaz (idealmente) de prever a classe de instâncias quaisquer do domínio em que ele foi treinado.

Em termos práticos, algoritmos de aprendizado de máquina têm como objetivo descobrir o relacionamento entre as variáveis de um sistema (entrada/saída) a partir de dados amostrados (Cherkassky & Mulier, 2007). Sendo assim, eles não são necessários quando os relacionamentos entre todas as variáveis do problema (entrada/saída) são completamente compreendidos. Este definitivamente não é o caso de muitos dos problemas reais com os quais nos deparamos em nosso dia-a-dia (Von Zuben, 2013).

Três paradigmas podem ser utilizados na geração de um preditor por meio de técnicas de AM: supervisionado, não-supervisionado e por reforço (Haykin, 2007). A escolha de um paradigma de aprendizado determina a maneira como o algoritmo de AM se relaciona com seu meio ambiente, ou seja, o modo como ocorrerá o seu aprendizado por meio de um conjunto de dados.

No aprendizado supervisionado, há a intervenção externa, a qual apresenta um conjunto de exemplos na forma de entrada e saída. De acordo com Monard & Baranauskas (2003), nesse método é fornecido ao algoritmo um conjunto de exemplos com rótulos de classe conhecida, sendo estes exemplos descritos como vetores de valores de características ou atributos e o rótulo da classe associada. Sousa (2019) afirma ainda que no aprendizado supervisionado, temos uma entrada e uma saída bem definida e já é de conhecimento que existe uma relação entre elas. Ressalta-se ainda que o aprendizado supervisionado é dividido em 2 métodos: classificação e regressão (Figura 2.1). A classificação consiste em problemas nos quais os dados de entrada já são conhecidos e o algoritmo tem a tarefa de atribuir esses dados em determinadas classes (Sousa, 2019). O método da regressão consiste em um problema no qual os dados de entrada e

saída se relacionam através de uma função contínua e, a saída desse modelo é, normalmente, um número (Sousa, 2019).

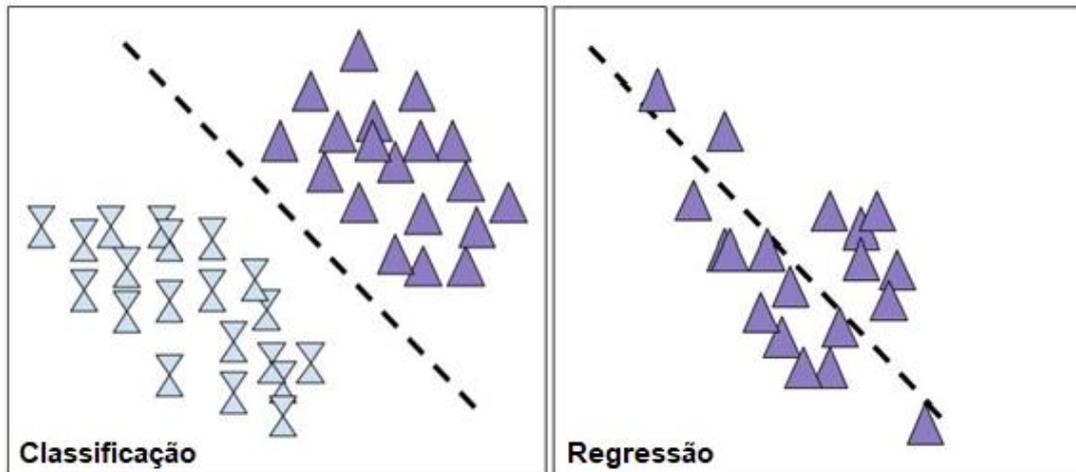


Figura 2.1. Classificação e regressão em aprendizado supervisionado. Fonte: Sousa, 2019.

No aprendizado não supervisionado não há a presença dessa intervenção externa, forçando o algoritmo a aprender e representar as entradas submetidas segundo uma medida de qualidade. Sousa (2019) expõe que nesse método somente os dados de entrada são conhecidos e o principal objetivo deste método é encontrar padrões ou regularidades no conjunto de entrada. Ao contrário do aprendizado supervisionado, neste método é gerado um agrupamento ou uma clusterização (Figura 2.2). Esse processo consiste no agrupamento dos dados de entrada os quais possuem informações em comum ou que sejam semelhantes entre si, formando diferentes grupos ou diferentes *clusters* (Sousa, 2019). A clusterização realiza esse agrupamento e analisa o grau de semelhanças e diferenças entres os diversos *clusters* gerados no processo.

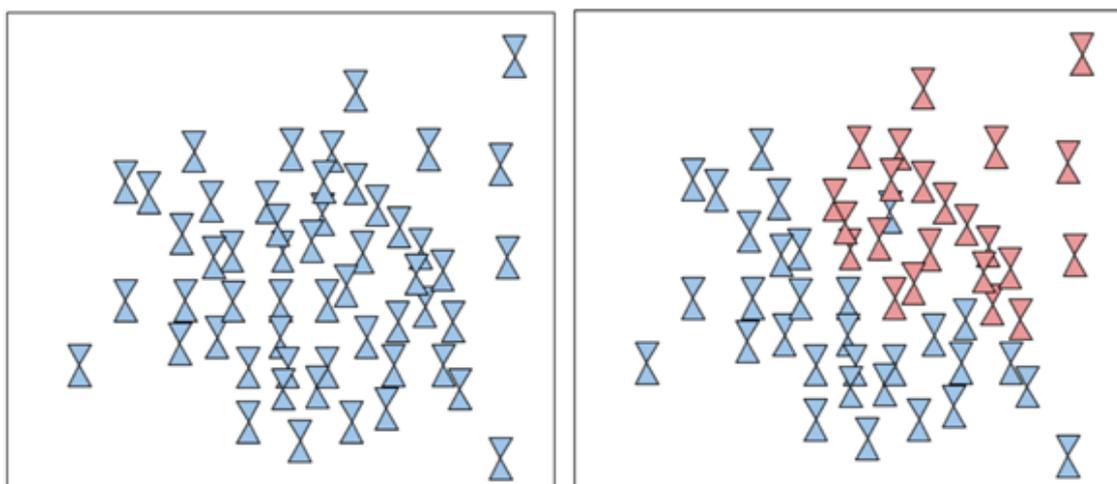


Figura 2.2. Clusterização em aprendizado não-supervisionado. Fonte: Sousa, 2019.

E no reforço, o aprendizado se dá por meio de recompensas ao indutor. Nesse método, assim como no aprendizado não-supervisionado, o agente não precisa de um modelo de dados de saída e entradas pré-definidos (Harmendani & Zanetti, 2019). Aqui, o aprendizado é estabelecido explorando-se um ambiente desconhecido e concentrando-se em receber recompensas após a execução de determinadas ações (Russel & Norvig, 2010). O que determina as recompensas e os próximos estágios do ambiente são as ações tomadas pelo agente (Harmendani & Zanetti, 2019).

No desenvolvimento de técnicas de AM, diversos pesquisadores se inspiraram nos sistemas biológicos. Como exemplo disso, têm-se as RNA, cujos conceitos são baseados nos mecanismos de aprendizado cerebrais e o seu processo de treinamento pode, então, ser caracterizado como aprendizado indutivo, sendo que o uso posterior da rede neural treinada para classificação, regressão ou agrupamento de dados é geralmente denominado de processo de inferência dedutiva (Von Zuben, 2013).

As técnicas de AM possuem inúmeros exemplos de aplicações de sucesso como em categorização de textos (Joachims, 2002), análise de imagens (Pontil & Verri, 1998; Kim *et al.*, 2002), bioinformática (Schölkopf *et al.*, 2003; Noble, 2004) e para mapeamentos de suscetibilidade a deslizamentos (Ballabio & Sterlacchini, 2012; Bui *et al.*, 2012; Conforti *et al.*, 2014; Kalantar *et al.*, 2018; Oliveira *et al.*, 2019; Quevedo *et al.*, 2019).

Para mapeamentos de suscetibilidade a deslizamentos, as bases mais utilizadas como entradas de modelos de AM são geologia, uso do solo, elevação, declividade e curvatura do relevo, sendo parâmetros que estão presentes em mais de 85% dos trabalhos a respeito de deslizamentos. Em sua grande maioria, as técnicas de AM utilizadas para mapeamentos de suscetibilidade a deslizamentos demonstram resultados satisfatórios, com taxas de acerto variando entre 75% e 95%, dependendo da precisão e resolução das bases usadas, dos parâmetros, da estrutura do modelo e do processo de amostragem, sendo este último um grande influenciador do modelo.

RNA é um dos métodos mais comuns utilizados na modelagem de deslizamentos (Kumar *et al.*, 2017). Vários estudos empíricos mostram que as RNA são técnicas úteis para o mapeamento de suscetibilidade a deslizamentos (Lee *et al.*, 2004; Pradhan & Lee, 2010; Losasso & Sdao, 2018).

De acordo com Nola (2015), as RNA são sistemas computacionais estruturados para receber informações, interpretá-las e tomar decisões apropriadas a partir de um treinamento. Foram desenvolvidas a partir da observação das atividades cerebrais humanas, em que as entidades básicas – neurônios, interconectados em redes permitem a troca de informação entre

eles (Nola, 2015). É um modelo útil para a resolução de problemas de forma semelhante às características principais do cérebro humano do ponto de vista do processamento de informação. Possui a capacidade de adquirir, armazenar e administrar o conhecimento (Nola, 2015), e é definida por Silva *et al.* (2010, p. 24) como sendo “um conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligadas por sinapses artificiais”. Este sistema de interconexões é representado por vetores e matrizes de pesos sinápticos.

Silva *et al.* (2010) e Matsunaga (2012) apresentam como principais características das RNA:

- i) Adaptação por experiência;
- ii) Capacidade de aprendizado;
- iii) Habilidade de generalização;
- iv) Organização dos dados;
- v) Tolerância a falhas (devido ao elevado número de interconexões);
- vi) Armazenamento distribuído (o processo é realizado de forma distribuída entre as diversas sinapses).

Os elementos computacionais ou unidades processadoras, denominadas neurônios artificiais são modelos simplificados dos neurônios biológicos, nos quais a aprendizagem do cérebro humano é alcançada através do processo de mudança química em conexões sinápticas, e acontece a partir de restrições inatas e experiências adquiridas (Kasabov, 1998).

Para Nola (2015) a RNA é uma técnica computacional que imita o sistema nervoso humano, mas que é dinâmica, sem memória e com capacidade de aprendizado baseado em exposição de exemplos. Thomé (2003) ressalta que não há um sequencia ordenada e definida de instruções para as redes, devendo a mesma aprender a tarefa que será executada através de exemplos prévios.

Um modelo de neurônio artificial simples (Figura 2.3) tem a função de coleta entre os sinais de entrada, a fim de agregá-los e produzir uma resposta de acordo com sua função de ativação (Silva *et al.*, 2010, p. 33). Assim, os sinais de entrada ($X_1, X_2...X_n$) são ponderados por seus pesos sinápticos ($W_1, W_2...W_n$) dando a eles uma relevância por meio de multiplicações. A soma ponderada das entradas (\sum) se transforma na saída do corpo celular artificial (u) após a introdução do limiar de ativação (θ). A função de ativação ($f(u)$) possui o objetivo de limitar a saída dentro de um intervalo de valores a serem assumidos.

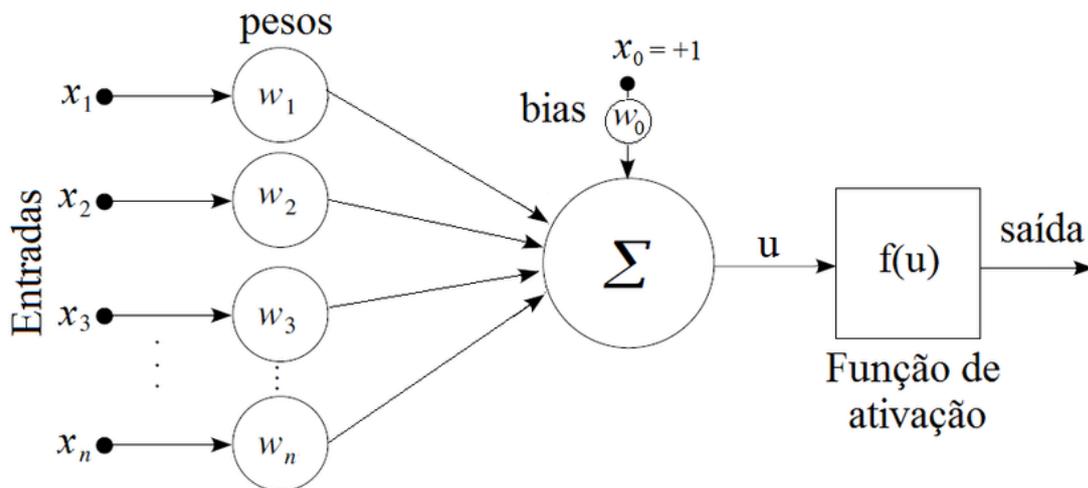


Figura 2.3. Modelo artificial de neurônio biológico. Fonte: Neto & Bonini, 2010.

A tabela 2.1 apresenta a descrição dos elementos que compõem um neurônio artificial.

Sinais de Entrada	Variáveis assumidas para uma aplicação específica. Cada valor de entrada é um tipo de dado associado à amostra. Ex: X1=valor de declividade.
Pesos Sinápticos	Valores associados a cada variável de entrada a fim de ponderá-las, permitindo quantificar as suas relevâncias. Os pesos se iniciam com valores aleatórios entre 0 e 1, e são ajustados durante o processo de cálculo da rede.
Combinador Linear	Tem a função de agrupar os produtos dos sinais de entrada e pesos sinápticos.
Limiar de Ativação	Determina um valor que deve ser atingido pelo combinador linear de forma a produzir um resultado pelo neurônio.
Potencial de Ativação	Diferença de valores entre o combinador linear e o limiar de ativação e determina o comportamento do neurônio (excitatório ou inibitório).
Função de Ativação	Normaliza a saída do neurônio para valores que possam ser interpretados.
Sinal de Saída	Resultado final do processo de cálculo. Pode ser o produto final geral da análise, bem como ser utilizado por outros neurônios da rede.

Tabela 2.1. Elementos de um neurônio artificial. Fonte: Nola, 2015.

Os pesos sinápticos são determinados primeiramente de forma randômica. A partir do início do processo, os pesos sofrem um ajuste de acordo com alguma regra visando o aprimoramento da rede neural na execução da sua tarefa. A representação matemática de um neurônio artificial é definida pelas equações 1 e 2.

$$\mathbf{u} = \sum_{i=1}^n \mathbf{W}_i \times \mathbf{X}_i - \boldsymbol{\theta} \quad \text{Equação 1}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{u}) \quad \text{Equação 2}$$

A arquitetura de uma RNA é a maneira pela qual os neurônios de uma rede neural estão estruturados. Esta estrutura está intimamente ligada ao algoritmo de aprendizagem a ser usado para treinar a rede (Haykin, 2007). Pode-se, em geral, classificar três tipos de arquiteturas de rede fundamentalmente diferentes.

Nas redes unidirecionais de uma única camada todas as conexões entre neurônios diferentes seguem a direção entrada→saída. Neste caso, a rede neural é organizada na forma mais simples: uma camada de entrada que se projeta para uma única camada de saída de camadas (Figura 2.4). Este tipo de rede também pode ser denominada como alimentada adiante ou acíclica ou de camada simples (Haykin, 2007 e Silva *et al.*, 2010).

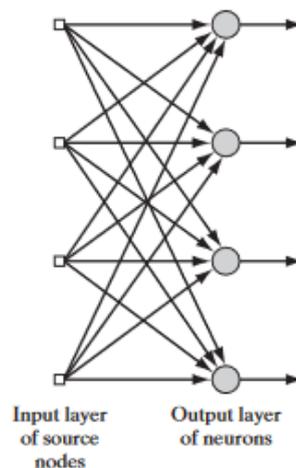


Figura 2.4. Rede unidirecional com única camada de neurônios Fonte: Haykin, 2007.

Existem também as redes unidirecionais, ou seja, que seguem uma única direção, que possuem mais de uma camada de nós computacionais, são as chamadas redes de múltiplas camadas. As camadas são ocultas e a função dos neurônios ocultos é intervir entre a entrada externa e a saída da rede (Haykin, 2007). Os sinais de entrada (Figura 2.5) são inseridos nos

neurônios da segunda camada (primeira camada oculta). Os sinais de saída desta camada são, então, utilizados como dados de entrada para a terceira camada e assim por diante até a finalização do processo.

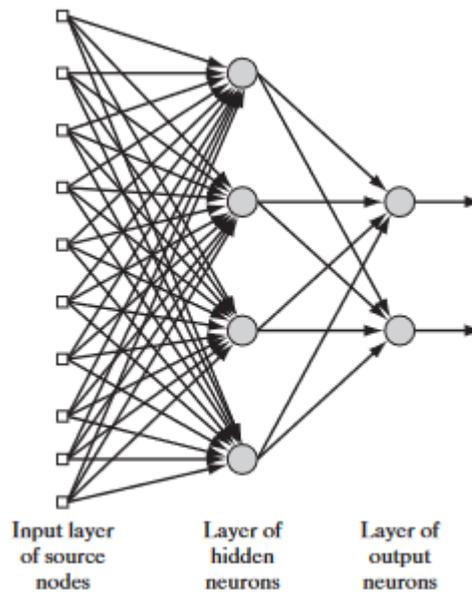


Figura 2.5. Rede unidirecional com multicamadas de neurônios. Fonte: Haykin, 2007.

O tipo de arquitetura chamado de rede recorrente apresenta pelo menos uma realimentação durante o processo de execução da rede. A presença da realimentação tem um impacto significativo na capacidade de aprendizagem da rede e no seu desempenho (Matsunaga, 2012). Uma rede recorrente tem uma camada de entrada de neurônios que alimenta seu sinal de saída de volta para as entradas de todos os outros neurônios, podendo ter ou não auto-realimentação (Haykin, 2007). A Figura 2.6 apresenta uma rede com neurônios ocultos onde as conexões de realimentação se originam destes neurônios bem como dos neurônios de saída, e possuem uma memória de atraso, os operadores de atraso unitário, que operam sobre todas as entradas $x(n)$ produzindo uma versão atrasada (valores passados - p) da seguinte forma: $x(n-1), x(n-2), \dots, x(n-p)$.

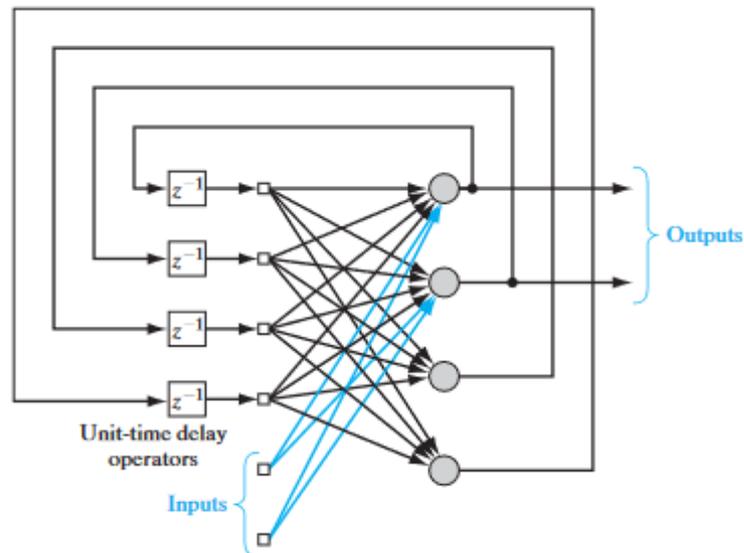


Figura 2.6. Rede recorrente com neurônios ocultos. Fonte: Haykin, 2007.

Entre os vários algoritmos presentes nas RNA, um dos mais utilizados é o de retropropagação (*back-propagation*), criado por Rumelhart, Hinton e Williams (1986), que visa encontrar o erro mínimo durante o treinamento da rede e ajusta os pesos de maneira iterativa (Pradhan & Lee, 2010). De acordo com Oliveira *et al.* (2013) o algoritmo de retropropagação atualiza os pesos de um neurônio através apenas dos erros, das entradas e da derivada do próprio neurônio. Segundo Gómez & Kavzoglu (2005, p. 22):

O algoritmo de retropropagação é popular devido à sua flexibilidade e adaptabilidade na modelagem de uma ampla gama de problemas em diferentes campos de aplicação. O algoritmo opera procurando uma superfície de erro, definida como uma função dos pesos, usando uma técnica de descida de gradiente para localizar o ponto com erro mínimo. Cada iteração no algoritmo de retropropagação possui dois movimentos básicos: orientado para frente e para trás. O ciclo de propagação direta começa com a apresentação de um conjunto de padrões de entrada na rede. A correção de erro inversa começa na camada de saída e o erro é realimentado através da camada intermediária em direção à camada de entrada, a fim de ajustar os pesos e, assim, reduzir o erro. A interação é mantida até que o erro seja reduzido a um nível aceitável em termos de um número predeterminado de vezes, ou um valor limite especificado (GÓMEZ; KAVZOGLU, 2005, p. 22).

Pradhan & Lee (2010) utilizaram RNA em três diferentes redes (11x24x2; 7x16x2; 4x10x2; Entrada x Neurônio Escondido x Saída, respectivamente) e os compararam com regressão logística e razão de frequência obtendo resultados de 90% e 93% na regressão e na

razão de frequência, respectivamente, e resultado de 94% na RNA com o algoritmo retropropagativo utilizando sete variáveis de entrada. Em estudo de Gómez & Kavzoglu (2005), foi utilizado o algoritmo retropropagativo em uma rede 9x28x1 com taxa de aprendizado de 0,2 e pesos da rede inicializados aleatoriamente em um intervalo de -0,25 até 0,25, acarretando em resultados de 87% de precisão para áreas com deslizamentos e 95% de precisão em áreas sem movimentos. Em Lee *et al.* (2007), com uma rede 14x30x2 e taxa de aprendizado de 0,01, os resultados do algoritmo também se mostraram satisfatórios e obtiveram alta performance em 65-90% das classes em estudo.

Para análise de suscetibilidade a deslizamentos, as RNA também vêm se mostrando eficientes nos seus mais diversos tipos. Chen *et al.* (2017c) utilizaram um conjunto de RNA para esse tipo de mapeamento na Índia e encontrou resultados de AUC de 0,82, enquanto que Bragagnolo *et al.* (2020), utilizando também um conjunto de RNA, obteve resultados para Porto Alegre e Rio de Janeiro semelhantes e até melhores do que os mapas de suscetibilidades divulgados pelos respectivos órgãos estaduais. Esses resultados mostram que as RNA, mesmo quando em conjunto com outros algoritmos, apresentam ainda elevada capacidade de mapeamento a deslizamento.

Yilmaz (2009) comparou a razão de frequência e regressão logística com uma rede neural de 8x17x1, com 2250 amostras de treinamento e obteve valores de AUC na faixa de 0,85 para as RNA, enquanto as outras duas técnicas ficaram abaixo de 0,84. Em estudos de suscetibilidade na Itália e utilizando uma rede 10x7x1, Conforti *et al.* (2014) conseguiu acurácias de até 0,93 através da AUC.

Oliveira *et al.* (2019) testou duas configurações de rede para a mesma área e encontrou bons resultados em ambas. Na primeira, com uma rede 9x11x1, considerada uma rede mais complexa, obteve acurácias de 0,93, enquanto que na segunda configuração, com 6x4x1, uma rede menos complexa, porém mais rápida e leve de ser rodada pelo modelo, obteve acurácias de 0,90. Isso demonstra que mesmo para modelos menos complexos de redes, com menores quantidades de entradas e de neurônios, conseguem obter eficientes resultados.

Bui *et al.* (2016) comparou as RNA com MVS, regressão logística e modelos logísticos de árvores. Apesar de todos os resultados terem sido satisfatórios, apenas as RNA obtiveram AUC maiores do que 0,90, e os autores ainda afirmam que o poder de previsão das RNA tem diferenças significantes do ponto de vista estatístico, evidenciando a sua ótima capacidade para estudos de suscetibilidade a deslizamentos.

Nefeslioglu *et al.* (2009) comparou o uso de RNA com regressão logística utilizando diferentes amostragens. Em todos os resultados as RNA obtiveram AUC maiores do que 0,91,

chegando a até 0,95 no melhor dos casos, enquanto a regressão obteve valores menores do que 0,86. Entretanto, os autores ressaltam que a capacidade de generalização das RNA sofre intensas mudanças dependendo do processo de amostragem e da distribuição dos valores dos atributos morfométricos utilizados.

Todos esses casos exemplificam a capacidade e a potencialidade das RNA para mapeamentos da suscetibilidade a deslizamentos, independente da estrutura da rede, dos parâmetros utilizados ou dos processos de amostragem em si.

CAPÍTULO 3: MATERIAIS E MÉTODOS

3.1 METODOLOGIA

A metodologia utilizada foi dividida em 4 partes, as quais são ilustradas na figura 3.1 e descritas abaixo.

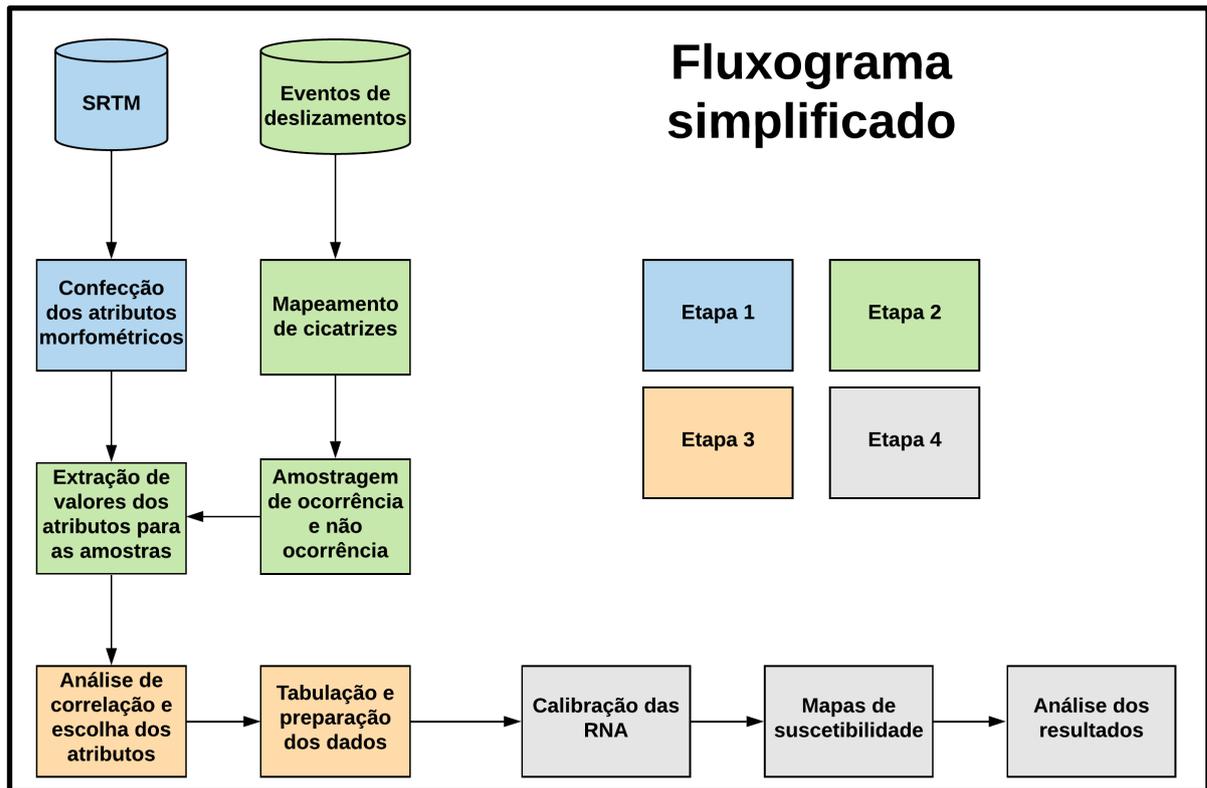


Figura 3.1. Fluxograma das etapas do trabalho.

Na etapa 1, utilizou-se um mosaico do MDE de 3 segundos de arco, com 90 m de resolução, obtido através da missão *Shuttle Radar Topography Mission* (SRTM) e que abrangesse toda a extensão da área de estudo. Esse MDE é disponibilizado pelo *United States Geological Survey* (USGS) e seus dados altimétricos foram processados por interferometria de radar de abertura sintética com base na banda C. Entre tantas versões disponíveis para uso, este estudo utilizou a versão *void-filled*, a qual é uma versão com preenchimentos de erros e falha que ocorrem na hora da obtenção dos dados. A escolha desse MDE se deve principalmente a 2 fatores:

- i) O modelo SRTM de 3 segundos de arco foi um dos primeiros a ser lançado e já possui confiabilidade comprovada, sendo utilizado em diversos estudos para diversas áreas e modalidades de trabalho;

- ii) A região em estudo possui grande extensão territorial e o uso de MDEs de alta resolução acabaria por demandar um alto custo computacional e maior tempo nos processos de modelagem, tornando a opção de 90 m adequada para a situação.

Foi realizado um mosaico das imagens desse MDE para que abrangesse toda a área de estudo e após isso, foram confeccionados os atributos morfométricos de terreno, através do *software* SAGA-GIS e da ferramenta *Basic Terrain Analysis*.

Para a etapa 2, foram selecionados 5 eventos de precipitação extrema ocorridos na Serra Geral no Sul do Brasil e que resultaram em deslizamentos.

- i) O evento mais antigo foi também o mais marcante e ocorreu no município de Timbé do Sul – SC, no final do ano de 1995, deixando centenas de desabrigados e dezenas de mortos. Registros pluviométricos marcaram 176 mm em 24 h e a área atingida foi superior a 3.500 hectares (Pellerin *et al.*, 1997). Nesta região, foram encontradas 469 cicatrizes de deslizamentos com área média de 0,56 ha.
- ii) Entre o final do ano 2000 e o meio de 2001, várias áreas próximas ao Arroio Forromeco foram atingidas por chuvas intensas, entre eles podemos citar os municípios de Alto Feliz e São Vendelino, nos quais foram registradas precipitação de 148 mm em um período de 2 h (Azambuja *et al.*, 2001). Foram encontradas 53 cicatrizes na área com área média de 0,23 ha;
- iii) No início de 2007, o município de Maquiné foi atingido por um período de chuvas intensas, o qual acarretou em deslizamentos e enxurradas no município, sendo declarado estado de emergência. O número de cicatrizes encontradas foi de 229 e a área média delas foi de 0,21 ha;
- iv) No vale do Rio Fão foi registrado um evento hidrometeorológico marcante nas datas de 03 e 04 de Janeiro de 2010, influenciado por jatos de baixos níveis e a formação de um corredor de umidade no qual resultado em uma precipitação superior a 250 mm em 24 h (Oliveira *et al.*, 2017) e o registro de 121 cicatrizes de deslizamento com área média de 0,22 ha;
- v) Em janeiro de 2017 ocorreu um evento hidrometeorológico no município de Rolante - RS que culminou em deslizamentos e inundações no município. Os pluviômetros locais registraram alturas pluviométricas de até 180 mm em poucas horas, e foram mapeadas 338 cicatrizes com área média de 0,34 ha.

Após a identificação e escolha desses 5 locais, foi realizado o mapeamento das cicatrizes através das imagens disponibilizadas pelo Google Earth Pro. Essas imagens foram as escolhidas

devido a multitemporalidade e a alta resolução das mesmas. A tabela 3.1 mostra as datas das imagens que foram utilizadas para o mapeamento.

	Data da imagem
Timbé do Sul	04/03/2005
Forromeco	06/04/2004
Maquiné	01/06/2010 e 06/08/2017
Fão	17/01/2012
Rolante	06/08/2017

Tabela 3.1. Datas das imagens utilizadas para mapeamento de cicatrizes.

As cicatrizes foram delimitadas através de polígonos e então foram exportadas para ambiente SIG para que pudesse ser realizado o processo de amostragem.

Dentro de um SIG, os polígonos das cicatrizes foram transformados em pontos, sendo 1 ponto por cada pixel, utilizando pixel de 90 m para padronizar com os atributos morfométricos extraídos. Cada ponto gerado foi considerado como uma amostra de ocorrência. Na sequência, foram gerados *buffers* de distância das cicatrizes (2 – 40 km) a fim de gerar as amostras de não ocorrência. Para cada *buffer* foi gerado o mesmo número de amostras de ocorrência, respeitando assim a proporção 1:1 entre amostras de ocorrência e não ocorrência.

De posse das amostras e dos atributos, os valores numéricos desses atributos foram extraídos para cada uma das amostras, obtendo valores para ocorrência e não ocorrência, os quais foram exportados para uma tabela de Excel.

Na terceira etapa foi feita uma análise de correlação nessa tabela a fim de estabelecer quais atributos teriam melhor correlação com as variáveis dependentes (ocorrência e não ocorrência), estabelecendo o limiar de – 0,08 até 0,08 para a exclusão desses atributos do processo de modelagem. Após a definição e exclusão desses atributos, o restante foi organizado de forma a ser lido pelo *script* da RNA dentro do programa MATLAB.

Já na etapa 4, a RNA foi treinada utilizando 50% das amostras para tal e o restante para validação cruzada (25%) e teste (25%). O resultado final foram 180 mapas de suscetibilidade, cada um utilizando amostras de distintas áreas e distintos *buffers* de distância. Esses mapas foram analisados de forma visual, destacando suas características e distinções e também através da análise da acurácia feito pela Curva ROC.

CAPÍTULO 4: REDES NEURAS ARTIFICIAIS APLICADAS À SUSCETIBILIDADE A DESLIZAMENTOS: O EFEITO DAS ÁREAS AMOSTRAIS NA CAPACIDADE DE GENERALIZAÇÃO E EXTRAPOLAÇÃO DO MODELO

RESUMO

As redes neurais artificiais (RNA) têm sido utilizadas para a identificação de áreas suscetíveis a deslizamentos, sendo um dos métodos mais precisos para essa finalidade. Diversos fatores podem interferir no desempenho dos modelos e de seus mapas resultantes, com destaque para a amostragem. Neste artigo, foi avaliada a influência das áreas amostrais na modelagem de suscetibilidade a deslizamentos e na capacidade de generalização e extrapolação dos dados. A partir de um inventário de cicatrizes, distribuídos em cinco áreas do Sul do Brasil, foram definidas amostras de não-ocorrência por meio de diferentes *buffers* (2-40 km) em relação aos deslizamentos, fazendo com que as amostras ocupem distintas unidades de relevo. Como variáveis de entrada do modelo, utilizou-se 16 atributos morfométricos do terreno, extraídos de um modelo digital de elevação. O treinamento da rede de múltiplas camadas foi realizado pelo algoritmo retropropagativo e a acurácia foi calculada por meio da *Area Under the Receiver Operating Characteristic Curve* (AUC). A acurácia dos modelos ficou entre 0,73 e 0,93. Essa variação foi explicada principalmente pelo *buffer* utilizado. O mapa de suscetibilidade resultante do modelo de maior acurácia foi obtido com um *buffer* de 40 km para coleta de amostras de não-ocorrência. A maior distância entre as amostras de ocorrência e de não-ocorrência facilita a modelagem, uma vez que amplia as diferenças morfométricas entre os grupos amostrais. Destaca-se também que mesmo utilizando amostras de treinamento em apenas uma das cinco áreas, foi possível extrapolar o mapa de suscetibilidade para toda a região do inventário de cicatrizes.

Palavras-chave: Amostragem; aprendizado de máquina; desastres naturais; atributos preditivos

ABSTRACT

Artificial neural networks (ANNs) have been used to identify areas susceptible to landslides and constitute one of the most precise methods for this purpose. Several factors can interfere in the performance of the models and their resulting maps (especially sampling). This article evaluated the influence of sampling areas on landslide susceptibility modelling and the capacity for generalization and extrapolation of data. Based on an inventory of scars, distributed in five areas of southern Brazil, non-occurrence samples were defined by means of different buffers (2-40 km) in relation to the landslides, making the samples occupy distinct relief units. A total of 16 morphometric attributes of the terrain (extracted from a digital elevation model) were used as input variables of the model. Multilayered

network training was carried out using a backpropagation algorithm and accuracy was calculated by means of the Area Under the Receiver Operating Characteristic Curve (ROC). Model accuracy was between 0,73 and 0,93. This variation was explained mainly by the buffer used. The susceptibility map resulting from the model of greater accuracy was obtained with a 40-km buffer in order to collect non-occurrence samples. The great distance between the occurrence and non-occurrence samples facilitates the modelling, since it increases the morphometric differences between the sampling groups. The study also highlights that even using training samples in only one of the five areas, a susceptibility map of the entire region was extrapolated from the inventory of scars.

Keywords: Sampling; Machine learning; Natural hazards; Predictive attributes

4.1 INTRODUÇÃO

Os desastres naturais são eventos naturais inesperados e/ou incontrolláveis de magnitude incomum e que podem ameaçar pessoas (Bokwa, 2013). Esses desastres vêm recebendo crescente atenção de entidades públicas e instituições de pesquisa, devido principalmente, ao aumento de suas ocorrências e dos danos e prejuízos causados (Tominaga *et al.*, 2009). Os principais desastres naturais no Brasil são as inundações e os deslizamentos. De acordo com o Relatório *The Human Cost of Weather Related Disasters* do Escritório das Nações Unidas para a Redução de Desastres e o Centro de Pesquisas de Epidemiologia em Desastres (UNISDR/CRED, 2015), entre 1980 e 2010 ocorreram 146 desastres naturais e 4.948 mortes, o que coloca o Brasil em 13º quanto aos riscos de inundações e 14º quanto a deslizamentos.

Comumente, os deslizamentos ocorrem em paisagens de relevos íngremes e acabam por causar danos e prejuízos sociais, ambientais e econômicos (Shahabi & Hashim, 2015; Huang & Zhao, 2018; Steger & Kofler, 2019) em áreas de ocupação humana. A fim de diminuir os danos e prejuízos causados por esses desastres e realizar uma gestão mais efetiva e focada na redução dos riscos, no ordenamento territorial e no uso do solo, pesquisadores tem utilizado mapeamentos de suscetibilidade, os quais tem se tornado um dos principais meios para inventariar, mapear e classificar áreas suscetíveis à ocorrência desses eventos (Catani *et al.*, 2013; Goetz *et al.*, 2015; Moosavi & Niazi, 2016; Gordo *et al.*, 2019; Nguyen *et al.*, 2019).

O uso de sistemas de informação geográficas (SIG) e sensoriamento remoto (SR), juntamente com a aplicação de modelos empíricos, estão entre as formas mais utilizadas e aceitas no mapeamento de áreas suscetíveis a deslizamentos. Entre os modelos mais utilizados, destacam-se aqueles baseados em aprendizagem de máquina (AM) como as redes neurais artificiais (RNA) (Kalantar *et al.*, 2018; Zhu *et al.*, 2018; Wang *et al.*, 2019), que costumam apresentar resultados coerentes com a distribuição dos eventos ocorridos. Estes modelos extraem conhecimentos a partir de amostras de cicatrizes e passam por uma fase de treinamento. As variáveis explicativas devem ser bem selecionadas a fim de gerar um treinamento adequado e utilizar toda a sua capacidade de aprendizagem.

O processo de amostragem é talvez a etapa mais importante na obtenção do resultado final. Comumente se usa a estratégia de amostragem 1:1 (Heckmann *et al.*, 2014), definindo o mesmo número de amostras de ocorrência e de não ocorrência. Apesar disso, distintas técnicas são utilizadas para amostragens de ocorrência e não ocorrência. Nas amostras de ocorrência, as técnicas mais utilizadas são as que utilizam apenas o centróide das cicatrizes (Van Den Eeckhaut *et al.*, 2006; Xu *et al.*, 2014; Chen *et al.*, 2018), as que utilizam o local de ruptura

(Steger *et al.*, 2016; Hong *et al.*, 2019) e as que utilizam todos os pixels presentes nas cicatrizes (Petschko *et al.*, 2014; Quevedo, 2019). As amostras de não ocorrência também são de extrema importância. De acordo com Zhu *et al.* (2018), a boa escolha das mesmas pode suprimir a superestimação de suscetibilidade, permitindo maior capacidade de generalização dos modelos. Xiao *et al.* (2010) afirmam que há 2 métodos comumente utilizados para gerar essas amostras, um deles é a aleatoriedade e abrange toda a área de estudo, excluindo apenas regiões consideradas como de ocorrência (Chen *et al.*, 2017; He *et al.*, 2019), o outro é o método *buffer controlled sampling* (BCS), usado geralmente para delimitar uma distância mínima que a amostragem de não ocorrência deve ter em relação as amostras de ocorrência, como demonstrado em Yao *et al.* (2008) e Xie *et al.* (2018), que estipularam uma distância mínima de 40 m e 300 m, respectivamente, a partir das amostras de ocorrências. Destaca-se, no entanto, que faltam estudos relativos à influência das distâncias máximas no processo de modelagem a deslizamentos.

Finalmente, a partir do inventário completo, é necessário definir a quantidade de amostras utilizadas no treinamento e validação. Este é um dos maiores problemas encontrados na confecção de bons mapas de suscetibilidade (Ada & San, 2018). A grande maioria dos trabalhos pesquisados utiliza a proporção de 70% das amostras para treinamento e 30% para validação (Bui *et al.*, 2016; Chen *et al.*, 2017; Pham & Prakash, 2017; Rahmati *et al.*, 2017; He *et al.*, 2019; Wang *et al.*, 2019). Alguns autores utilizam 50% / 50% (Pradhan, 2013; Zêzere *et al.* 2017; Pham *et al.*, 2019), e outros 80% / 20% (Shahabi & Hashim, 2015; Shirzadi *et al.*, 2018), mostrando assim a grande variabilidade estatística que pode ser usada, com influência direta no resultado final dos modelos e em sua capacidade de generalização.

A partir dessas considerações, o presente estudo tem como objetivo principal analisar a influência das áreas amostrais na capacidade de generalização e extrapolação de modelos de RNA para mapear áreas suscetíveis a deslizamentos.

4.2 METODOLOGIA

4.2.1 Área de estudo

A área de estudo (Figura 4.1) localiza-se nas bordas da unidade geomorfológica do Planalto Meridional, ao sul do Estado de Santa Catarina e na região central e nordeste do Estado

do Rio Grande do Sul. Essa região corresponde à denominada Serra Geral, caracterizada por altitudes de até 1.400 m e acentuadas declividades.

Essa área constitui uma das maiores províncias de basaltos continentais do mundo (Neto, 2006). Dentre as rochas vulcânicas dominantes, destacam-se basaltos, andesitos e andesitos basálticos (95%) seguidas de riolitos e riolitos (5%) (Hartmann, 2014). Destaca-se ainda a presença de arenitos intercalados da Fm. Botucatu (Roisenberg & Viero, 2000; Machado, 2005). É descrita geomorfologicamente como a borda do Planalto Meridional, originada pela erosão diferencial e pela tectônica. Possui lineamentos estruturais em distintas direções, a macroforma de seu relevo é do tipo falhado e os solos possuem espessura média de 50 centímetros (Betiollo, 2006).

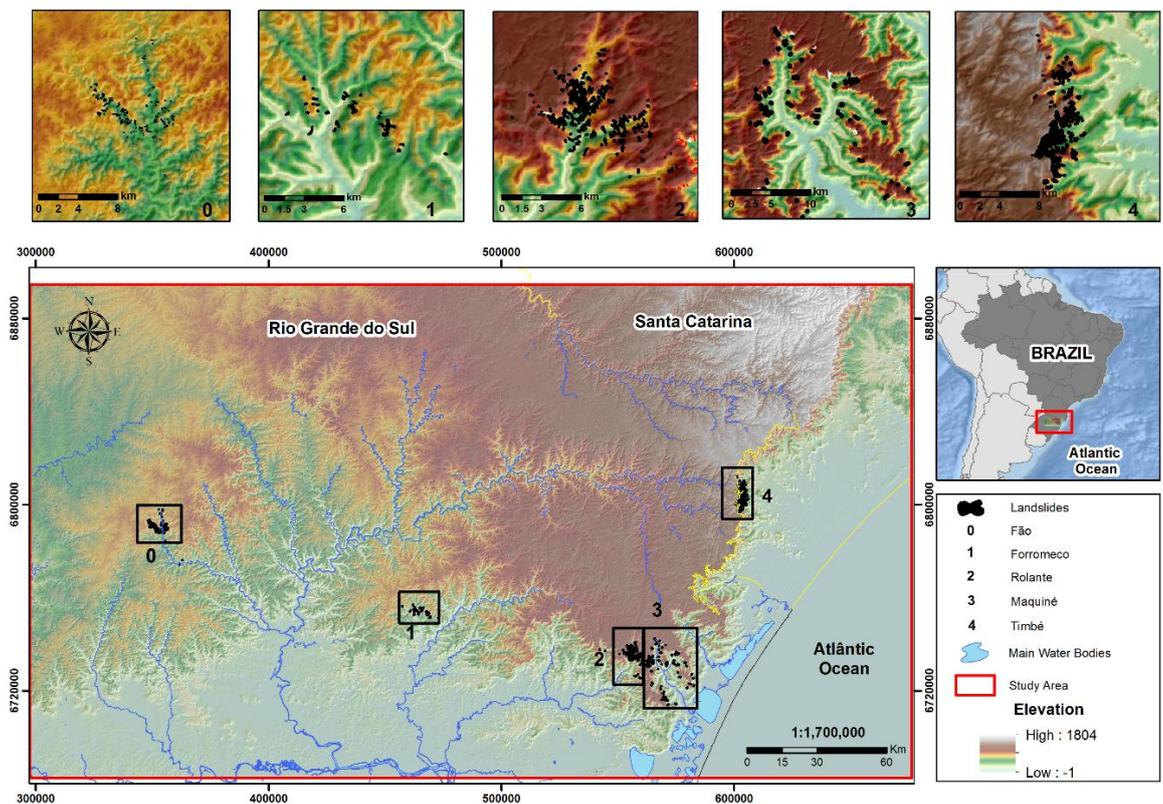


Figura 4.1. Localização da área de estudo e das amostras de cicatrizes.

O clima da região varia em função de fatores como altitude e relevo e de acordo com Rossato (2011), se enquadra em dois tipos distintos: um subtropical úmido com variação longitudinal das temperaturas médias na região centro norte e outro subtropical muito úmido com inverno frio e verão fresco na região a nordeste, juntamente na divisa de Estados entre Rio Grande do Sul e Santa Catarina. As precipitações entre os meses de maio e outubro estão

associadas preferencialmente à passagem de sistemas frontais, formados pelo ingresso de massas de ar frias, sendo mais volumosas e homogêneas sobre o Sul do Brasil, normalmente ocasionando inundações. Já no verão, costumam ocorrer precipitações localizadas, intensas e de curta duração, devido ao deslocamento de sistemas convectivos formados na região tropical, normalmente sendo responsáveis por enxurradas em pequenas bacias e deslizamentos nas áreas mais declivosas.

4.2.2 Preparação dos dados

A figura 4.2 mostra as etapas metodológicas adotadas no estudo, incluindo o inventário de cicatrizes, a elaboração dos *buffers* e a amostragem, a extração de parâmetros morfométricos, a modelagem matemática com RNA e a validação através da Area Under the Receiver Operating Characteristic Curve (ROC).

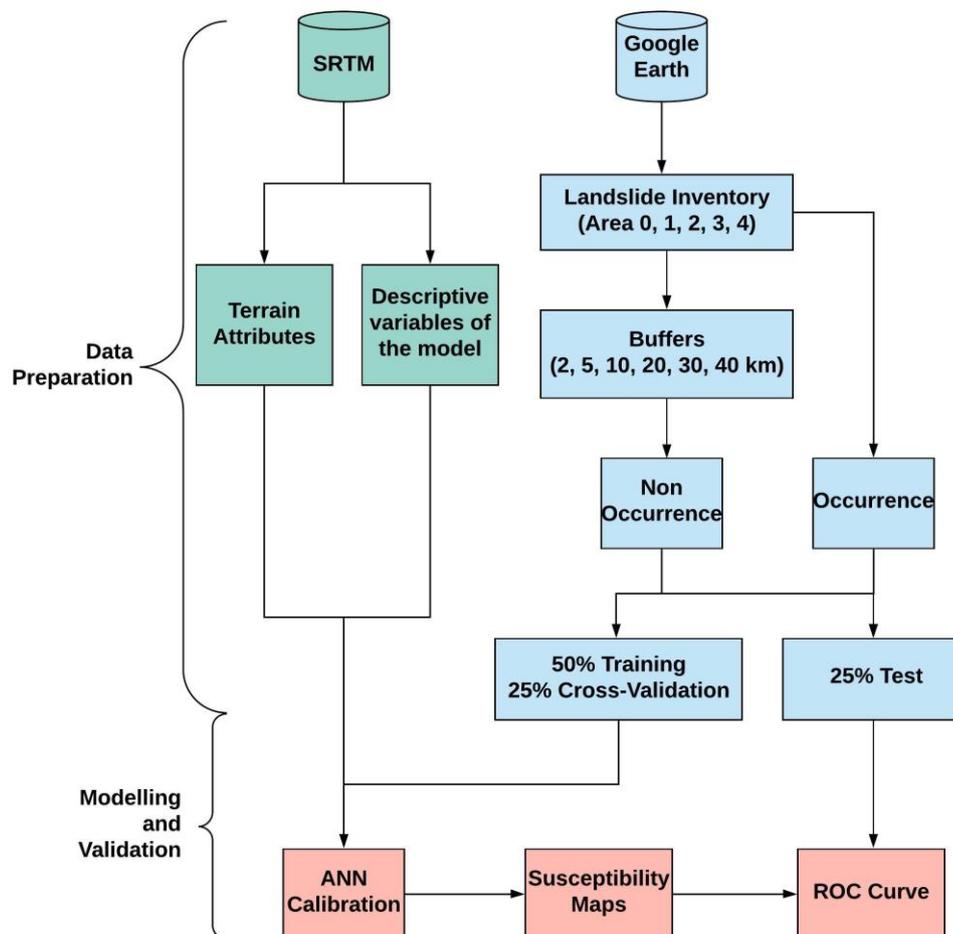


Figura 4.2. Fluxograma da metodologia utilizada.

4.2.3 Inventário de cicatrizes

Foi realizado o inventário de cicatrizes em 5 diferentes áreas da Serra Geral, nas quais havia o registro de um grande número de cicatrizes de deslizamentos causadas por eventos extremos de precipitação entre os anos de 1995 e 2017, sendo elas:

- o vale do rio Fão, na Bacia Hidrográfica do Taquari-Antas, localizado nos municípios de Fontoura Xavier, Barros Cassal e Soledade, nomeado como área 0;
- o vale do arroio Forromeco, na Bacia Hidrográfica do Caí, localizado nos municípios de São Vendelino, Alto Feliz e Carlos Barbosa, nomeado como área 1;
- o vale do rio Mascarada, na Bacia Hidrográfica do Sinos, localizado nos municípios de Rolante, Riozinho e São Francisco de Paula, nomeado como área 2;
- o vale do rio Maquiné, na bacia hidrográfica e município de Maquiné, nomeado como área 3;
- o vale do rio da Rocinha, na Bacia Hidrográfica do Araranguá, no município de Timbé do Sul e Jacinto Machado, nomeado como área 4.

Para realizar o inventário, foram utilizadas imagens multitemporais do Google Earth Pro, a fim de obter o maior número de cicatrizes na região. As cicatrizes foram delimitadas a partir de polígonos e transferidas para ambiente SIG, no qual foram convertidas para pontos, sendo um ponto por cada pixel presente dentro das cicatrizes. Foram utilizados os pixels do *Shuttle Radar Topography Mission (SRTM)* com resolução espacial de 90 m, mesma resolução adotada para posterior confecção dos atributos de terreno.

De acordo com Bui *et al.* (2016) o mapeamento de suscetibilidade a deslizamentos é um caso clássico de classificação binária, no qual a ocorrência de um deslizamento é resultante do valor 1 e a não ocorrência de um valor 0. Seguindo esta lógica, cada ponto gerado foi definido como uma amostra de ocorrência, num total de 1624 pontos de ocorrência. Para as amostras de não ocorrência, gerou-se o mesmo montante de pontos, distribuídos de forma aleatória no terreno (com distância mínima de 150 m entre si para garantir que um pixel do MDE possua no máximo uma amostra) e com limitações máximas de distância, atribuídas através da criação de *buffers* com distâncias de 2 km, 5 km, 10 km, 20 km, 30 km e 40 km das amostras de ocorrência, conforme demonstrado na figura 4.3.

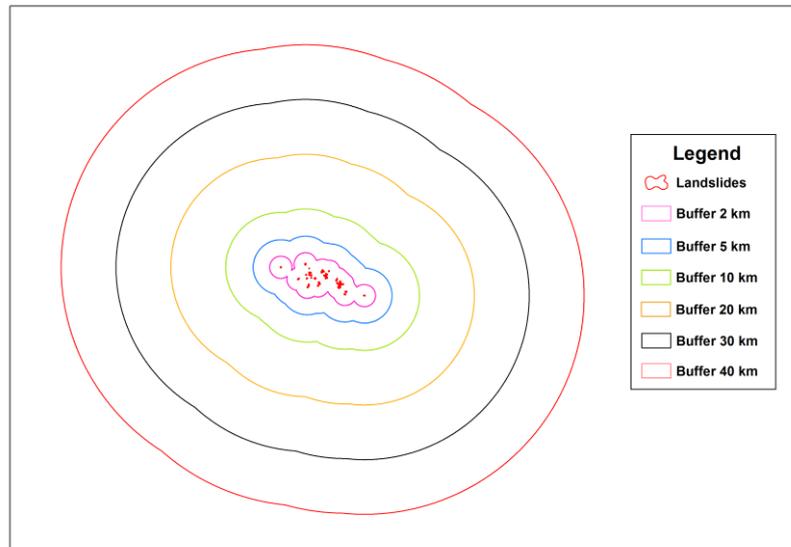


Figura 4.3. Exemplo de *buffers* utilizados para amostragem de não ocorrência.

4.2.4 Fatores condicionantes de deslizamentos

Foram utilizadas como entrada no modelo 16 atributos morfométricas do terreno (Figura 4.4), elaboradas no *software* SAGA-GIS 7.0 através da ferramenta *Basic Terrain Analysis*, a partir do MDE SRTM com 90 m de resolução (3 *arc-seconds*), sendo elas:

Elevation (ELV) representa as altitudes presentes na região e é um parâmetro amplamente utilizado em processos de modelagem. Sua variação tem efeito notável em modelos de suscetibilidade dependendo da região (Nourani *et al.*, 2014; Hong *et al.*, 2016; Al-Najjar *et al.*, 2019).

Slope (SLP) é o ângulo de inclinação da encosta e, assim como a elevação, é constantemente utilizado para modelagens de suscetibilidade a deslizamentos, sendo muitas vezes considerado o fator de maior relevância para a modelagem (Lee & Min, 2001; Kavzoglu *et al.*, 2014; Nourani *et al.*, 2014).

Analytical Hillshading (AHS) calcula o ângulo em que a luz vinda da posição da fonte de luz atingiria a superfície (Tarini *et al.*, 2006).

Aspect (ASP) é a orientação ou direção da inclinação, com significativa importância nos processos de modelagem (Galli *et al.*, 2008).

Channel Network Base Level (CNBL) considera-se a distância vertical até o nível do canal de base da rede local. Para Bock & Kothe (2008), o CNBL é essencial para a previsão da profundidade de solos hidromórficos influenciados pelo escoamento de águas subsuperficiais e subterrâneas.

Vertical Distance to Channel Network (VDCN) distância vertical até a rede de drenagem, expressa em metros (Möller *et al.*, 2012). Valores mais baixos indicam maior proximidade com a rede de drenagem e, conseqüentemente, menor suscetibilidade a deslizamentos (Samia *et al.*, 2017).

Longitudinal Curvature (LOC) refere-se à taxa de variação de fluxo em áreas declivosas, afetando a aceleração e desaceleração do fluxo (Zevenbergen & Thorne, 1987; Chen *et al.*, 2017).

Cross-section curvature (CSC) A curvatura da superfície transversal à direção da inclinação, influenciando velocidade de fluxo na superfície (Zevenbergen & Thorne, 1987).

Flow Accumulation (FAC) área de drenagem a montante de cada célula, possui relação direta com a saturação do solo (Fernandes *et al.*, 2004).

Topographic Wetness Index (TWI) estima a quantidade de umidade presente no solo (Al-Najjar *et al.*, 2019). Influencia diretamente nos processos de instabilidade de encosta devido a pressão na água nos poros e diminuição da resistência dos solos (Kavzoglu *et al.*, 2014).

Topographic Position Index (TPI) compara a elevação de cada célula em um DEM para a elevação média de uma vizinhança especificada em torno dessa célula. (Guisan *et al.*, 1999). TPI positivos são locais maiores do que média de seu ambiente, enquanto TPI negativos representam locais inferiores aos seus arredores. Valores próximos a zero são áreas planas ou com grandes inclinações (Weiss, 2001).

Convergence Index (CIN) calcula um índice de convergência/divergência em relação ao escoamento superficial, os valores negativos correspondem a convergente, os positivos correspondem às condições de fluxo divergentes (Koethe & Lehmeier, 1993).

Relative Slope Position (RSP) representando a posição relativa da inclinação nas células entre o fundo do vale e o topo (Samia *et al.*, 2019).

Mass Balance Index (MIB) expressa o saldo entre deposição ou liberação de massa nas encostas (Zanin *et al.*, 2017). Valores negativos de MBI desempenham áreas de deposição líquida, como depressões e várzeas; valores positivos de MBI representam áreas de erosão líquida, como encostas, e valores de MBI próximos de zero indicam áreas onde existe um equilíbrio entre erosão e deposição, como declives lentos e áreas planas (Möller *et al.*, 2008);

Valley Depth (VDP) representa a distância vertical a uma rede de canais de nível de base sendo realizado pela interpolação entre a elevação desse nível de base e a subtração desse nível de base das elevações originais (Oliveira *et al.*, 2019a).

LS Factor (LSF) representa o potencial de perda do solo pelas propriedades combinadas comprimento e direção da inclinação (Hong *et al.*, 2016). Pode ser um fator de alta importância

para suscetibilidade a deslizamentos e os valores mais altos representam terrenos com maior declividade e maior comprimento de rampa (Samia *et al.*, 2017; Oliveira *et al.*, 2019a).

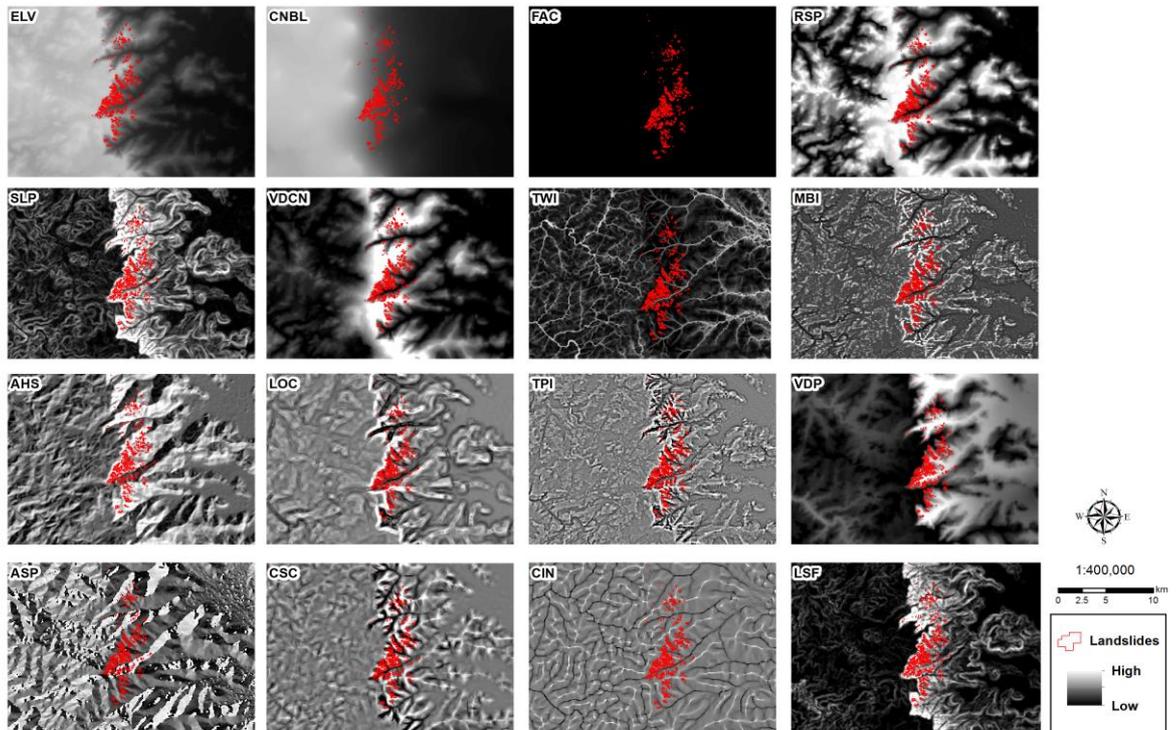


Figura 4.4. Atributos morfométricos de terreno utilizados como entrada na modelagem.

Foi realizada uma análise de correlação linear entre as variáveis independentes do modelo (atributos do terreno) e a dependente (ocorrência ou não ocorrência de deslizamentos), a fim de verificar preliminarmente quais atributos possuem maior influência no processo e quais são redundantes, e que se repetem sistematicamente entre duas ou mais variáveis de entrada. Elaborou-se ainda estatísticas básicas (mínima, média e máxima) de todas as amostras com relação aos *buffers*, a fim de evidenciar a mudança nos valores dos parâmetros conforme a área de amostragem se distanciava das cicatrizes.

4.3 MODELAGEM E VALIDAÇÃO

RNA é um dos métodos mais comuns utilizados na modelagem de deslizamentos (Kumar *et al.*, 2017). Vários estudos mostram sua utilidade e eficiência no mapeamento de suscetibilidade a deslizamentos (Lee *et al.*, 2004; Pradhan & Lee, 2010; Losasso & Sdao, 2018, Kalantar *et al.*, 2018, Oliveira *et al.*, 2019a).

No processo de modelagem foi utilizado um script de RNA, no software MATLAB, sendo o treinamento realizado pelo método retropropagativo com múltiplas camadas (Rumelhart; Hinton; Willians, 1986). Foi utilizada ainda uma série de validação cruzada paralela à série de treinamento para evitar o superajustamento do modelo e garantir sua capacidade de generalização. A proporção de amostras foi estabelecida em 50% para treinamento, 25% para validação cruzada e 25% para teste.

As amostras de treinamento e de validação cruzada, usadas em cada rodada da modelagem, foram separadas para cada uma das cinco áreas, isto é, o treinamento das RNA considerou individualmente cada área amostral. Essa restrição no treinamento foi definida para que fosse possível analisar a capacidade de extrapolação dos modelos para toda a Serra Geral, Sul do Brasil. Apenas as amostras de teste foram consideradas na totalidade da área de estudo.

Para a configuração inicial, o número de neurônios na camada oculta da rede seguiu a fórmula $n-2$, em que n é o número de variáveis de entrada utilizadas na modelagem. Foi testado o aumento do número de neurônios de forma automatizada até um modelo com $n+6$ neurônios na camada oculta, de modo a avaliar o aumento de desempenho dos modelos em função da maior complexidade do modelo e seus pesos sinápticos. Estabeleceu-se como número máximo de ciclos de aprendizagem de 10.000 por iteração, a partir de testes iniciais. Devido à inicialização de pesos aleatória, executou-se 5 repetições em cada configuração do modelo.

Para validação, o índice de acurácia foi calculado por meio da curva ROC, análise baseada na distribuição final de um método de classificação que diferencia previsões corretas e falhas, de acordo com uma tabela de contingência 2×2 (Cantarino *et al.*, 2019), considerada como a representação bidimensional de desempenho de um modelo.

Por fim, para analisar a influência das variáveis de entrada na saída dos modelos, foi utilizado o algoritmo proposto por Oliveira *et al.* (2015), que calcula a importância relativa dos atributos de entrada por meio de pequenas perturbações nos valores das variáveis e medição da sensibilidade destas na saída do modelo.

4.4 RESULTADOS

4.4.1 Análise de correlação e importância das variáveis

A análise de correlação linear dos atributos morfométricos (Tabela 4.1) serviu como pré-seleção para a modelagem com RNA. Atributos com correlação muito baixa com a variável

dependente (ocorrência ou não ocorrência de deslizamentos) foram descartados do processo de modelagem, resultando em maior agilidade no processo de treinamento da rede. Após a eliminação das variáveis menos correlacionadas, realizou-se ainda a verificação dos valores mínimos, médios e máximos (Tabela 4.2) das variáveis de entrada para todas as amostras, separadas por *buffers* de não ocorrência.

	SLP	AHS	ASP	CNBL	CSC	LOC	LSF	TPI	TWI	VDP	VDCN	ELV	FAC	CIN	RSP	MBI
2km	0.49	0.19	-0.04	-0.16	-0.13	-0.22	0.49	-0.17	-0.14	0.25	-0.09	-0.18	-0.02	-0.10	-0.24	-0.04
5km	0.60	0.24	-0.06	-0.20	-0.15	-0.23	0.60	-0.18	-0.22	0.30	0.09	-0.14	-0.03	-0.10	-0.17	0.01
10km	0.64	0.28	-0.07	-0.17	-0.15	-0.23	0.64	-0.17	-0.28	0.32	0.19	-0.05	-0.04	-0.08	-0.08	0.06
20km	0.69	0.32	-0.08	-0.15	-0.18	-0.26	0.70	-0.21	-0.30	0.38	0.27	-0.01	-0.04	-0.11	-0.03	0.04
30km	0.73	0.33	-0.06	-0.10	-0.18	-0.24	0.73	-0.19	-0.36	0.36	0.38	0.07	-0.05	-0.11	0.06	0.10
40km	0.75	0.37	-0.07	-0.05	-0.17	-0.24	0.74	-0.19	-0.39	0.36	0.42	0.14	-0.03	-0.10	0.11	0.11
Média	0.65	0.25	-0.06	-0.14	-0.16	-0.23	0.65	-0.19	-0.28	0.33	0.21	-0.03	-0.04	-0.10	-0.06	0.05

Tabela 4.1. Correlação linear entre atributos morfométricos e a variável dependente do modelo.

	SLP	AHS	CNBL	CSC	LOC	LSF	TPI	TWI	VDP	VDCN	ELV	CIN	RSP	MIB	
MIN	1.3	0.02	87	-	-	0.0	31.8	0.0	0	0	127	-55.6	0.0	-	Occurrence
MÉAN	27.8	1.26	421	0.008	0.008	11.5	-2.2	6.7	261	193	614	-2.2	0.4	-	
MAX	59.0	2.18	823	0.005	0.004	43.8	20.4	16.9	628	664	1089	69.2	1.0	-	
MIN	0.0	0.03	42	-	-	0.0	20.3	0.0	0	0	75	-66.6	0.0	-	Non Occurrence 2 km
MÉAN	16.8	1.05	476	0.000	0.000	6.3	-0.1	7.3	186	219	694	0.1	0.5	-	
MAX	58.5	2.15	1123	0.006	0.005	28.3	29.3	17.8	654	772	1234	70.6	1.0	-	
MIN	0.0	0.02	4	-	-	0.0	29.3	4.2	0	0	4	-66.0	0.0	-	Non Occurrence 5 km
MÉAN	13.5	1.00	513	0.000	0.000	4.8	-0.1	7.6	172	167	681	0.0	0.5	-	
MAX	56.1	2.05	1140	0.005	0.006	29.4	16.6	19.9	696	793	1233	83.3	1.0	-	
MIN	0.0	0.00	3	-	-	0.0	38.0	0.0	0	0	-	-76.9	0.0	-0.50	Non Occurrence 10 km
MÉAN	11.8	0.97	505	0.000	0.000	4.2	-0.2	8.0	168	138	-	-0.4	0.4	0.08	
MAX	58.4	2.13	1219	0.007	0.005	23.7	20.2	22.1	701	692	-	66.5	1.0	2.14	
MIN	0.0	0.03	4	-	-	0.0	29.0	4.5	0	0	-	-92.0	-	-	Non Occurrence 20 km
MÉAN	10.2	0.93	508	0.001	0.000	3.4	0.2	8.2	152	114	-	0.3	-	-	
MAX	55.6	2.13	1218	0.005	0.005	31.2	20.6	21.1	675	733	-	80.0	-	-	
MIN	0.0	0.04	2	-	-	0.0	18.6	4.4	0	0	-	-65.9	-	-0.39	Non Occurrence 30 km
MÉAN	9.0	0.92	480	0.000	0.001	3.0	-0.1	8.7	160	89	-	0.2	-	0.06	
MAX	50.9	2.10	1243	0.004	0.004	28.3	15.0	22.4	689	805	-	84.5	-	2.09	
MIN	0.0	0.05	-	-	-	0.0	22.0	4.6	0	0	0	-79.8	0.0	-0.51	Non Occurrence 40 km
MÉAN	8.0	0.80	-	0.000	0.001	2.7	-0.1	9.1	160	78	526	0.2	0.3	0.06	
MAX	45.5	2.04	-	0.005	0.003	42.6	15.8	22.0	682	640	1318	78.3	1.0	2.15	

Tabela 4.2. Valores mínimos médios e máximos das variáveis de entrada, separados por *buffer* de amostras de não ocorrência.

A partir da análise da correlação e eliminação de atributos com valores muito baixos, foram confeccionados gráficos em formato boxplot (Figura 4.5) da distribuição dos valores de

cada atributo para os pontos de ocorrência e não ocorrência, a qual permite inferir a maior relevância dos atributos para a ocorrência de deslizamentos.

A separação entre as classes de ocorrência e de não ocorrência é facilitada pelo uso de atributos do terreno que apresentam diferentes padrões de valores entre as duas classes. Quanto mais distintos os valores entre ocorrência e não ocorrência, mais importante o atributo se torna na separação das áreas suscetíveis. Ao analisar a Tabela 4.2, pode-se notar que os valores das variáveis explicativas das amostras de não ocorrência se distanciam das amostras de deslizamentos à medida que aumenta o *buffer* de 2 para 40 km. Por exemplo, o atributo SLP, que apresentou uma média de 8° nas amostras coletadas dentro do *buffer* de 40 km, chega a 16,8° quando o *buffer* é reduzido para apenas 2 km. Isto é, a redução do *buffer* fez com que as amostras de não ocorrência tivessem valores de declividade mais próximos dos valores observados nas cicatrizes de deslizamentos, com média de 27,8°. O mesmo ocorreu com atributos como AHS, LSF e TWI.

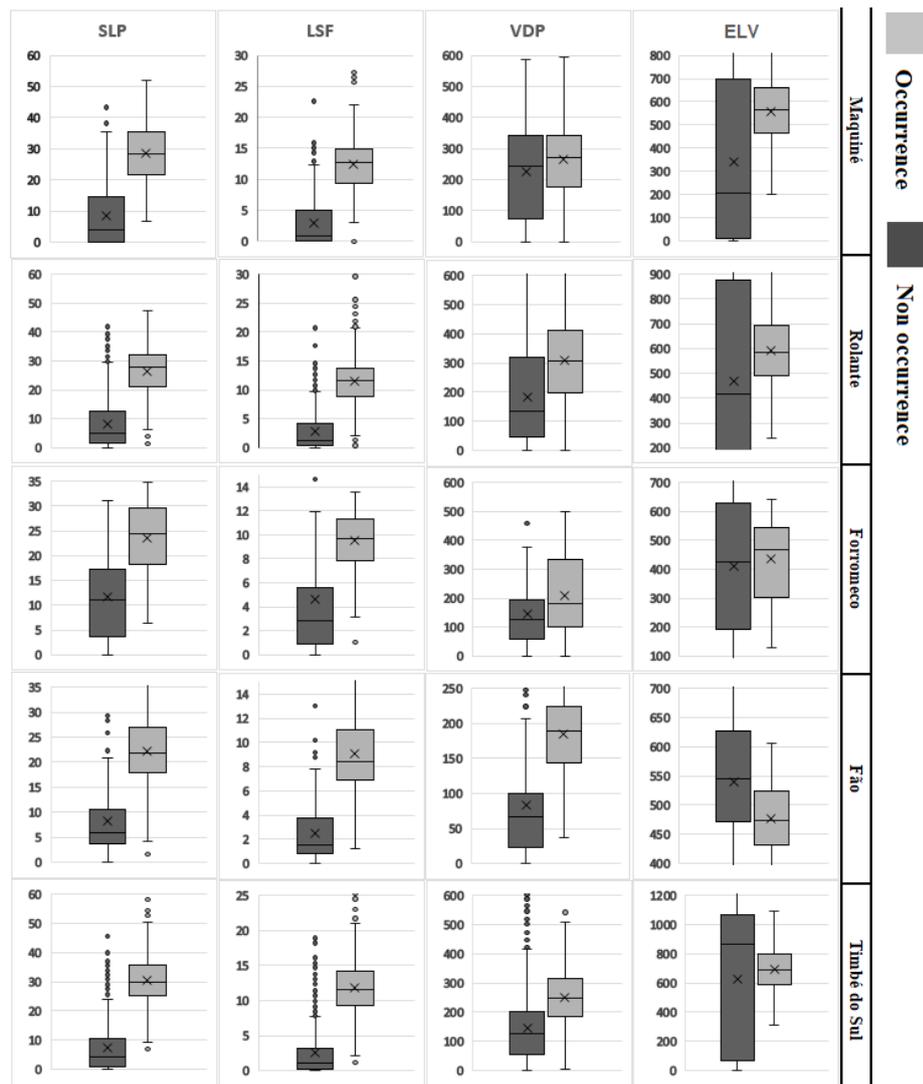


Figura 4.5. Gráficos de boxplots com os parâmetros mais importantes para o *buffer* de 40 km.

Os atributos do terreno que apresentaram maiores distinções foram SLP, LSF e VDP, corroborando com o resultado da correlação das amostras, na qual obtiveram maiores valores (0,65, 0,65 e 0,33, respectivamente). O atributo SLP como um dos mais importantes confirma os resultados encontrados por Quan & Lee (2012), Kavzoglu *et al.* (2014), Hong *et al.* (2017). Hong *et al.* (2016) concluíram que o LSF possui alta correlação com a ocorrência de deslizamentos e ainda, Gameiro *et al.* (2019) e Quevedo *et al.* (2019) encontraram o VDP como um dos parâmetros mais importantes para modelar a suscetibilidade na bacia do Rolante.

Valores de AHS, CNBL, CSC, LOC, TPI, TWI, VDCN e CIN também se apresentam como atributos de distinção entre as amostras de ocorrência e não ocorrência. Entretanto, devido aos valores muito próximos entre as amostras de ocorrência e não ocorrência, esses atributos se tornam menos efetivos na diferenciação de áreas suscetíveis a deslizamentos. Os atributos MBI, RSP, FAC, ELV e ASP apresentaram correlações lineares menores do que 0,06, sendo considerados pouco eficientes para diferenciar áreas propensas.

A elevação, atributo usado na maioria dos modelos de suscetibilidade e que, geralmente, apresenta elevada importância na composição do mapa final, não resultou em uma boa separabilidade das classes de ocorrência e de não ocorrência a deslizamentos (Figura 4.5). Isso se deve ao fato de terem sido incorporados eventos de deslizamentos em diferentes áreas da Serra Geral. Os valores de elevação da escarpa erosiva do planalto apresentam uma alta variabilidade ao longo da área de estudo, fazendo com que os pontos de ruptura dos deslizamentos apresentem diferenças altimétricas para cada uma das cinco áreas. Nota-se ainda que a importância deste atributo reduz à medida que o *buffer* aumenta e novas unidades de relevos são incorporadas nas amostras de não ocorrência. Isso se deve pelo fato de que com áreas mais distantes, os valores de elevação para as amostras de não ocorrência acabam por ter uma maior amplitude. Apesar disso, o modelo com maior acurácia atribuiu uma importância relativa considerável para este atributo (Tabela 4.3), mostrando que mesmo não tendo bons resultados na distinção por boxplots e na análise de correlação, o atributo da elevação ainda deve ser considerado de importante influência na modelagem e mapeamento da suscetibilidade a deslizamentos.

Atributo	SLP	AHS	CSC	LOC	LSF	TPI	TWI	VDP	VDCN	ELV	CIN	RSP	MBI
Importância (%)	35	3	3	1	14	4	7	5	0	14	0	0	1

Tabela 4.3. Importância relativa de cada atributo no melhor modelo.

4.4.2. Análise de acurácia

A Tabela 4.4 mostra a melhor configuração de modelo de RNA (entradas e neurônios na camada oculta) para cada conjunto amostral (área de treinamento e *buffer*), sendo considerada a acurácia nas amostras de testes de toda a Serra Geral. A Figura 6 compara visualmente os resultados dos modelos destacados na Tabela 4.4 em um recorte da Serra Geral.

<i>Buffer</i>	<i>Áreas da Serra Geral</i>				
	<i>Maquiné</i>	<i>Rolante</i>	<i>Forromeco</i>	<i>Fão</i>	<i>Timbé</i>
<i>2 km</i>	0,788	0,759	0,813	0,810	0,791
<i>5 km</i>	0,879	0,817	0,868	0,861	0,848
<i>10 km</i>	0,912	0,739	0,874	0,896	0,871
<i>20 km</i>	0,916	0,824	0,874	0,897	0,861
<i>30 km</i>	0,926	0,786	0,919	0,919	0,872
<i>40km</i>	0,912	0,873	0,907	0,931	0,872

Tabela 4.4. Acurácia do melhor modelo, considerando apenas as amostras de teste, separados por área de treinamento e *buffer* de amostras de não ocorrência.

Nota-se uma clara influência da distância das amostras de não ocorrência para as áreas com deslizamentos no desempenho dos modelos. À medida que o *buffer* aumenta, foi observado um gradativo aumento também no valor de acurácia, o que também ocorreu em estudos de Oliveira *et al.* (2019b). Considerando um *buffer* de apenas 2 km, a acurácia dos modelos ficou entre 0,759 e 0,813. Ao usar um *buffer* de 40 km, a acurácia foi de 0,872 (Timbé) a 0,931 (Fão). Isso se deve ao fato de que quanto mais distantes as amostras de não ocorrência estão das cicatrizes de deslizamentos, maior é a separabilidade entre as classes, uma vez que aumenta o número de amostras em outros compartimentos do relevo, resultando em valores de atributos do terreno diferentes. Logo, quanto maior o *buffer*, menor a dificuldade para a RNA estabelecer pesos sinápticos capazes de separar áreas suscetíveis das demais.

As áreas amostrais mais representativas para o mapeamento da suscetibilidade na Serra Geral foram as do Maquiné, Forromeco e Fão. Treinando as redes neurais com as amostras dessas três áreas, obteve-se uma maior capacidade de extrapolação dos modelos para toda a área de estudo, com acurácia dos melhores modelos superior a 0,9 considerando as amostras de teste bem distribuídas em toda Serra Geral (Tabela 4.4). A explicação para isso é que as cicatrizes identificadas nestas três áreas foram mais abrangentes no que tange os compartimentos geomorfológicos e os atributos morfométricos, ainda que em menor número

absoluto de deslizamentos. Nas regiões de Rolante e de Timbé do Sul, as cicatrizes de deslizamentos estão mais concentradas no espaço, na sua totalidade em áreas mais declivosas e próximas da borda superior da escarpa erosiva, resultando em menor variabilidade dos índices morfométricos. Consequentemente, isso induz o modelo a ser mais restritivo e com menor capacidade de extrapolação para grandes áreas. Isso fica claro na Figura 4.6, onde se observa que os modelos treinados apenas com as amostras de Forromeco (Figura 4.6C) e Fão (Figura 4.6D), foram os que resultaram em mapas com maior abrangência das áreas suscetíveis, em contraponto aos mapas resultantes da calibração do modelo com as amostras de Rolante e de Timbé do Sul.

Algumas estatísticas corroboram com essa afirmação. A elevação e a declividade média das áreas com ocorrência de deslizamentos na Serra Geral foi de 614 m e 27,8°, respectivamente. Em Timbé do Sul, por exemplo, os deslizamentos ocorreram em altitudes mais elevadas (média de 692 m) e vertentes mais íngremes (média de 30,3°). Isso ajuda a explicar porque esse modelo foi mais restritivo em relação às áreas suscetíveis, tendo maior dificuldade para produzir um mapa de áreas suscetíveis com qualidade para toda a Serra Geral (Figura 4.6E). O modelo calibrado com as amostras de Rolante resultou também em um mapa restritivo de áreas suscetíveis a deslizamentos (Figura 4.6B). Neste mapa, as áreas suscetíveis ficaram preferencialmente em locais com altitude maior (média de 1.000 m), próximo da borda superior da escarpa do planalto. Por outro lado, considerando as amostras de Maquiné, os deslizamentos ocorreram preferencialmente em altitudes em torno de 560 m e declividades de 28°. Esses valores se aproximam do perfil geral das amostras da área de estudo, resultando em modelos calibrados com maior capacidade de extrapolação e melhores índices de acurácia quando submetidos a teste em outras áreas da Serra Geral.

Outro aspecto importante é que, na análise visual dos mapas de suscetibilidade apresentados na Figura 4.6, fica claro que dois ou mais modelos de RNAs, com valores próximos e/ou superiores a 0,9 de acurácia, podem apresentar um padrão espacial bem diferente entre si. Esse aspecto já havia sido abordado no estudo de Oliveira *et al.* (2019a), que aplicou modelos empíricos baseados na aprendizagem de máquinas para o mapeamento de áreas suscetíveis a deslizamentos. Esse resultado reforça a importância da análise visual dos mapas oriundos da aplicação de modelos matemáticos, ainda mais quando estes são empíricos e se baseiam exclusivamente na relação entre variáveis de entrada e a saída esperada.

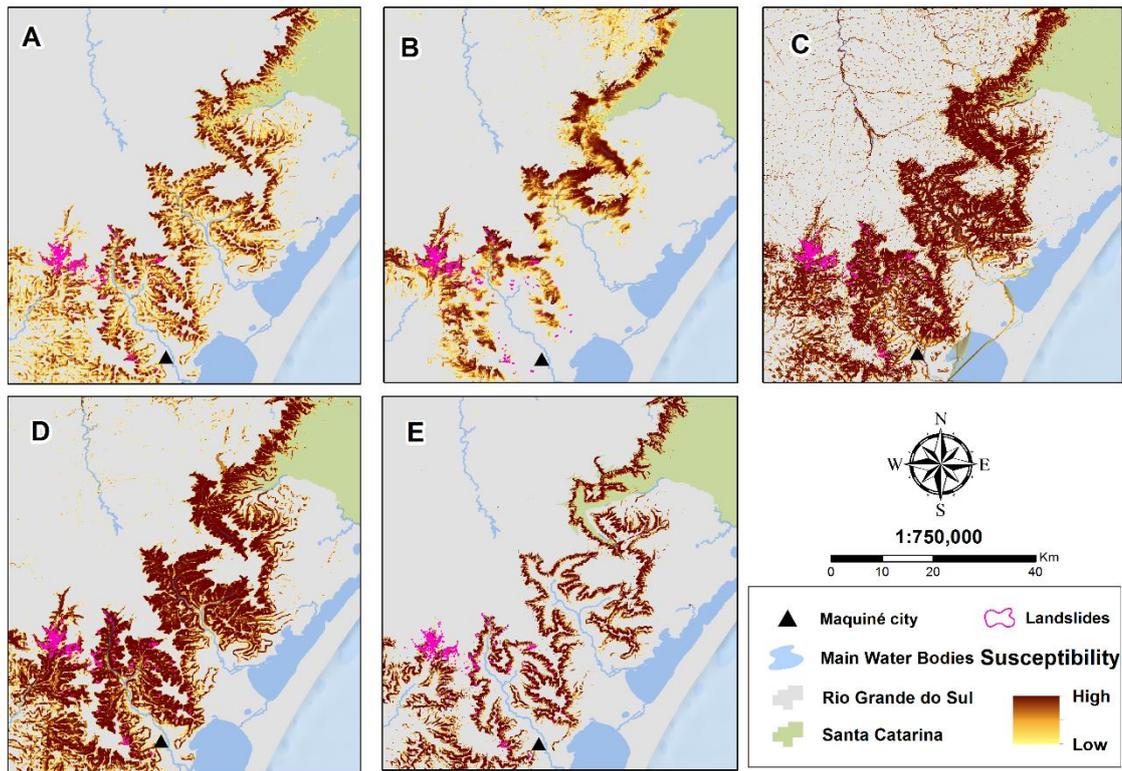


Figura 4.6. Comparação visual dos melhores modelos com amostras de treinamento de cada área: (A) Maquiné; (B) Rolante; (C) Forromeco; (D) Fão; (E) Timbé do Sul.

O modelo com a maior média de acurácia foi treinado com as amostras do Fão e *buffer* de 40 km, atingindo valor de 0,931 (Tabela 4.4). Apesar de ser uma das áreas com menor quantidade de cicatrizes, as amostras abrangem distintas características do relevo, permitindo gerar modelos mais representativos quando extrapolados para as áreas adjacentes. Nesse modelo foram excluídos os atributos ASP, CNBL e FAC, os quais tiveram correlação linear próxima a 0. A partir desse modelo de maior acurácia, foi gerado o mapa de suscetibilidade para a Serra Geral (Figura 4.7) e, para melhor análise e visualização, foi realizada uma aproximação nas áreas amostrais com presença de deslizamentos, a fim de se analisar como esse modelo se comportou nas áreas de teste (Figura 4.8).

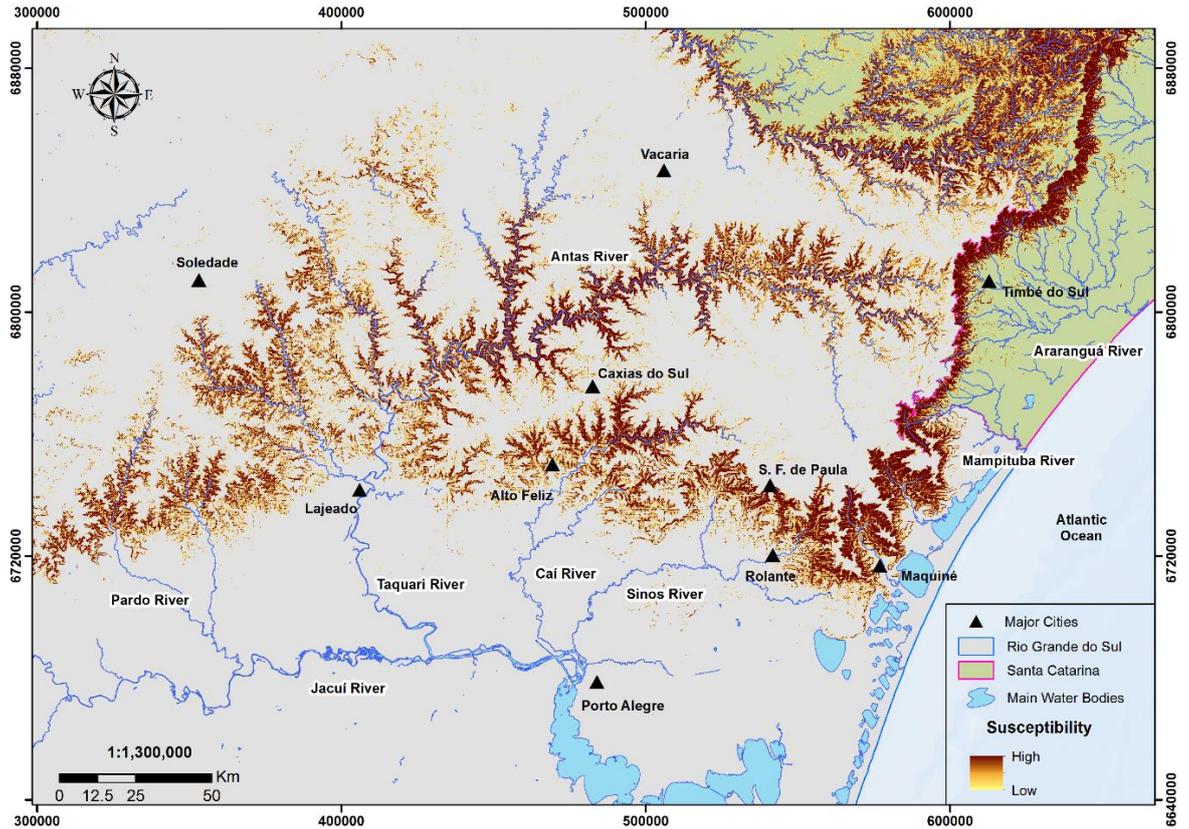


Figura 4.7. Mapa de suscetibilidade a deslizamentos, na Serra Geral, sul do Brasil.

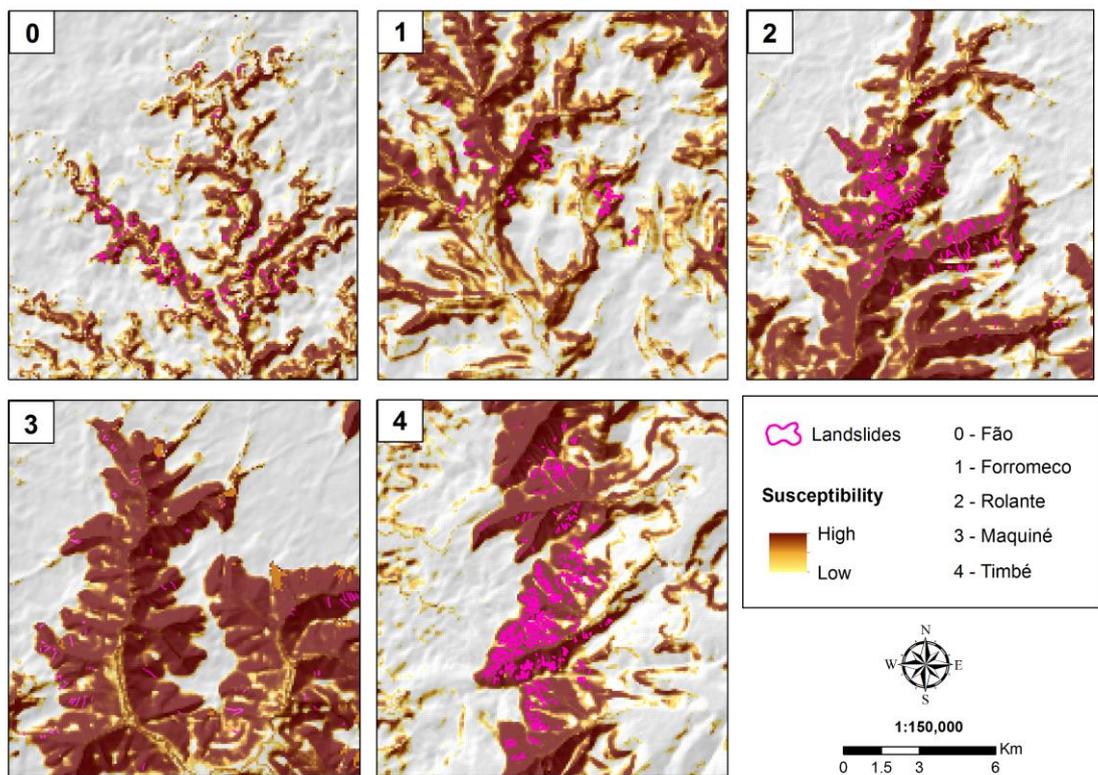


Figura 4.8. Suscetibilidade do melhor modelo para as áreas amostrais dos deslizamentos

Na Figura 4.8 fica claro o excelente ajuste do modelo, com mais de 90% de concordância na indicação da suscetibilidade nas áreas utilizadas apenas para testar a qualidade do mapa final. De modo geral, os locais com ocorrência de deslizamentos foram devidamente destacados pelo modelo, mesmo nas quatro áreas em que o modelo não havia recebido informações para o treinamento, reforçando a capacidade de generalização e de extrapolação dos modelos de RNAs, tanto para análise a deslizamento quanto para outros tipos de modelagens, como simulações hidrológicas Oliveira *et al.* (2014) e problemas estruturais (Araújo, 2015)

Cabe destacar que, mesmo no pior dos modelos, a acurácia foi superior a 0,739. A maioria dos testes resultou em acurácia superior a 0,85. Essas taxas de concordância são muito boas, uma vez que o teste foi realizado incluindo áreas bem distantes do local onde o modelo foi calibrado. Além disso, as acurácias obtidas são compatíveis e até mais eficientes quando comparadas com outros estudos que aplicaram as RNAs, tais como Kalantar *et al.* (2018) e Sameen *et al.* (2020) que obtiveram valores máximos de 0,79 na curva ROC, e também Rahmati *et al.* (2017) e Zhu *et al.* (2018), os quais obtiveram acurácias máximas de 0,91 e 0,92, respectivamente. Tudo isso comprova que a metodologia de usar combinados um inventário de cicatrizes, um MDE e um modelo baseado em RNA pode ser aplicada com excelentes resultados na identificação de outras áreas suscetíveis a deslizamentos.

Por fim, deve-se salientar um aspecto importante em relação à afirmação de que os modelos aqui apresentados foram capazes de extrapolar as áreas suscetíveis para toda a Serra Geral, Sul do Brasil. Ainda que os resultados apresentados corroborem com essa afirmação, é importante mencionar de que essa extrapolação foi apenas de domínio espacial e não numérico dos atributos morfométricos usados como entrada dos modelos. Isto é, constatou-se de que as RNAs são capazes de gerar mapas de suscetibilidade a deslizamentos em grandes áreas como a do presente estudo, desde que o conjunto de treinamento (ainda que restrito espacialmente) seja representativo do domínio dos valores dos atributos de entrada na totalidade da área. É evidente que isso se aplica apenas para áreas com relativa homogeneidade geomorfológica e geológica, como afirmado também por Wang *et al.* (2020). Uma vez que nos afastamos da área de treinamento, a tendência é de que essa homogeneidade diminua, alterando significativamente o domínio numérico dos atributos de entrada do modelo, reduzindo consideravelmente a sua capacidade de extrapolação espacial e, conseqüentemente, a qualidade e exatidão do mapa.

4.5 CONCLUSÕES

A principal contribuição da pesquisa aqui apresentada foi ilustrar como a amostragem afeta o mapeamento de áreas suscetíveis a deslizamentos a partir da modelagem com RNAs, bem como influencia na sua capacidade de generalização e de extrapolação. A partir de dezenas de aplicações das redes neurais, utilizando pequenas áreas amostrais para treinamento, e testando o modelo em outras áreas adjacentes, foi possível concluir que:

- i) as variáveis de entrada mais relevantes no processo de modelagem foram a declividade (SLP), o fator LS (LSF) e a elevação (ELV), ainda que o último tenha uma importância reduzida em comparação com estudos que utilizaram essa variável em pequenas áreas;
- ii) a acurácia dos modelos para o mapeamento das áreas suscetíveis a deslizamentos aumentou proporcionalmente à distância (*buffer*) utilizada para obtenção de amostras de não ocorrência, uma vez que, para *buffers* de apenas 2 km o melhor modelo obteve acurácia de 0,813, enquanto que para um *buffer* de 40 km foi obtido uma acurácia de 0,931. Isso significa que, quanto maior o *buffer*, menor a dificuldade para a RNA estabelecer pesos sinápticos capazes de separar áreas suscetíveis das demais;
- iii) a representatividade das amostras interferiu significativamente na capacidade dos modelos para extrapolar os resultados com qualidade para locais mais distantes. As cicatrizes localizadas em áreas mais representativas da Serra Geral, como as observadas nas regiões do Fão, Forromeco e Maquiné, foram mais eficientes no processo de treinamento do modelo, resultando em maiores taxas de acerto quando aplicado e testado em outras áreas;
- iv) as RNAs, combinadas com um inventário de cicatrizes e um MDE, se mostraram eficientes para a modelagem e o mapeamento da suscetibilidade a deslizamentos, além de terem demonstrado excelente capacidade de generalização e de extrapolação espacial. Essa conclusão se refere a áreas com relativa homogeneidade geomorfológica, uma vez que a alteração do domínio numérico dos atributos de entrada do modelo tende a reduzir consideravelmente a sua capacidade de extrapolação espacial e, conseqüentemente, a qualidade e exatidão do mapa de áreas suscetíveis.

4.6 AGRADECIMENTOS

Essa pesquisa foi financiada pela CAPES (processo 88882.438888/2011-01) e pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) Edital 01/2017 - ARD, processo 17/2551-0000894-4.

4.7 REFERÊNCIAS

ADA, M.; SAN, B. T. Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. **Nat Hazards** 90:237–263, 2018

AL-NAJJAR, H. A. H.; KALANTAR, B.; PRADHAN, B.; SAEIDI, V. Conditioning factor determination for mapping and prediction of landslide susceptibility using machine learning algorithms. **Proc. SPIE 11156, Earth Resources and Environmental Remote Sensing/GIS Applications X**, 111560K, 2019. doi: 10.1117/12.2532687

ARAÚJO, T. M. A. **Redes neurais incorporadas a métodos numéricos para solução de problemas de engenharia estrutural**. Dissertação (Mestrado em Engenharia Civil) – Centro de Tecnologia, Universidade Federal de Alagoas, p. 91, 2015.

BETIOLLO, L. M. **Caracterização estrutural, hidrogeológica e hidroquímica dos sistemas aquíferos guarani e serra geral no nordeste do Rio Grande do Sul, Brasil**. Dissertação (Mestrado em Geociências) – Instituto de Geociências, Universidade Federal do Rio Grande do Sul, p. 117, 2006.

BOCK, M.; KÖTHE, R. **Predicting the Depth of hydromorphic Soil Characteristics influenced by Ground Water**, in: *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie-20gie (SAGA – Seconds Out)*, edited by: Böhner, J., Blaschke, T., and Montanarella, L., Univ.Hamburg, Inst. für Geographie, 13–22, 2008.

BOKWA, A. **Natural Hazard**. In: Bobrowsky P.T. (eds) *Encyclopedia of Natural Hazards*. Encyclopedia of Earth Sciences Series. Springer, Dordrecht, 2013.

BUI, D. T.; TUAN, T. A., KLEMPE, H.; PRADHAN, B.; REVHAUG, I. Spatial Prediction Models for Shallow Landslide Hazards: a Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree. **Landslides** 13 (2), 361–378, 2016. <https://doi.org/10.1007/s10346-015-0557-6>.

CANTARINO, I.; CARRION, M. A.; GOERLICH, F.; IBÁÑEZ, V. M. A ROC analysis-based classification method for landslide susceptibility maps. **Landslides** 16:265–282, 2019. <http://doi.org/10.1007/s10346-018-1063-4>

CATANI, F.; LAGOMARSINO, D.; SEGONI, S.; TOFANI, V. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. **Nat. Hazards Earth Syst. Sci.** 13, 2815–2831, 2013.

CHEN, W.; POURGHASEMI, H. R.; KORNEJADY, A.; ZHANG, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. **Geoderma** 305, 314–327, 2017.

CHEN, W.; XIE, X.; WANG, J.; PRADHAN, B.; HONG, H.; BUI, D. T.; DUAN, Z.; MA, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. **Catena** 151, 147–160, 2017.

CHEN, W.; PENG, J.; HONG, H.; SHAHABI, H.; PRADHAN, B.; LIU, J.; ZHU, A-X.; PEI, X.; DUAN, Z. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. **Science of the Total Environment** 626, 1121–1135, 2018.

FERNANDES, N. F.; GUIMARÃES, R. F.; GOMES, R. A. T.; VIEIRA, B. C.; MONTGOMERY, D. R.; GREENBERG, H. Topographic controls of landslides in Rio de Janeiro: field evidence and modeling. **Catena** 55 (2), 163-181, 2004.

GALLI, M.; ARDIZZONE, F.; CARDINALI, M.; GUZZETTI, F.; REICHENBACH, P. Comparing landslide inventory maps. **Geomorphology** 94, 268–289, 2008.

GAMEIRO, S.; QUEVEDO, R. P.; OLIVEIRA, G. G.; RUIZ, L. F. C.; GUASSELLI, L. A. Análise e correlação de atributos morfométricos e sua influência nos deslizamentos ocorridos na Bacia do Rio Rolante, RS. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, XIX, 2019 Santos. **Anais...** 17, 2880-2883, 2019.

GOETZ, J. N.; BRENNING, A.; PETSCHKO, H.; LEOPOLD, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. **Computers & Geosciences** 81, 1–11, 2015.

GORDO, C.; ZÊZERE, J. L.; MARQUES, R. Landslide Susceptibility Assessment at the Basin Scale for Rainfall- and Earthquake-Triggered Shallow Slides. **Geosciences** 9 , 268, 2019.

GUISAN, A.; WEISS, S. B.; WEISS, A. D. GLM versus CCA spatial modeling of plant species distribution. **Plant Ecology** 143, 107-122, 1999.

HARTMANN, L. A. A história natural do Grupo Serra Geral desde o Cretáceo até o Recente. **Ciência e Natura**, Santa Maria, 36 Ed. Especial, 173–182, 2014.

HE, Q.; SHAHABI, H.; SHIRZADI, A.; LI, S.; CHEN, W.; WANG, N.; CHAI, H.; BIAN, H.; MA, J.; CHEN, Y.; WANG, X.; CHAPI, K.; AHMAD, B. B. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms, **Science of the Total Environment** 663, 1–15, 2019.

HECKMANN, T.; GEGG, K.; BECHT, M. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flow. **Nat. Hazards Earth Syst. Sci.**, 14, 259–278, 2014.

HONG, H.; NAGHIBI, S. A.; POURGHASEMI, H. R.; PRADHAN, B. GIS-based landslide spatial modeling in Ganzhou City, China. **Arab J Geosci**, 9, 112, 2016. <https://doi.org/10.1007/s12517-015-2094-y>

HONG, H.; LIU, J.; ZHU, A-X.; SHAHABI, H.; PHAM, B. T.; CHEN, W.; PRADHAN, B.; BUI, D. T. A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). **Environ Earth Sci** 76, 652, 2017. <https://doi.org/10.1007/s12665-017-6981-2>

HONG, H.; MIAO, Y.; LIU, J.; ZHU, A-X. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. **Catena** 176, 45–64, 2019.

HUANG, Y.; ZHAO, L. Review on landslide susceptibility mapping using support vector machines. **Catena** 165, 520–529, 2018.

KALANTAR, B.; PRADHAN, B.; NAGHIBI, S. A.; MOTEVALLI, A.; MANSOR, S. Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). **Geomatics, Natural Hazards and Risk**, 9 (1), 49-69, 2018. <https://doi.org/10.1080/19475705.2017.1407368>

KAVZOGLU, T.; SAHIN, E. K.; COLKESEN, I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. **Landslides** 11, 425–439, 2014. <https://doi.org/10.1007/s10346-013-0391-7>

KOTHE, R.; LHEMEIER, F. SAGA – Ein Programmsystem zur Automatischen Relief-Analyse. **Zeitschrift für Angewandte Geographie**, 4, 11-21, 1993.

KUMAR, D.; THAKUR, M.; DUBEY, S. C.; SHUKLA, P. D. Landslide Susceptibility Mapping & Prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India. **Geomorphology**, 295, 115-125, 2017.

LEE, S.; MIN, K. Statistical analysis of landslide susceptibility at Yongin, Korea. **Environ. Geol.** 40, 1095–1113, 2001.

LEE, S.; RYU, J-H.; WON, J-S.; PARK, H-J. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. **Engineering Geology** 71 (3), 289–302, 2004.

LOSASSO, L.; SDAO, F. The artificial neural network for the rockfall susceptibility assessment. A case study in Basilicata (Southern Italy). **Geomatics, Natural Hazards and Risk** 9 (1), 737-759, 2018.

MACHADO, F. B. **Geologia e aspectos petrológicos das rochas intrusivas e efusivas mesozóicas de parte da borda leste da bacia do paraná no estado de São Paulo**. Dissertação (Mestrado em Geociências) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, p. 214, 2005.

MÖLLER, M.; VOLK, M.; FRIEDRICH, K.; LYMBURNER, L. Placing soil-genesis and transport processes into a landscape context: A multiscale terrain-analysis approach. **Journal of Plant Nutrition and Soil Science** 171 (3), 419-430, 2008.

MÖLLER, M.; KOSCHITZKI, T.; HARTMANN, K. J.; JAHN, R. Plausibility test of conceptual soil maps using relief parameters. **Catena**, 88 (1), 57–67, 2012. <https://doi.org/10.1016/j.catena.2011.08.002>

MOOSAVI, V.; NIAZI, Y. Development of hybrid wavelet packet-statistical models (WP-SM) for landslide susceptibility mapping. **Landslides**, 13 (1), 97-114, 2016.

NETO, O.F.; OLIVEIRA, A. G. S. **Caracterização geotécnica do basalto da formação serra geral para utilização na construção civil**. Simpósio Brasileiro de Jovens Geotécnicos – GeoJovem, 2006.

NGUYEN, P. T.; TUYEN, T. T.; SHIRZADI, A.; PHAM, B. T.; SHAHABIM, H.; OMIDVAR, E.; AMINI, A.; ENTEZAMI, H.; PRAKASH, I.; PHONG, T. V.; VU, T. B.; THANH, T.; SARO, L.; BUI, D. T. Development of a Novel Hybrid Intelligence Approach for Landslide Spatial Prediction **Appl. Sci.** 9, 2824, 2019. <https://doi.org/10.3390/app9142824>

NOURANI, V.; PRADHAN, B.; GHAFFARI, H.; SHARIFI, S. S. Landslide susceptibility mapping at Zonouz plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. **Nat. Hazards** 71 (1), 523–547, 2014.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. O desempenho das redes neurais artificiais (RNAs) para Simulação Hidrológica Mensal. **RBRH–Revista Brasileira de Recursos Hídricos**, 19 (2), 251-265, 2014.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. Simplifying artificial neural network models of river basin behavior by an automated procedure for input variable selection. **Eng. Appl. Artif. Intel.** 40, 47–61, 2015. <https://doi.org/10.1016/j.engappai.2015.01.001>

OLIVEIRA, G. G.; RUIZ, L. F. C.; GUASSELLI, L. A.; HAETINGUER, C. (2019a). Random forest and artificial neural networks in landslide susceptibility modeling: a case study of the Fão River Basin, Southern Brazil. **Nat Hazards** 99, 1049, 2019a. <https://doi.org/10.1007/s11069-019-03795-x>.

OLIVEIRA, G. G.; RUIZ, L. F. C.; QUEVEDO, D. M. Redes neurais artificiais e geotecnologias para o mapeamento de áreas suscetíveis a inundações. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, XXIII, Foz do Iguaçu. **Anais...** p. 10, 2019b.

PETSCHKO, H.; BRENNING, A.; BELL, R.; GOETZ, J.; GLADE, T. Assessing the quality of landslide susceptibility maps — case study Lower Austria. **Nat. Hazards Earth Syst. Sci.** 14 (1), 95–118, 2014.

PHAM, B. T.; PRAKASH, I. A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide susceptibility assessment. **Bulletin of Engineering Geology and the Environment**, 78, 1911–1925, 2017. <https://doi.org/10.1007/s10064-017-1202-5>

PHAM, B. T.; PRAKASH, I.; SINGH, S. K.; SHIRZADI, A.; SHAHABI, H.; TRAN, T. T.; BUI, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. **Catena** 175, 203–218, 2019.

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 & Geosciences** 51, 350–365, 2013.

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

QUAN, H. C.; LEE, B. G. GIS-Based landslide susceptibility mapping using analytic hierarchy process and artificial neural network in Jeju (Korea). **KSCE Journal of Civil Engineering**, 16 (7), 1258-1266, 2012.

QUEVEDO, R. P. **Mapeamento de suscetibilidade a deslizamentos a partir de aprendizado de máquina**. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre, p. 71, 2019.

QUEVEDO, R. P.; GUASSELLI, L. A.; OLIVEIRA, G. G.; RUIZ, L. F. C. Modelagem de áreas suscetíveis a deslizamentos: avaliação comparativa de técnicas de amostragem,

aprendizado de máquina e modelos digitais de elevação. São Paulo, UNESP, **Geociências**, 38 (3), 781 – 795, 2019.

RAHMATI, O.; TAHMASEBPOUR, N.; HAGHIZADEH, A.; POURGHASEMI, H. R.; FEIZIZADEH, B. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. **Geomorphology** 298, 118–137, 2017.

ROISENBERG, A.; VIERO, A. P. **O vulcanismo Mesozóico da Bacia do Paraná no Rio Grande do Sul**. In: HOLZ, M.; DE ROS, L. F. (Org.). *Geologia do Rio Grande do Sul*. 2.ed. Porto Alegre: Universidade Federal Rio Grande do Sul, 1, 355-374, 2000.

ROSSATO, M. S. **Os climas do Rio Grande do Sul: variabilidade, tendências e tipologia**. Tese (Doutorado), Instituto de Geociências, Programa de Pós-Graduação em Geografia da Universidade Federal do Rio Grande do Sul, p. 253, 2011.

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

SAMEEN, M. I.; PRADHAN, B.; LEE, S. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. **Catena** 186, 104249, 2020. <https://doi.org/10.1016/j.catena.2019.104249>

SAMIA, J.; TEMME, A.; BREGT, A.; WALLING, J.; GUZZETTI, F.; ARDIZZONE, F.; ROSSI, M. Characterization and quantification of path dependency in landslide susceptibility. **Geomorphology** 292, 16–24, 2017. <https://doi.org/10.1016/j.geomorph.2017.04.039>

SAMIA, J.; TEMME, A.; BREGT, A.; WALLING, J.; GUZZETTI, F.; ARDIZZONE, F. Dynamic path dependent landslide susceptibility modelling. **Nat. Hazards Earth Syst. Sci.**, 20, 271–285, <https://doi.org/10.5194/nhess-20-271-2020>, 2020.

SHAHABI, H.; HASHIM, M. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. **Sci. Rep.**, 5, 9899, 2015. <https://doi.org/10.1038/srep09899>.

SHIRZADI, A.; SOLIAMANI, K.; HABIBNEJHAD, M.; KAVIAN, A.; CHAPI, K.; SHAHABI, H.; CHEN, W.; KHOSRAVI, K.; PHAM, B. T.; PRADHAN, B.; AHMAD, A.; AHMAD, B. B.; BUI, D. T. Novel GIS Based Machine Learning Algorithms for Shallow Landslide Susceptibility Mapping. **Sensors** 18, 3777, 2018. <https://doi.org/10.3390/s18113777>

STEGER, F.; BRENNING, A.; BELL, R.; GLADE, T. The propagation of inventory-based positional errors into statistical landslide susceptibility models. **Nat. Hazards Earth Syst. Sci.** 16, 2729–2745, 2016.

STEGER, S.; KOFLER, C. **Statistical Modeling of Landslides: Landslide Susceptibility and Beyond**. In: POURGHASEMI, H. R.; GOKCEOGLU, C. Spatial Modeling in GIS and R for Earth and Environmental Sciences, Elsevier, 2019. <https://doi.org/10.1016/B978-0-12-815226-3.00024-7>

TARINI, M.; CIGNONI, P.; MONTANI, C. Ambient occlusion and edge cueing for molecular visualization. **IEEE Transactions on Visualization and Computer Graphics** 12 (5), 1237-1244, 2006.

TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. **Desastres Naturais: Conhecer para Prevenir**. São Paulo: Instituto Geológico, 2009.

UNISDR/CRED. **The Human Cost of Weather-Related Disasters 1995-2015**, 2015.

VAN DEN EECKHAUT, M.; VANWALLEGHEM, T.; POESEN, J.; GOVERS, G.; VERSTRAETEN, G.; VANDEKERCKHOVE, L. Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). **Geomorphology** 76 (3–4), 392–410, 2006.

WANG, Y.; FANG, Z.; HONG, H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. **Science of the Total Environment** 666, 975–993, 2019.

WANG, Y.; FENG, L.; LI, S.; REN, F.; DU, Q. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. **Catena** 188, 104425, 2020. <https://doi.org/10.1016/j.catena.2019.104425>

WEISS, A. **Topographic Position and Landforms Analysis**, Posterpresentation, ESRI User Conference, San Diego, CA, 2001.

XIAO, C-C.; TIAN, Y.; SHI, W-Z.; GUO, Q-H.; WU, L. A new method of pseudo absence data generation in landslide susceptibility mapping with a case study of Shenzhen. *Science China*, **Technological Sciences** 53 (1), 75-84, 2010. <https://doi.org/10.1007/s11431-010-3219-x>

XIE, P.; WEN, H.; MA, C.; BAISE, L. G.; ZHANG, J. Application and comparison of logistic regression model and neural network model in earthquake-induced landslides susceptibility mapping at mountainous region, China. **Geomatics, Natural Hazards and Risk** 9 (1), 501-523, 2018. <https://doi.org/10.1080/19475705.2018.1451399>

XU, C.; SHYU, J. B. H.; XU, X-W. Landslides triggered by the 12 January 2010 Mw 7.0 Port-au-Prince, Haiti, earthquake: visual interpretation, inventory compiling and spatial distribution statistical analysis **Nat. Hazards Earth Syst. Sci. Discuss.**, 2, 1259–1331, 2014. <https://doi.org/10.5194/nhess-14-1789-2014>

YAO, X.; THAM, L. G.; DAI, F. C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. **Geomorphology** 101, 572–582, 2008.

ZANIN, P. R.; BONUMÁ, N. B.; MINELLA, J. P. G. Determinação do Fator Topográfico em Bacias Hidrográficas. **Rev. Bras. Geomorfol.** (Online), São Paulo, 18 (1), 19-36, 2017.

ZEVENBERGEN, L. W.; THORNE, C. R. Quantitative analysis of land surface topography. **Earth surface processes and landforms**, 12, 47-56, 1987.

ZÊZERE, J. L.; PEREIRA, S.; MELO, R.; OLIVEIRA, S. C.; GARCIA, R. A. C. Mapping landslide susceptibility using data-driven methods. **Science of Total Environment** 589, 250–267, 2017. <http://dx.doi.org/10.1016/j.scitotenv.2017.02.188>

ZHU, A. X.; MIAO, Y.; YANG, L.; BAI, S-B.; LIU, J.; HONG, H. Comparison of the presence-only method and presence-absence method in landslide susceptibility mapping. **Catena**, 171, 222-233, 2018.

CAPÍTULO 5: COMPARAÇÃO ENTRE DISTINTAS ÁREAS DE AMOSTRAGEM NO MAPEAMENTO DE SUSCETIBILIDADE A DESLIZAMENTOS COM REDES NEURAIS ARTIFICIAIS.

RESUMO

Os deslizamentos de terra podem provocar sérias consequências ambientais, econômicas e sociais. O uso de redes neurais artificiais (RNA) para o mapeamento desses deslizamentos tem se tornado mais frequente a cada ano, sendo um dos métodos mais confiáveis para tal. Entre os principais influenciadores dos mapas gerados, destacam-se as áreas amostrais, que influenciam diretamente no resultado final. Este artigo se propôs a avaliar critérios da amostragem de cicatrizes de deslizamento e modelagem por RNA analisando a influência da distância nos processos de amostragem, o uso de múltiplos eventos de deslizamento e a relação entre os atributos de terreno e os modelos de suscetibilidade. Para tal, foram utilizadas 5 áreas amostrais na Serra Geral no sul do Brasil, *buffers* de distância no processo de amostragem de não ocorrência e 16 atributos de terreno. O treinamento da rede de múltiplas camadas foi realizado pelo algoritmo retropropagativo e a acurácia foi calculada por meio da *Area Under the Receiver Operating Characteristic Curve* (ROC). Os resultados mostraram que quanto maiores as distâncias das coletas das amostras de não ocorrência, maior será a acurácia do modelo, sendo que os melhores modelos foram obtidos utilizando um *buffer* de 40 km. Mostraram ainda que o uso de múltiplos eventos (*full model*) resultou em melhores resultados do que quando usado cada evento separadamente (*single model*), obtendo acurácias de 0,954 e 0,931 respectivamente. Isso se deve principalmente pelo fato de haver uma maior diferenciação entre as amostras de ocorrência e não ocorrência quando se utilizam múltiplos eventos, facilitando assim a distinção entre áreas mais e menos suscetíveis. Entre os atributos de maior destaque, ressalta-se a declividade, fator LS, elevação e profundidade de vale, os quais obtiveram resultados de relação linear (R^2) acima de 0,35, sendo, por isso, considerados de extrema importância para o mapeamento da suscetibilidade a deslizamentos.

Palavras-chave: Amostragem; aprendizado de máquina; desastres naturais; atributos preditivos

ABSTRACT

Landslides can have serious environmental, economic, and social consequences. Using artificial neural networks (ANN) to map these landslides is becoming more frequent every year, being one of the most reliable methods for this. Among the prime influences on the generated maps, sample areas are significantly interesting, since they directly influence the final result. This article aims to evaluate the criteria for landslide scar sampling and ANN modeling by analyzing the influence of distance on the

sampling processes, the use of multiple landslide events, and the relationship between terrain attributes and susceptibility models. To this end, were used five sampling areas in the Serra Geral, southern Brazil, distance buffers in the non-occurrence sampling process, and 16 terrain attributes. The training of the multilayer network was carried out by backpropagation algorithm, and the accuracy was calculated using the Area Under the Receiver Operating Characteristic Curve (ROC). The results showed that the greater the distances of the non-occurring samples, the greater the accuracy of the model, with a 40 km buffer resulting in the best models. They also showed that the use of multiple events (full model) produced better results than each event used separately (single model), obtaining accuracies of 0.954 and 0.931, respectively. This is mainly because there is greater differentiation between occurrence and non-occurrence samples when using multiple events, thus facilitating the distinction between more and less susceptible areas. Among the most prominent attributes, slope angle, LS factor, elevation, and valley depth stand out, as their linear relation (R^2) is above 0.35, and, therefore, considered of extreme importance for mapping landslide susceptibility.

Keywords: Sampling; Machine learning; Natural hazards; Predictive attributes

5.1 INTRODUÇÃO

Deslizamentos resultam de mudanças nas variáveis ambientais de determinada região, com influência direta do ambiente (geologia, geomorfologia, declividade, etc.) e das atividades humanas, podendo provocar consequências econômicas e sociais (Pandey & Sharma, 2017; Zhao *et al.*, 2019). O mapeamento de suscetibilidade a deslizamentos é de essencial importância para entender e prever futuros deslizamentos, contribuindo para a mitigação das suas consequências e para o planejamento do uso e ocupação, principalmente em áreas de elevadas altitudes e declividades (Yalcin, 2008; Chen *et al.*, 2019).

Com o avanço de técnicas de Sistemas de Informações Geográficas (SIG) e sensoriamento remoto (SR), diversas metodologias para modelagem têm sido utilizadas para mapeamentos de suscetibilidades. Entre as mais utilizadas estão a regressão logística (Yilmaz, 2009; Hong *et al.*, 2015; Bui *et al.*, 2016), pesos de evidência (Pourghasemi *et al.*, 2012; Hong *et al.*, 2018a) e índice de estatística (Regmi *et al.*, 2014; Aghdam *et al.*, 2016). Nos últimos anos, as técnicas de aprendizado de máquina (AM) vêm se destacando na geração de modelos de suscetibilidade a deslizamentos, com destaque para redes neurais artificiais (Chen *et al.*, 2017a; Oliveira *et al.*, 2019a; Lucchese *et al.*, 2020; Wang *et al.*, 2020), árvore de decisão (Oliveira *et al.*, 2019a; Sachdeva *et al.*, 2019;) e máquinas de vetores de suporte (Sachdeva *et al.*, 2019; Wang *et al.*, 2020). Esses modelos funcionam extraindo conhecimentos a partir de amostras previamente selecionadas e passam, antes de tudo, por uma fase de treinamento através de variáveis explicativas, as quais devem ser cuidadosamente selecionadas a fim de gerar um treinamento concordante e conseguir utilizar toda a capacidade de aprendizagem que a técnica permite.

Como primeiro passo para utilizar esses modelos é necessário um inventário de cicatrizes preciso e completo para a confecção de modelos confiáveis (Görüm, 2019). Esse inventário é empregado para realizar o processo de amostras de ocorrência e não ocorrência dos eventos e, comumente, se utiliza a estratégia 1:1 (Heckmann *et al.*, 2014). Ainda no processo de amostragem, muito se debate sobre as proporções de amostras para treinamento e validação. De acordo com Baeza *et al.* (2010) pelo menos 50% das amostras devem ser usadas para o treinamento do modelo a fim de se obter uma análise discriminante. O que ainda é pouco discutido nesse processo é a abrangência das amostras de não ocorrência. Quevedo *et al.* (2019) afirmam que áreas amostrais mais abrangentes resultam em melhores resultados de acurácia e generalização do que quando se utiliza amostras mais restritivas. Segundo Hong *et al.* (2019) para um número fixo de amostras de não ocorrência, se o espaço de amostragem for pequeno,

haverá superestimação do modelo e, para intervalos de áreas de amostragem de não ocorrência fixos, haverá uma subestimação se o número de amostras for muito grande, mostrando a intensa relação entre quantidade de amostras e espaço amostral. Oliveira *et al.* (2019b) observaram que para mapeamentos de inundações, quanto maior a distância de coleta das amostras de não ocorrência, maior é a acurácia do modelo e que a proximidade entre amostras de ocorrência e de não ocorrência torna os modelos mais restritivos e com menor capacidade de extrapolação.

Outro fator de destaque e que pouco é abordado em estudos de suscetibilidade a deslizamentos é o conjunto amostral considerando distintos eventos. A grande maioria dos estudos utilizam amostras de apenas uma área (Shirzadi *et al.*, 2019; Sameen *et al.*, 2019) ou amostras históricas de diferentes épocas e que ainda possuam registros (Van Den Eeckhaut *et al.*, 2012; Valenzuela *et al.*, 2017; Wang *et al.*, 2019). Com isso, ainda existem dúvidas se a ampliação do conjunto amostral, considerando múltiplos eventos de deslizamentos em diferentes áreas de uma mesma região geomorfológica exerce uma influência significativa na qualidade visual e na acurácia dos mapas de suscetibilidade a deslizamentos.

Como último ponto pouco destacado nos processos de modelagem de suscetibilidade a deslizamentos, ressalta-se a relação entre as RNAs e as variáveis de entrada. Mesmo com as RNAs sendo modelos empíricos e complexos, com expressivas conexões sinápticas e sendo consideradas caixas pretas por muitos pesquisadores, é possível analisar e interpretar os seus resultados a ponto de identificar as relações entre os resultados finais da suscetibilidade e as variáveis que foram usadas como entrada para os modelos?

Esta pesquisa se propõe a avaliar critérios da amostragem de cicatrizes de deslizamentos e da modelagem com redes neurais artificiais para o mapeamento de áreas suscetíveis a deslizamentos na Serra Geral, sul do Brasil. Entre as análises realizadas, destacam-se:

- i) a definição de *buffers* para a coleta de amostras de não ocorrência;
- ii) a utilização de simples ou múltiplos eventos de deslizamentos para o treinamento das redes neurais;
- iii) a avaliação da influência dos atributos do terreno (variáveis independentes) na composição do resultado final dos mapas de suscetibilidade.

5.2 METODOLOGIA

5.2.1 Área de estudo

A área adotada como estudo de caso (Figura 5.1), denominada Serra Geral, é caracterizada por declives acentuados e altitudes de até 1.400 m. Localiza-se nas bordas do Planalto Meridional, e tem sua origem na erosão diferencial e da tectônica. Abrange a região central e nordeste do Estado do Rio Grande do Sul e o sul do Estado de Santa Catarina.

É constituída predominantemente por rochas vulcânicas, com destaques para basaltos e andesitos, e pequenas porções de riolitos e riodacitos (Hartmann, 2014). Devido a sua grande extensão, é considerada uma das maiores províncias basálticas do mundo (Neto, 2006). Possui lineamentos estruturais em distintas direções, relevo do tipo falhado e solos com média de 50 centímetros de espessura (Betiollo, 2006).

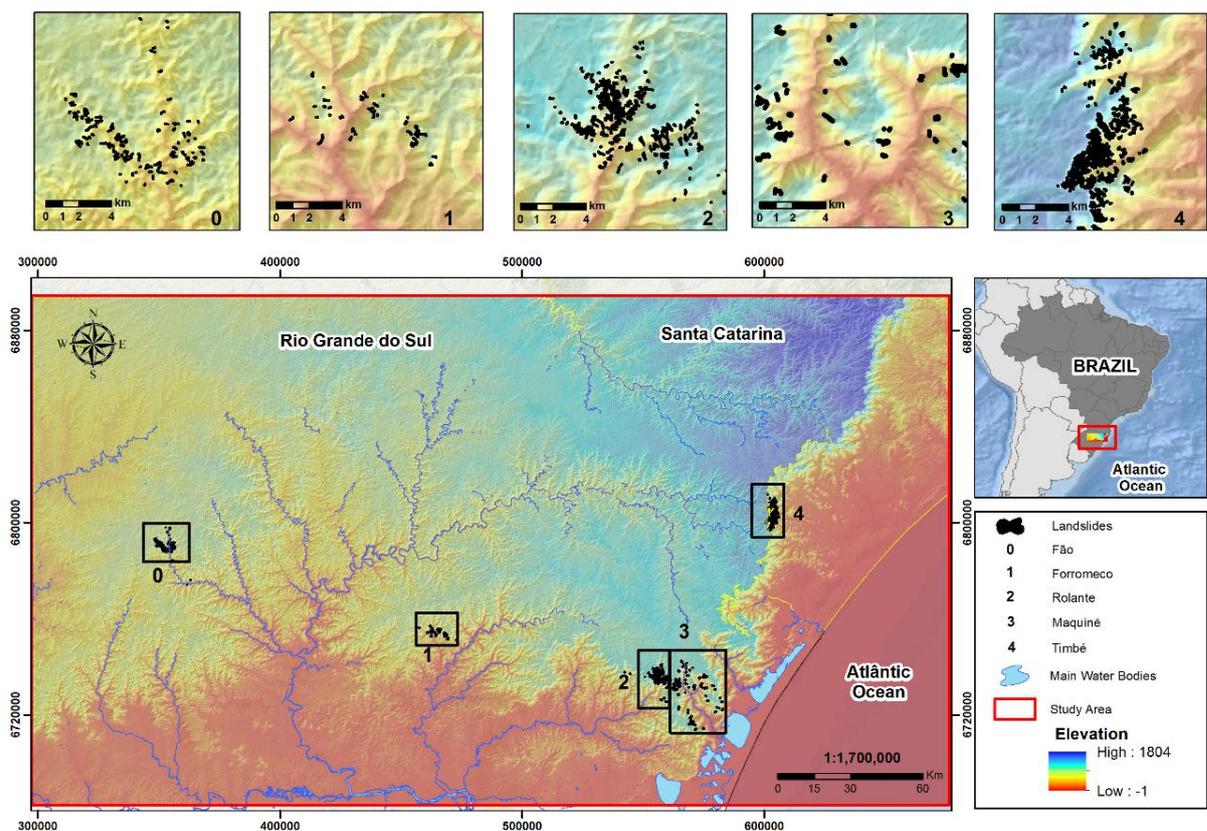


Figura 5.1 Área de estudo e localização das áreas amostrais de cicatrizes a deslizamentos.

O clima da região varia em função de fatores como altitude e relevo e, de acordo com Rossato (2011), se enquadra em subtropical úmido com variação longitudinal das temperaturas médias na região centro norte e subtropical muito úmido com inverno frio e verão fresco na região a nordeste, juntamente na divisa de Estados entre Rio Grande do Sul e Santa Catarina. O verão é a época mais suscetível a ocorrência de deslizamentos em áreas declivosas, quando ocorrem precipitações localizadas, intensas e de curta duração.

5.2.2 Conjunto de dados

A preparação dos dados contou primeiramente com a busca por eventos extremos ocorridos na região da Serra Geral, Sul do Brasil e que tivessem acarretado em consideráveis deslizamentos. Foi confeccionado um inventário de cicatrizes de deslizamento utilizado como entrada no processo de modelagem.

Foram utilizadas imagens SRTM (90 m) para confecção de um mosaico que abrangesse toda a extensão da área em estudo, e confeccionados atributos de relevo usados como valores de ocorrência e não ocorrência das amostras de cicatrizes.

Foi utilizado um script de RNA contendo os valores dos atributos morfométricos contidos em cada uma dessas amostras de ocorrência e não ocorrência para a geração de modelos de suscetibilidade a deslizamento, os quais foram analisados através da curva ROC.

5.2.3 Inventário de cicatrizes

Um inventário de cicatrizes confiável é de vital importância para a análise de suscetibilidade a deslizamentos (Jebur *et al.*, 2014). Para realização do inventário de cicatrizes, foram utilizadas imagens multitemporais do Google Earth Pro, devido ao seu fácil acesso e a resolução espacial e temporal de suas imagens, as quais facilitam a identificação das cicatrizes. Esse inventário foi confeccionado através das cicatrizes presentes em 5 áreas amostrais na Serra Geral (Tabela 5.1), nas quais há o registro de cicatrizes de deslizamentos relacionados a eventos de extrema precipitação entre os anos de 1995 e 2017:

Área Amostral	Bacia Hidrográfica	Municípios	Código da Área
Vale do rio Fão	Taquari-Antas	Fontoura Xavier/Soledade	0
Vale do arroio Forromeco	Caí	São Vendelino/Alto Feliz/Carlos Barbosa	1
Vale do rio Mascarada	Sinos	Rolante/Riozinho/São Francisco de Paula	2
Vale do rio Maquiné	Maquiné	Maquiné	3
Vale do rio Rocinha	Araranguá	Timbé do Sul/Jacinto Machado	4

Tabela 5.1 Localização das áreas amostrais de deslizamentos na Serra Geral.

Para delimitação das cicatrizes foram definidos polígono, transferidos para ambiente SIG e convertidas para pontos, constando um ponto por cada pixel dentro das cicatrizes, com

resolução espacial de 90 m, mesma resolução utilizada para a confecção dos atributos de terreno.

Para Bui *et al.* (2016) o mapeamento de suscetibilidade a deslizamentos é um caso clássico de classificação binária, no qual a ocorrência de um deslizamento é resultante do valor 1 e a não ocorrência de um valor 0. Seguindo esta lógica, cada ponto gerado foi definido como uma amostra de ocorrência, num total de 1638 pontos de ocorrência, sendo 164 amostras na área 0 (Fão), 60 na área 1 (Forromeco), 492 na área 2 (Rolante), 209 na área 3 (Maquiné) e 713 na área 4 (Timbé do Sul). Para as amostras de não ocorrência, para cada área gerou-se o respectivo montante de pontos, resultando em 1638 pontos distribuídos de forma aleatória no terreno, com distâncias mínimas de 150 m entre si e com limitações máximas de distância, atribuídas através da criação de *buffers* com distâncias de 2 km, 5 km, 10 km, 20 km, 30 km e 40 km das amostras de ocorrência, conforme figura 5.2. Ou seja, para cada *buffer* utilizado, foram geradas 1638 amostras aleatórias de não ocorrência.

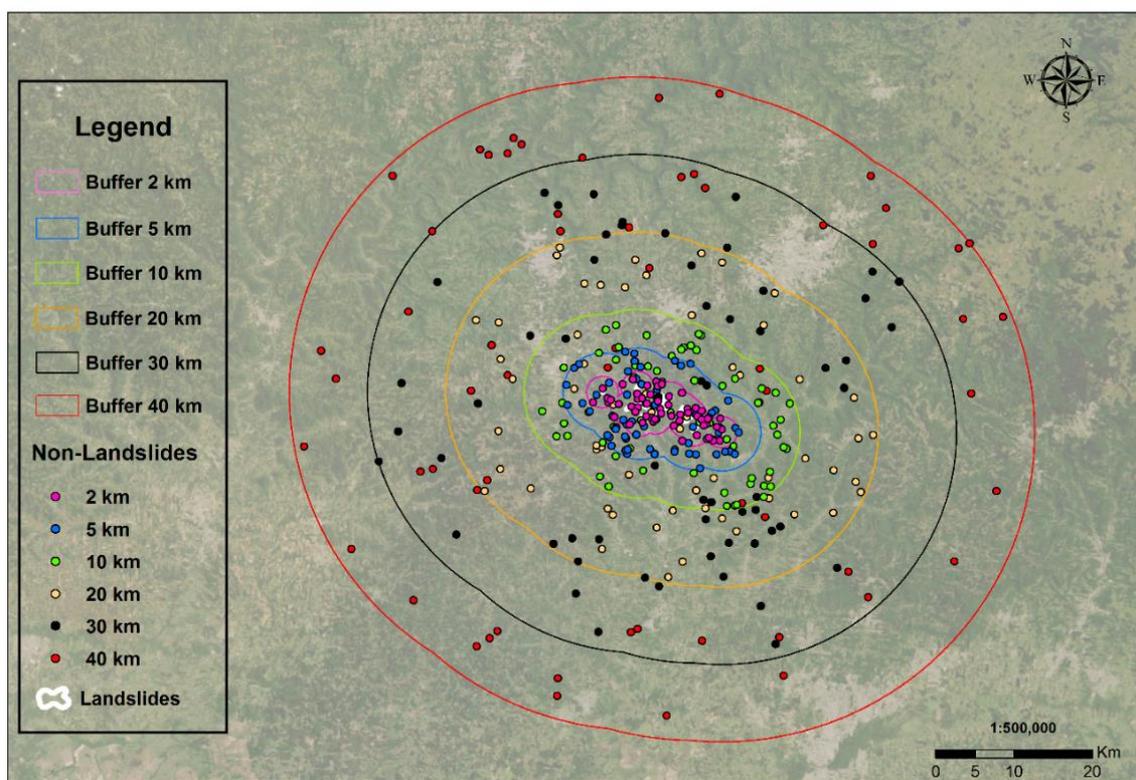


Figura 5.2 Exemplo de distâncias de *buffers* utilizados para amostragem de não ocorrência.

5.2.4 Fatores condicionantes de deslizamento

De acordo com Jebur *et al.* (2014), não há uma regra ou uma estrutura específica para seleção e definição da importância dos fatores condicionantes. Foram utilizados 16 atributos morfométricos de relevo (Figura 5.3), extraídos através da ferramenta Basic Terrain Analysis, no software SAGA GIS 7.0: elevation (ELV), slope (SLP), Analytical Hillshading (AHS), aspect (ASP), channel network base level (CNBL), vertical distance to channel network (VDCN), longitudinal curvature (LOC), cross-section curvature (CSC), flow accumulation (FAC), topographic wetness index (TWI), topographic position index (TPI), convergence index (CIN), relative slope position (RSP), mass balance index (MBI), valley depth (VDP) e LS factor (LSF).

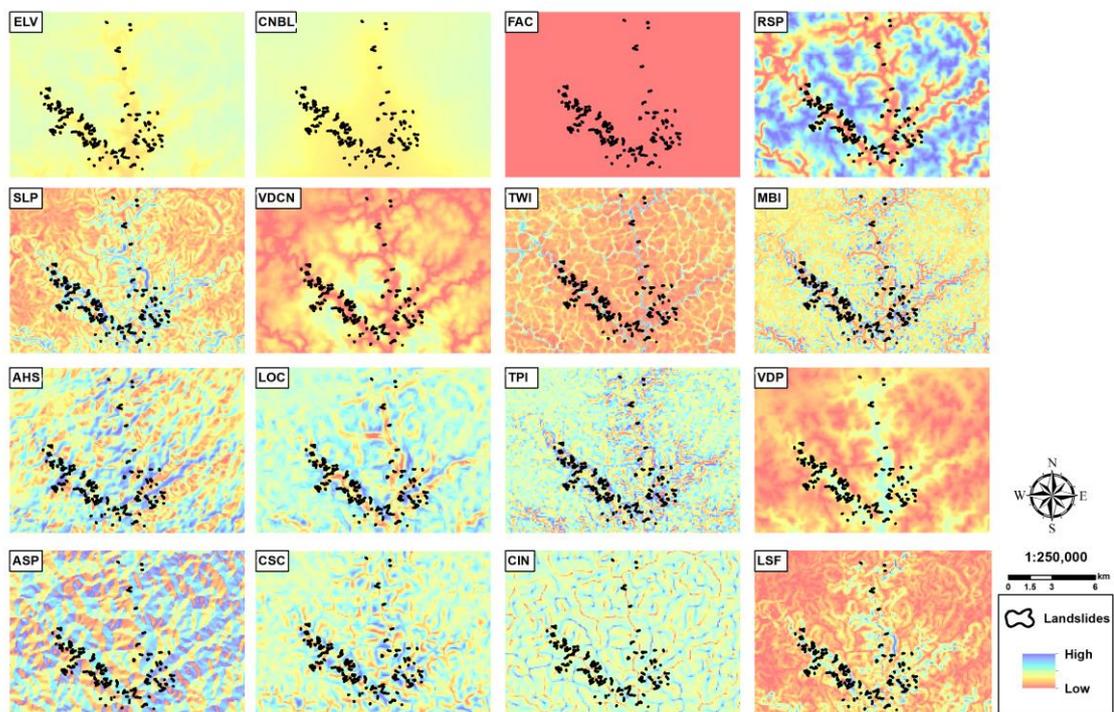


Figura 5.3 Atributos morfométricos de terreno utilizados como entrada na modelagem.

Atributos como ELV, SLP, ASP e TWI já são amplamente usados para modelagens de deslizamentos (e.g. Jebur *et al.*, 2014; Chen *et al.*, 2017b; Hong *et al.*, 2018). Os outros atributos também são fatores que podem influenciar significativamente os modelos por se tratarem de características importantes do relevo. Esses atributos, assim como as cicatrizes, foram gerados através de uma modelo digital de elevação com resolução de 90 m, padronizando assim todos os dados utilizados para a modelagem. A tabela 5.2 mostra a estatística descritiva básica de cada atributo, contendo valores mínimos, máximos, médios e desvios padrão.

<i>ATTRIBUTES</i>	<i>MIN</i>	<i>MEAN</i>	<i>MAX</i>	<i>STD DEV.</i>
ELV	-1.0	452.6	1804	355.7
SLP	0.0	5.9	62.5	6.7
AHS	0.0	0.8	2.1	0.2
ASP	0.0	179.3	360	99.7
CNBL	-1.0	398.5	1540	336.6
VDCN	0.0	54.1	880.7	77.6
LOC	-0.01	-0.00	0.04	0.0
CSC	-0.01	-0.00	0.01	0.0
FAC	7654	12577	25014	2300
TWI	3.5	9.6	27.0	3.8
TPI	-61.6	-1.8	47	2.3
CIN	-98.7	0.0	99.1	13.6
RSP	0.0	0.3	1	0.3
MBI	-0.67	0.05	2.16	0.15
VDP	0.0	113.2	755.1	99.6
LSF	0.0	1.8	133.4	2.7

Tabela 5.2 Valores estatísticos dos atributos morfométricos de terreno.

Após a confecção desses atributos, muitos estudos realizam uma análise para saber como os mesmos se relacionam com as variáveis de entrada do modelo (amostras de ocorrência e não ocorrência), utilizando métodos como correlação de Pearson (Bui *et al.*, 2016; Wang *et al.*, 2020; Lucchese *et al.*, 2020). A fim de agilizar e melhorar o processo de modelagem foi utilizado a análise de correlação de Pearson para identificar os atributos menos correlacionados com as amostras de ocorrência e não ocorrência e excluí-los do processo de modelagem.

5.2.5 Modelagem e validação

As RNA vêm sendo amplamente utilizadas para a modelagem de deslizamentos. Vários estudos revelam a utilidade e eficiência dessas redes no mapeamento de suscetibilidade (Chen *et al.*, 2017a; Kalantar *et al.*, 2018; Oliveira *et al.*, 2019a).

No processo de modelagem foi utilizado um script de RNA, no software MATLAB, com treinamento realizado pelo método retropropagativo com múltiplas camadas (Rumelhart; Hinton; Williams, 1986). A fim de evitar o superajustamento do modelo, foi utilizada uma série de validação cruzada paralela à série de treinamento. As proporções de amostras correspondem a 50%/25%/25%, treinamento, validação cruzada e teste, respectivamente.

O processo de treinamento e validação cruzada foi realizado em 2 etapas. Na primeira etapa, em cada rodada da modelagem, as amostras foram separadas para cada uma das cinco áreas amostrais e cada um dos seis *buffers*. O treinamento das RNA considerou cada área

amostral e seus *buffers* individualmente, gerando um total de 150 modelos, mostrando como cada área amostral se comporta quando extrapolada para toda a região em estudo. Na etapa seguinte, as amostras das cinco áreas foram utilizadas conjuntamente no processo de treinamento, ampliando a variabilidade dos parâmetros utilizados. Nessa etapa, foram gerados 30 distintos modelos, referentes apenas aos seus respectivos *buffers*, mostrando o comportamento de cada um quando utilizados com amostras de não ocorrência em diferentes distancias de coleta.

Para a configuração inicial, o número de neurônios na camada oculta da rede seguiu a fórmula $n-2$, em que n é o número de variáveis de entrada utilizadas na modelagem. Foi testado o aumento do número de neurônios de forma automatizada até um modelo com $n+6$ neurônios na camada oculta, de modo a avaliar o aumento de desempenho dos modelos em função da maior complexidade do modelo e seus pesos sinápticos. Estabeleceu-se como número máximo de ciclos de aprendizagem de 10.000 por iteração. Devido à inicialização de pesos aleatória, executou-se 5 iterações em cada configuração do modelo.

No processo de validação, a acurácia foi calculada através da curva ROC, a qual de acordo com Jaafari *et al.* (2019) é o procedimento científico mais popular para validar a eficiência de predição de modelos, sendo utilizado para estudos de inundações (Hong *et al.*, 2018a; Termeh *et al.*, 2018) incêndios (Nami *et al.*, 2018;) e deslizamentos (Chen *et al.*, 2019; Oliveira *et al.*, 2019a). A curva ROC é uma análise baseada na distribuição final de um método de classificação que diferencia previsões corretas e falhas (Cantarino *et al.*, 2019), considerada como a representação bidimensional de desempenho de um modelo.

Utilizou-se ainda o índice de contribuição relativa (RCI), proposto por Oliveira *et al.* (2015) o qual nos dá os valores de importância atribuídos pela RNA para cada uma das variáveis independentes. Esse índice é calculado através da divisão do índice de contribuição (IC) pela soma do IC sobre todas as entradas do modelo e é um método para avaliar e definir variáveis mais e menos importantes para cada caso.

5.3. RESULTADOS E DISCUSSÕES

5.3.1 Inventário de cicatrizes e morfometria das áreas amostrais

O mapeamento de cicatrizes de deslizamentos ocorrido na Serra Geral do Sul do Brasil resultou em 1210 cicatrizes, sendo 121 no vale do Fão, 53 no vale do Forromeco, 338 em Rolante, 229 em Maquiné e 469 em Timbé do Sul. A tabela 5.3 mostra as áreas médias atingidas por cicatrizes e os valores dos atributos morfométricos mais importantes.

Quando transformadas em amostras de ocorrência, foram encontradas 1638 amostras. Essa mesma quantidade de amostras foi gerada para serem usadas como amostras de não ocorrência, resultando em 3276 amostras. Como foram gerados seis diferentes *buffers* de distância e em cada um foi gerado o mesmo montante de amostras de não ocorrência, o trabalho total resultou em 9828 amostras de não ocorrência.

	<i>OCCURRENCE</i>					
	<i>Mean Area (ha)</i>	<i>Máx Area (ha)</i>	ELV	SLP	LSF	VDP
FÃO	0.22	1.0	477.35	22.10	9.10	184.55
FORROMEÇO	0.23	1.2	437.08	23.57	9.49	210.01
ROLANTE	0.34	4.0	590.13	26.52	11.49	307.93
MAQUINÉ	0.21	2.5	559.25	28.44	12.48	266.54
TIMBÉ DO SUL	0.56	6.1	692.26	30.27	11.87	248.91

Tabela 5.3 Descrição das áreas e atributos morfométricos mais importantes nas amostras de cicatrizes

Nas regiões de Timbé do Sul e Rolante ocorreram os eventos de precipitação mais extremos, tanto devido ao número de cicatrizes presentes quanto a sua dimensão. Em Timbé do Sul, a área média das cicatrizes é de 0,56 ha e a maior atinge 6,1 ha. Em Rolante, esses valores são de 0,34 e 4,0 ha, respectivamente. Fão, Forromeco e Maquiné obtiveram médias semelhantes, entre 0,21 e 0,23 ha, destacando apenas que em Maquiné a maior das cicatrizes ocupa 2,5 ha, enquanto em Fão e Forromeco atingiu 1,0 e 1,2 ha.

Além dos eventos extremos em si, os valores de ELV e SLP também influenciaram na quantidade e dimensões das cicatrizes. As regiões de Timbé, Maquiné e Rolante, que possuem maiores quantidades e tamanho de cicatrizes, são também as que possuem grande distinção em sua ELV e SLP, quando comparadas com Fão e Forromeco. Nos 3 primeiros, a ELV variou de 559 até 692 m, e a SLP acima de 26.5°. Nos outros 2, a ELV não superou os 480 m e as declividades médias foram de 23.5°. Isso demonstra que a morfometria do relevo está intimamente ligada aos deslizamentos, tanto quanto a intensidade do evento de precipitação que possa ter ocorrido.

As médias dos atributos (tabela 5.4) são fatores essenciais para entender o resultado final do modelo pois explicam a grande diferença entre eles, de acordo com os valores usados para treinamento.

O atributo SLP, para o modelo do Fã, apresentou valor médio de 22°, enquanto o modelo total obteve uma média de 25°. Isso pode explicar a razão do modelo do Fã possuir mais áreas suscetíveis nas regiões de declive, pois ele considerou declividades menores como mais suscetíveis a haver deslizamentos, enquanto o modelo total, com média maior, restringiu o modelo a mapear apenas regiões com declividades médias de 25. Apesar de a declividade ser um dos fatores mais importantes para modelagem de suscetibilidade a deslizamentos (Pradhan & Lee, 2010), e Meten *et al.* (2015) afirmarem que quanto maior a declividade, maior será a suscetibilidade de movimentação de material nas vertentes, deve-se tomar cuidado ao analisar os seus valores. Altas declividades nem sempre são um indício de área suscetível a deslizamentos.

	<i>BEST SINGLE MODEL</i>	<i>BEST FULL MODEL</i>
	<i>Mean</i>	<i>Mean</i>
SLP	22,10	25,87
AHS	1,13	1,16
CSC	-0,00	-0,00
LOC	-0,00	-0,00
LSF	9,18	10,71
TPI	-3,37	-2,55
TWI	7,31	6,99
VDP	184,55	246,79
VDCN	63,02	160,33
ELV	477,35	547,26
CIN	-4,25	-2,75
RSP	0,26	0,38
MBI	0,01	0,09

Tabela 5.4 Médias dos atributos em cada um dos melhores modelos para as amostras de ocorrência.

O LSF teve o mesmo comportamento do SLP. Isso se deve ao fato de serem 2 atributos semelhantes de terreno, sendo que a declividade é utilizada no cálculo do LSF. Para Quevedo *et al.* (2019) devido a SLP e LSF serem altamente correlacionados, possuindo informações comum entre eles, ambos tendem a agir da mesma forma no processo de modelagem e muitas vezes, quando um desses fatores obtém uma elevada importância, o outro terá uma relevância menor em todo o processo.

Cabe ressaltar ainda o atributo VDP, que também obteve maiores valores médios no *full model*. Apesar de ter sido apenas o quinto atributo mais importante na escala da RNA, para o modelo de amostras totais esse atributo foi o terceiro de maior influência, ficando abaixo somente de SLP e ELV. Ele demonstra que profundidades de vale menores também são de significativa importância nos processos de deslizamentos.

A ELV, assim como SLP, é um dos mais eficazes atributos para esse tipo de modelagem (Kawabata & Bandibas, 2009; Chen *et al.*, 2017a). Assim como o SLP, a ELV teve vital importância na confecção desses modelos. O modelo do Fão apresentou médias mais baixas desse atributo, variando em torno de 477 m e classificando como suscetíveis regiões mais baixas, o inverso do que ocorreu no modelo total, onde as elevações foram de 547 m. Essa variação fica clara no *single model*, onde em certas regiões da área de estudo foram amplamente classificadas como suscetíveis, mesmo tendo características de altitudes moderadas, baixas declividades e amplas redes de drenagem.

5.3.2 Seleção dos atributos

Os atributos foram analisados através da correlação de Person (tabela 5.5), sendo que os resultados variam de -1 a 1. Quanto mais próximo do 0, menor a correlação entre o atributo e a ocorrência de deslizamentos, ou seja, menor é a sua influência no processo de deslizamento. Com isso, estabeleceu-se valores entre -0,8 e +0,8 como limiar de escolha entre os atributos, resultando que os valores de correlação entre -0,8 e +0,8, foram excluídos do processo de modelagem.

	<i>OCURRENCE</i>						
	2km	5km	10km	20km	30km	40km	Média
SLP	0,49	0,60	0,64	0,69	0,73	0,75	0,65

AHS	0,19	0,24	0,28	0,32	0,33	0,37	0,25
ASP	-0,04	-0,06	-0,07	-0,08	-0,06	-0,07	-0,06
CNBL	-0,16	-0,20	-0,17	-0,15	-0,10	-0,05	-0,14
CSC	-0,13	-0,15	-0,15	-0,18	-0,18	-0,17	-0,16
LOC	-0,22	-0,23	-0,23	-0,26	-0,24	-0,24	-0,23
LSF	0,49	0,60	0,64	0,70	0,73	0,74	0,65
TPI	-0,17	-0,18	-0,17	-0,21	-0,19	-0,19	-0,19
TWI	-0,14	-0,22	-0,28	-0,30	-0,36	-0,39	-0,28
VDP	0,25	0,30	0,32	0,38	0,36	0,36	0,33
VDCN	-0,09	0,09	0,19	0,27	0,38	0,42	0,21
ELV	-0,18	-0,14	-0,05	-0,01	0,07	0,14	-0,03
FAC	-0,02	-0,03	-0,04	-0,04	-0,05	-0,03	-0,04
CIN	-0,10	-0,10	-0,08	-0,11	-0,11	-0,10	-0,10
RSP	-0,24	-0,17	-0,08	-0,03	0,06	0,11	-0,06
MBI	-0,04	0,01	0,06	0,04	0,10	0,11	0,05

Tabela 5.5 Coeficientes de correlação linear das variáveis independentes com a variável dependente do modelo (ocorrência ou não ocorrência de deslizamentos), por *buffer*.

Dentre os 16 atributos gerados, 2 deles se destacam positivamente, tendo os maiores resultados de correlação e para todos os *buffers*, e 2 se destacam negativamente, não tendo praticamente nenhuma relação com os deslizamentos.

Os atributos mais correlacionados foram SLP e LSF, tendo resultados praticamente iguais em todas as análises e atingindo uma correlação média de 0,65, resultados semelhantes ao encontrados por Gameiro *et al.* (2019) e Lucchese *et al.* (2020). Cabe destacar também que os atributos AHS, CSC, LOC, TPI, TWI, VDP, VDCN e CIN estiveram presentes em todos os processos de modelagem, por resultarem em valores de correlação sempre superiores aos limites pré-estabelecidos, indicando que são ainda atributos a serem considerados nas análises de suscetibilidade a deslizamentos.

Atributos como FAC, ASP e MBI não se mostraram de grande valia para a análise de suscetibilidade, visto que não foram considerados como correlacionados na maioria dos modelos, sendo apenas o MBI utilizado nos modelos com distâncias de 30 e 40 km. A ELV,

atributo utilizado em praticamente todos os estudos de suscetibilidade a deslizamentos, apresentou variações conforme os *buffers* utilizados. Este atributo se mostrou eficaz ao ser utilizado para pequenas áreas (2 e 5 km) e também para grandes áreas (40 km) mas não se mostrou eficaz quando utilizado em distâncias moderadas. Isso está relacionado em grande parte a grande variabilidade altimétrica presente em algumas regiões da Serra Geral, mesmo que em espaços reduzidos, o que faz com que a amplitude dos valores de não ocorrência varie muito em distintas regiões.

5.3.4 Análise e comparação dos modelos

Para comparação dos modelos, foi confeccionada a matriz de confusão, a qual indica os valores dos verdadeiros positivos (VP), falsos positivos (FP), falsos negativos (FN) e verdadeiros negativos (VN). Os VP mostram o valor do número de pixels que foram classificados corretamente como suscetíveis enquanto FP são os pixels classificados erroneamente como suscetíveis. VN é o número de pixels classificados corretamente como não suscetível e FN aqueles classificados incorretamente como não suscetível. A tabela 5.6 mostra os valores de VP, FP, FN e VN para todos os modelos com amostrais totais e também o índice de acurácia global e o índice de *recall*. A tabela 5.7 fornece as mesmas informações, só que para os melhores modelos de cada uma das áreas em particular. Através da matriz de confusão, foi possível calcular também a AUC de cada um dos modelos.

<i>MODEL</i>	VP (%)	FP (%)	FN (%)	VN (%)	<i>GLOBAL ACCURACY</i>	<i>RECALL</i>
2 km full	36,69	13,31	8,36	41,64	0,783	0,814
5 km full	39,26	10,74	5,40	44,60	0,838	0,879
10 km full	41,61	8,39	4,43	45,57	0,871	0,903
20 km full	43,35	6,65	3,54	46,46	0,898	0,924
30 km full	43,47	6,53	1,71	48,29	0,917	0,962
40 km full	44,84	5,16	1,43	48,57	0,934	0,968

Tabela 5.6 Indicadores de acurácia global e erros de omissão e comissão para os modelos com amostras totais. VP = verdadeiro positivo; FP = falso positivo; FN = falso negativo; VN = verdadeiro negativo.

<i>MODEL</i>	VP (%)	FP (%)	FN (%)	VN (%)	<i>GLOBAL ACCURACY</i>	<i>RECALL</i>
Maquiné	42,98	6,93	4,80	45,29	0,882	0,899
Rolante	41,88	8,05	6,57	43,49	0,853	0,864
Forromeco	41,48	8,41	5,97	44,14	0,850	0,874
Fão	42,26	7,64	1,98	48,11	0,903	0,955

Timbé	46,17	3,83	3,56	46,44	0,926	0,928
--------------	-------	------	------	-------	-------	-------

Tabela 5.7 Indicadores de acurácia global e erros de omissão e comissão para os melhores modelos com amostras de cara área. VP = verdadeiro positivo; FP = falso positivo; FN = falso negativo; VN = verdadeiro negativo.

Nota-se que os dois melhores modelos foram os que tiveram valores mais baixos de FN, o que significa que foram os modelos que tiveram menor quantidade de pixels classificados incorretamente como não suscetível. Esse dado é de extrema relevância pois mostra através de valores a quantidade de pixels que são suscetíveis, mas que o modelo acabou classificando como não suscetível. No melhor modelo com amostras específicas, denominado de *single model* e que corresponde ao modelo Fão 40 km, a taxa de FN foi de 1,98%, ou seja, menos de 2% dos pixels da imagem foram classificados erroneamente como não suscetíveis. Para o melhor modelo com amostras totais, denominado de *full model*, esse valor foi menor ainda, atingindo 1,43% dos pixels da imagem. Ressalta-se ainda que essa porcentagem de pixels errados apresentou queda significativa conforme se aumentou o *buffer* de distância, mostrando que áreas de coleta mais distância favorecem a confecção de melhores modelos. Cabe salientar também que apesar da distância influenciar positivamente nos valores de pixels corretos, as amostras de ocorrência também possuem influência, como no caso de Rolante, que mesmo com um *buffer* de 40 km, acarretou em um modelo com muitos FN e FP, devido à pouca representatividade que suas amostras de ocorrência tem quando comparadas com as outras regiões e com toda a área da Serra Geral em análise.

Percebe-se também um aumento gradativo da acurácia global e do recall conforme se aumentou a distância da coleta das amostras de não ocorrência, demonstrando que quanto mais distante a coleta é feita, mais representativas são as amostras e modelos mais precisos são gerados. Os maiores valores desses índices foram para o *full model*, atingindo 0,934 e 0,968, respectivamente. Nos modelos com amostras de cada área, o *single model* foi o que obteve o maior *recall*, com 0,955, entretanto, quando se analisa a acurácia global, o modelo de Timbé foi o que obteve maior valor.

Para cada modelo, foram realizadas 5 iterações, assim cada região de *buffer* gerou 5 distintos modelos, num total de 30 para cada área amostral e 30 para a área total. A cada nova iteração, a inicialização do modelo era feita de forma diferente, gerando assim distintos resultados. Para padronização dos modelos, foi calculada a média das acurácias das iterações, o que resultou em uma acurácia geral para cada modelo.

Percebe-se uma influência marcante da distância no desempenho dos modelos, tanto visualmente quanto estatisticamente. Na análise da curva ROC, as acurácias médias dos modelos (Tabela 5.8) de amostras totais variaram de 0,847 (modelo 2 km) até 0,954 (modelo 40 km), enquanto com o uso de amostras separadas, os valores ficaram entre 0,872 (Timbé – 40 km) e 0,931 (Fão – 40 km). Isso ressalta a qualidade de ambas as técnicas no processo de modelagem, mas evidencia que acurácias melhores podem ser obtidas quando se utiliza um conjunto amostral com todas as áreas juntas, além de serem melhor distinguidos e analisados visualmente, gerando mapas com menos ruídos e mais acertos. Apesar disso, o resultado com amostras do Fão se mostrou bem semelhante, estatisticamente, com os resultados quando se utilizou todas as amostras. Isso indica uma boa capacidade de extrapolação espacial das RNA quando os domínios dos valores de entrada dos atributos morfométricos são respeitados, ou seja, a extrapolação ocorre de maneira eficiente em áreas com características geológicas e geomorfológicas mais homogêneas.

Salienta-se que os melhores modelos utilizaram amostras de não ocorrência mais distante, assim como também ocorreu em estudos de Oliveira *et al.* (2019b). Isso indica que quanto mais distante o *buffer*, mais facilidade a RNA terá para discernir entre áreas suscetíveis e não suscetíveis, aumentando assim a acurácia dos modelos.

BUFFER	ÁREAS DA SERRA GERAL					
	Maquiné	Rolante	Forromeco	Fão	Timbé	Área Total
2 KM	0,788	0,759	0,813	0,81	0,791	0,847
5 KM	0,879	0,817	0,868	0,861	0,848	0,902
10 KM	0,912	0,739	0,874	0,896	0,871	0,932
20 KM	0,916	0,824	0,874	0,897	0,861	0,947
30 KM	0,926	0,786	0,919	0,919	0,872	0,942
40 KM	0,912	0,873	0,907	0,931	0,872	0,954

Tabela 5.8 Acurácias médias dos modelos, considerando apenas as amostras de teste, separados por área de treinamento e *buffer* de amostras de não ocorrência.

Conhecendo o melhor modelo de cada área, realizou-se a quantificação (em km²) das regiões com alta suscetibilidade a deslizamento (Tabela 5.9). Através disso, nota-se que os modelos de Forromeco e Fão foram altamente abrangentes, classificando como suscetível 11.647 km² e 13.191 km² respectivamente. Enquanto isso o modelo mais restritivo de todos foi

utilizando amostras de Timbé, com 2.634 km² de área suscetível. Os modelos de Maquiné, Rolante e com amostras totais foram os intermediários, com áreas suscetíveis de 5.324, 8.968 e 5.596 km², respectivamente. Apesar de Rolante ficar mais próximo da média geral entre todos os modelos, a falta de representatividade de suas amostras fez com que o modelo não fosse eficaz, tanto com relação a acurácia quanto espacialmente. Ressalta-se ainda que o modelo com amostras totais, o qual atingiu a maior acurácia, foi um modelo restritivo, limitando as áreas suscetíveis principalmente a metade superior das escarpas da Serra Geral, local onde a maioria dos deslizamentos realmente acontecem.

<i>MODEL</i>	Maquiné	Rolante	Forromeco	Fão	Timbé	Total
Área (km²)	5324	8968	11647	13191	2634	5596

Tabela 5.9. Área com elevada suscetibilidade a deslizamento para cada um dos melhores modelos.

A Figura 5.4 mostra os melhores modelos de cada *buffer* nas respectivas áreas onde foram extraídas as amostras de ocorrência. Nota-se uma melhor eficiência visual nos modelos com *buffers* mais distantes, devido ao fato de serem mais abrangentes na questão dos valores de atributos utilizados, tendo uma maior variação nesses valores e, conseqüentemente, facilitando a distinção entre as áreas suscetíveis pelas RNA. Ressalta-se que esses modelos com *buffers* mais distantes apresentam menos “ruídos” em pequenas áreas próximas a escarpa da Serra Geral, são menos restritivos e evidenciam um maior grau de suscetibilidade nas regiões onde há o registro de deslizamentos.

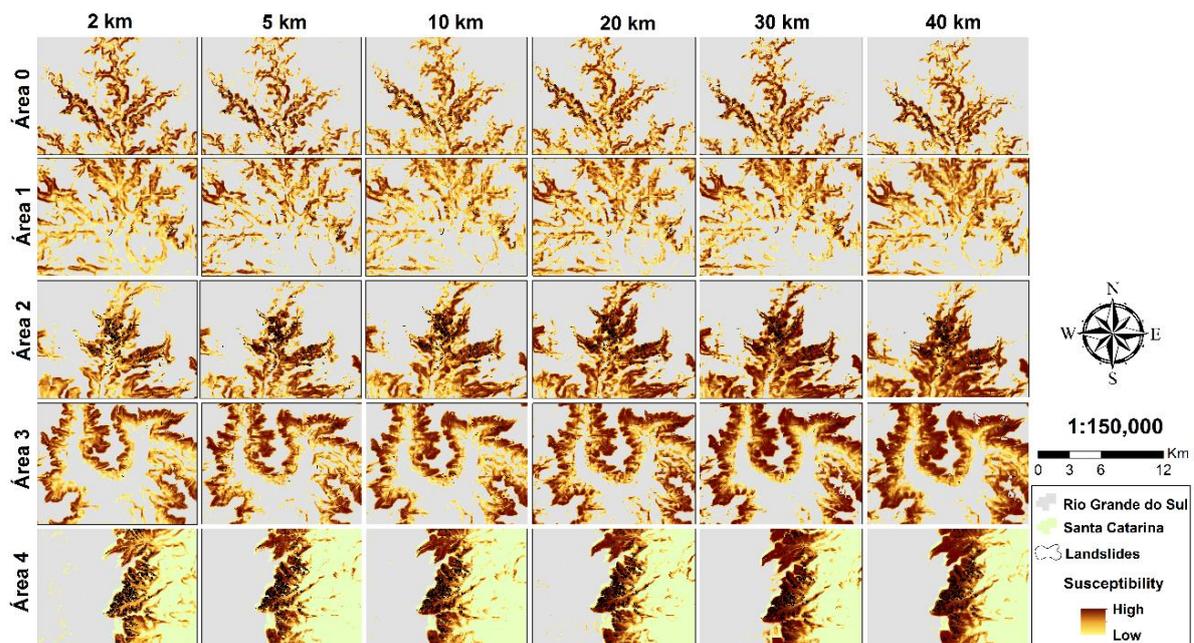


Figura 5.4 Comparação entre os modelos que utilizaram amostras totais.

A Figura 5.5 compara os 2 melhores modelos de cada técnica utilizada e destaca as 5 áreas de deslizamentos utilizadas para as amostragens. Percebe-se claramente o efeito visual em cada um dos modelos gerados. O *single model*, apresenta um mapa mais ruidoso com relação a suscetibilidade, registrando como suscetíveis regiões com baixa ou nenhuma presença de deslizamentos. Enquanto que o *full model*, mostra um mapa mais limpo e mais restritivo, se comparado ao com amostras do Fão. Isso se deve, em grande parte, aos valores dos atributos que foram utilizados para confecção desses modelos, conforme mostrado na Tabela 5.5.

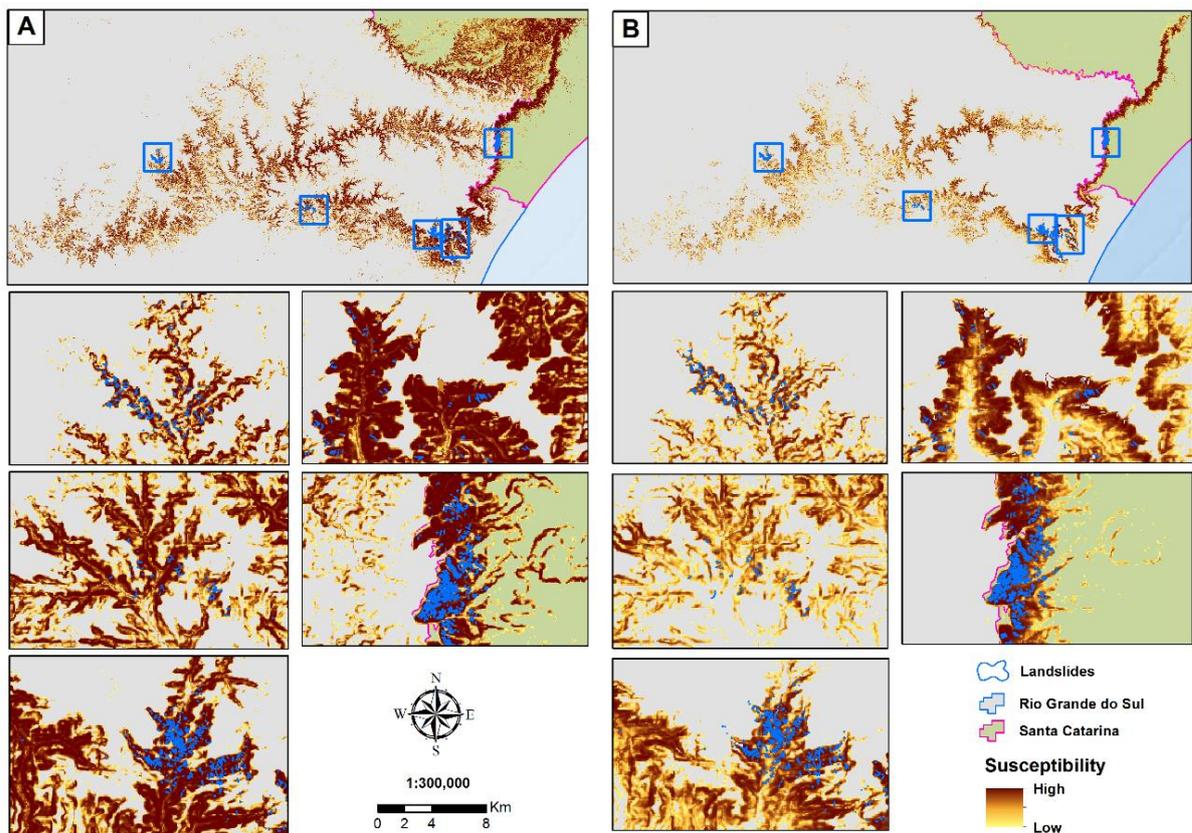


Figura 5.5 Comparação das áreas de cicatrizes entre os melhores modelos: A) melhor modelo com amostras separadas, *single model*; B) melhor modelo com amostras conjuntas, *full model*.

Ainda para analisar as semelhanças e diferenças entre os dois melhores modelos, foi realizada uma álgebra de mapas a fim de encontrar as áreas que mais se sobressaíram em cada um dos modelos. Foi aplicada a subtração do *full model* pelo *single model* (Figura 5.6). Essa álgebra de mapas resultou em um mapa com valores entre -1 e 1. Quanto maior o valor, maior é a suscetibilidade a deslizamentos de acordo com o *full model*. Quanto menor o valor, maior é a suscetibilidade de acordo com o *single model*. Se os resultados são próximos a 0, significa que a suscetibilidade dos dois modelos foi a mesma.

Foram realizadas ainda 10 aproximações para evidenciar as diferenças entre os modelos, sendo as 5 regiões onde foram encontradas cicatrizes de deslizamento (0, 1, 2, 3, 4 – Fão, Forromeco, Rolante, Maquiné e Timbé, respectivamente) e 5 regiões sem cicatrizes (5, 6, 7, 8, 9) e escolhidas para evidenciar a eficácia dos modelos em diferentes regiões da Serra Geral.

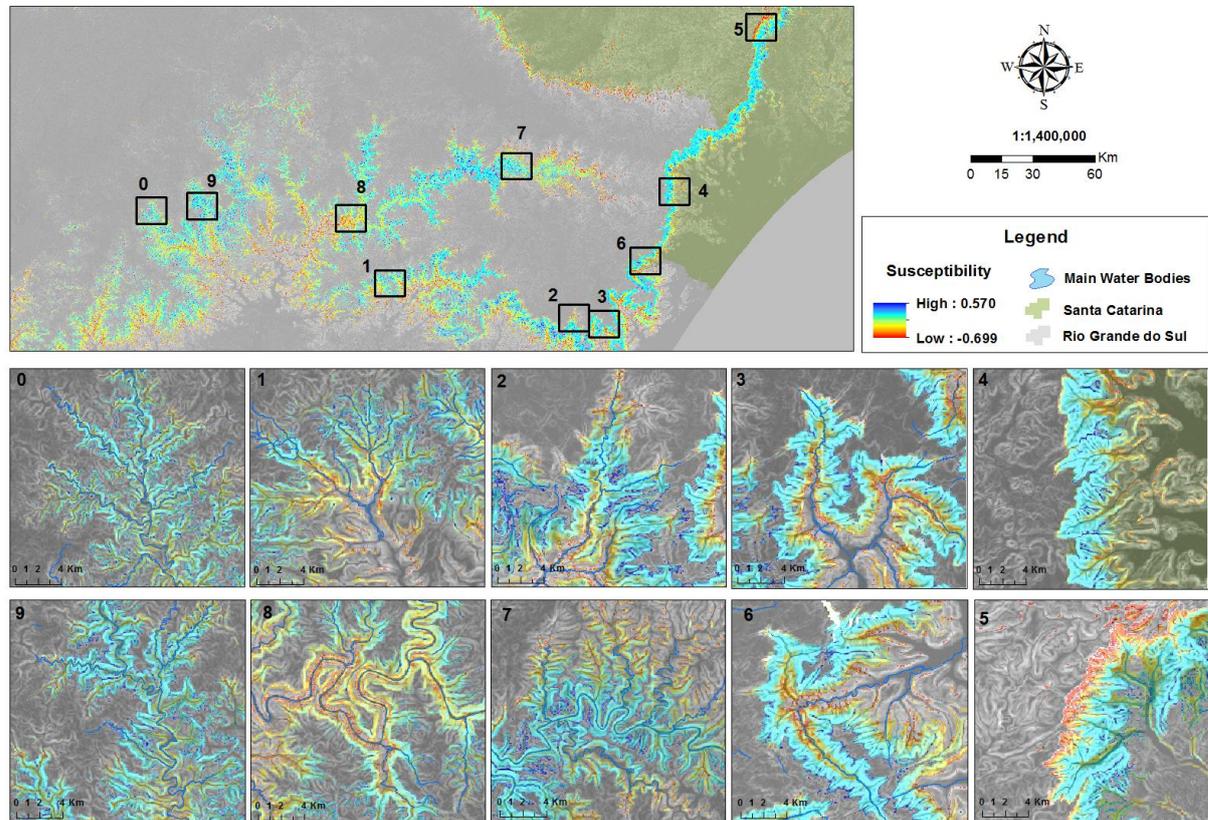


Figura 5.6 Subtração entre o *full model* e o *single model*.

O modelo gerado apresenta maior predominância dos valores positivos, indicando que o *full model* atingiu valores maiores de suscetibilidade em distintas regiões. A presença dos valores positivos tem amplo destaque principalmente nas porções intermediárias da Serra Geral, ressaltando que são áreas muito suscetíveis a deslizamentos. Em áreas de maiores altitudes e baixas declividades (4, 5 e 6) como a escarpa em Santa Catarina, e nos fundos de vales e regiões mais próximas verticalmente à drenagem (2, 3, 4, 7 e 8), caracterizadas por baixas altitudes e declividades, o *single model* mostrou valores de suscetibilidade mais altos.

A razão para essa grande diferença entre os resultados finais dos modelos se deve, principalmente, as amostras utilizadas nos treinamentos. As amostras de ocorrência do *single model*, que utilizou a área do Fão, possui menores declividades (22°), elevações (477 m) e distâncias verticais de drenagem (63 m), fazendo com que áreas com essas características sejam classificadas como mais suscetíveis. Apesar dessas pequenas discrepâncias, ressalta-se a

notável capacidade de extrapolação da RNA ao utilizar apenas amostras de uma área restrita para modelar uma região mais ampla.

No *full model*, utilizando amostras de treinamento de maiores amplitudes e mais representativas de toda a Serra Geral, a RNA conseguiu realizar um treinamento mais eficiente, definindo como suscetíveis áreas já identificadas como suscetíveis a deslizamentos.

Isso comprova que o uso de amostras de distintas regiões na Serra Geral torna processo de modelagem mais complexo, visto que utiliza dados de ocorrência de diferentes áreas amostrais, aumentando assim o poder de diferenciação da RNA para modelar a suscetibilidade a deslizamento.

Salienta-se ainda que há a necessidade de que o conjunto de amostras de treinamento seja representativo de toda a região, possuindo valores de atributos distintos entre todas as áreas e unidades de relevo presentes no seu entorno, aumentando assim a eficácia dos modelos.

5.3.5 Importância dos atributos morfométricos

A partir das RNA foi feita uma avaliação da importância dos atributos, utilizando o cálculo do RCI (Oliveira *et al.*, 2019). A Tabela 5.10 mostra os valores atribuídos para cada atributo dentro do melhor modelo para cada *buffer*.

Atributos	Distâncias de cada <i>buffer</i>					
	2 km	5 km	10 km	20 km	30 km	40 km
SLP	21	32	25	17	10	13
AHS	02	00	04	04	04	01
CNBL	14	13	19	13	21	-
CSC	04	06	06	02	06	00
LOC	01	01	01	02	01	01
LSF	04	02	02	05	07	02
TPI	02	03	04	07	03	03
TWI	05	02	04	03	02	01
VDP	08	03	08	08	07	06
VDCN	02	01	01	01	02	01
ELV	13	15	-	-	-	37

CIN	01	02	01	01	00	01
RSP	02	01	02	-	-	02
MBI	-	-	-	-	01	00

Tabela 5.10 Valores de importância atribuídos pela RNA

Assim como demonstrado na correlação de Pearson, o atributo SLP é um dos fatores principais que influenciam na suscetibilidade a deslizamentos. Entretanto, houve um contraste entre esses valores, visto que conforme o *buffer* de distância aumenta o valor de SLP também aumenta na correlação de Pearson, mas diminui na análise da RNA. CNBL também apresentou mudança nas duas análises, sendo muito valorizado pela RNA, principalmente no *buffer* de 30 km, enquanto na correlação apresentou valores medianos e que diminuía conforme se aumentava a distância.

O atributo LSF também foi muito discrepante nas duas análises. Obteve os maiores valores na correlação, semelhante ao SLP, enquanto que para as RNA se mostrou de pouca relevância, contrastando resultados obtidos por Oliveira *et al.* (2019). A ELV, atributo só utilizado para 2, 5 e 40 km, de acordo com a análise de correlação, foi o atributo de maior importância entre todos quando utilizado no *buffer* de 40 km. Isso corrobora os resultados de Youssef *et al.* (2016), Oliveira *et al.* (2019) e de Quevedo *et al.* (2019), no qual a ELV obteve importância maior do que 30%. Demonstra a importância desse atributo quando se utilizam amostras de não ocorrência mais variadas e que atingem unidades de relevo muito distintas, auxiliando muito a RNA na distinção entre as áreas suscetíveis e não suscetíveis. Os atributos MBI, RSP, CIN e VDCN se mostraram como os de menores influências pela rede, obtendo valores abaixo de 0,02. Apesar disso, a correlação entre as ocorrências e VDCN se mostrou de grande valia, principalmente para maiores distâncias.

Para análise da relação da importância entre os atributos e as áreas, foram gerados 500 pontos aleatórios dentro das 10 áreas (0 – 9), plotadas em um gráfico entre suscetibilidade de cada um dos melhores modelos e os principais atributos encontrados, sendo utilizados LSF, VDP, VDCN, SLP e ELV (Figura 5.7).

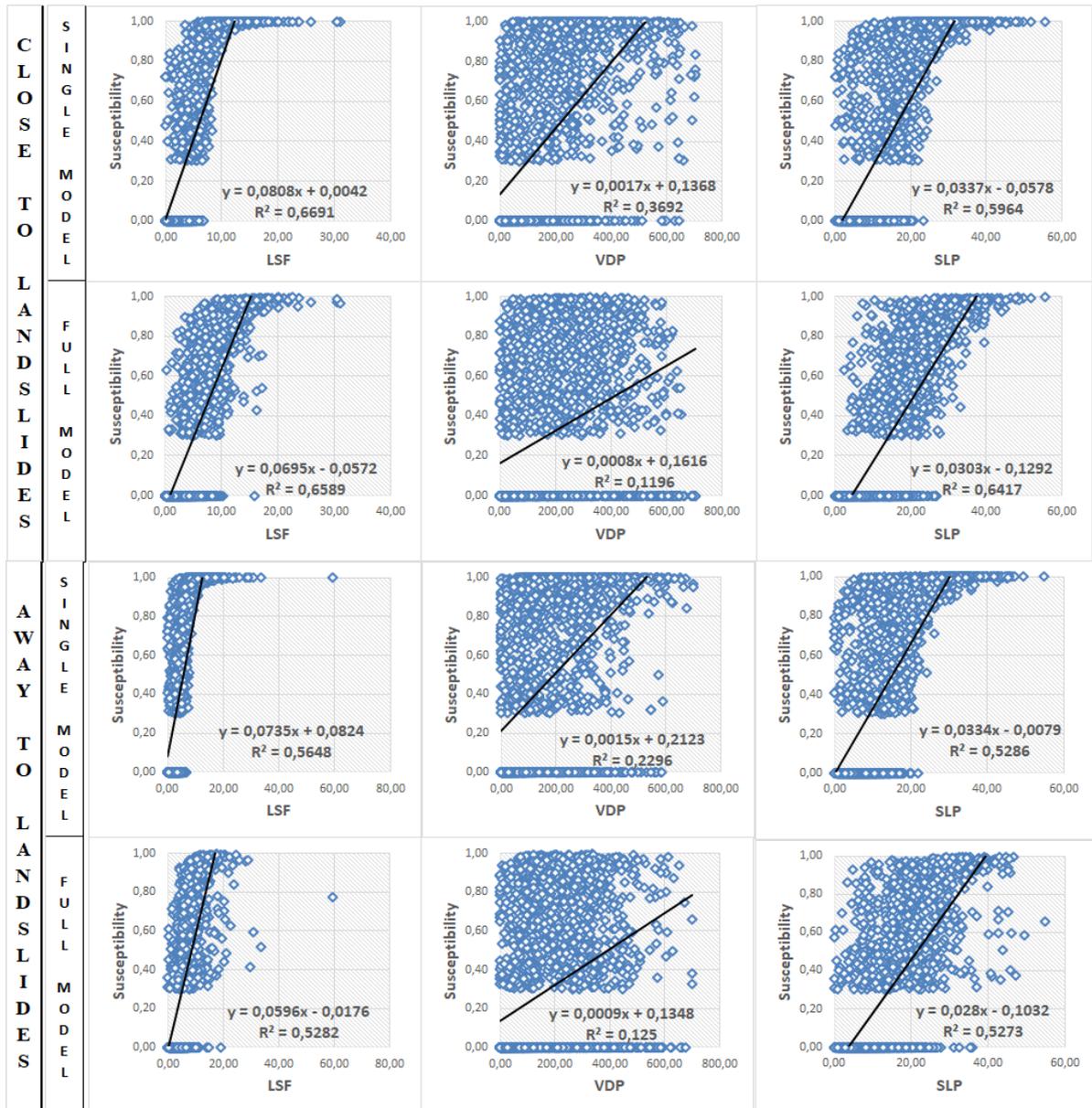


Figura 5.7. Gráficos da relação entre o índice de suscetibilidade a deslizamentos dos melhores modelos e os principais atributos morfométricos.

Os gráficos da relação entre suscetibilidade dos melhores modelos e os principais atributos mostram a relevância desses atributos em uma escala linear e o quanto são representativos no processo de modelagem. Destacam-se os atributos LSF e SLP, com R^2 acima de 0,52, demonstrando uma relação linear crescente em relação a suscetibilidade, ou seja, a suscetibilidade tende a aumentar conforme os valores destes atributos também aumentam. Essa relação se repetiu em ambos os modelos e áreas estudadas, porém o maior valor de LSF (0,66) ocorreu no *single model*, enquanto que para o SLP (0,64) ocorrendo no *full model*, ambos para

áreas próximas aos deslizamentos, reforçando assim a ideia de que ambos atributos possuem destacada importância no processo de modelagem a deslizamentos.

Ainda sobre esses dois atributos, o LSF apresentou maior suscetibilidade a cada aumento de valor, com taxa de aumento média de suscetibilidade de 0,07 para cada aumento do atributo. Com o SLP, essa média de aumento foi na faixa de 0,03 para cada grau de declividade. Assim como os atributos anteriores, o VDP também se mostrou representativo para a modelagem. Apesar de obter valores menores do que SLP e LSF, atingiu valores de relação linear de até 0,36 para o *single model*, demonstrando a sua importância e confirmando resultados de Oliveira et al. (2019), o qual encontrou importâncias de até 24% para este atributo.

Os atributos VDCN e ELV não apresentaram significativa relevância em sua relação linear com a suscetibilidade. Entretanto, quando a ELV foi analisada individualmente, para cada área e modelo, percebe-se a sua importância. A figura 5.8 mostra a análise da ELV por modelo e por área coletadas, sendo as imagens A e B o uso dos pontos de todas as áreas juntas e C e D apenas com os pontos da área 0. Ainda se ressalta que A e C são referentes ao *single model* e B e D referentes ao *full model*.

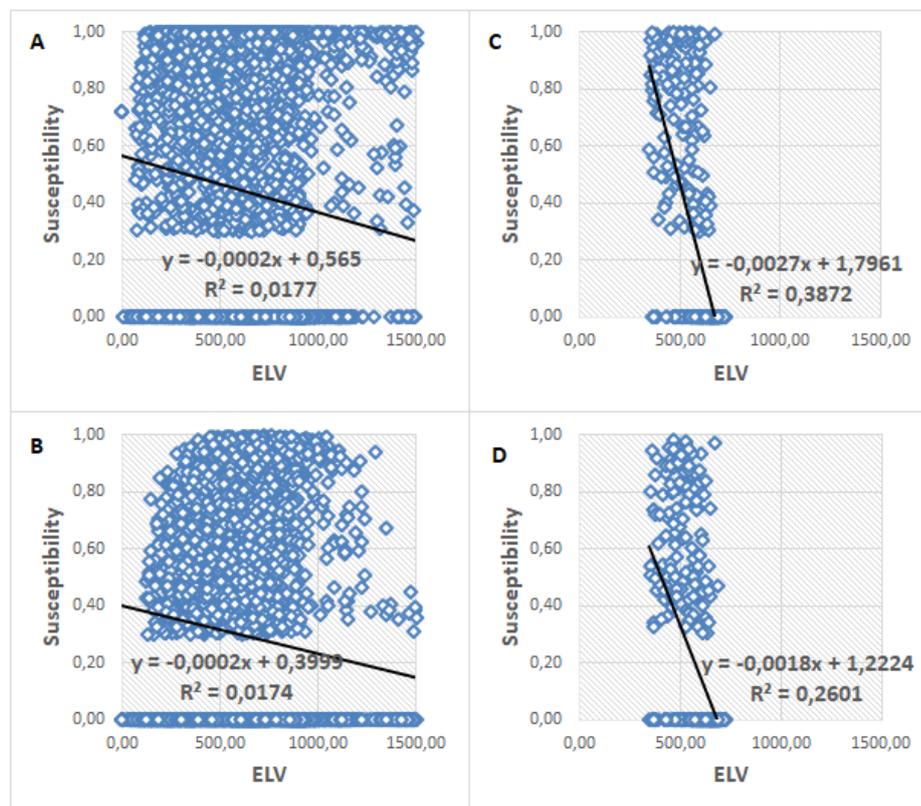


Figura 5.8. Relação entre o índice de suscetibilidade e a elevação para os melhores modelos; A) Uso de todas as amostras para o *single model*; B) Uso de todas as amostras para o *full model*;

C) Uso das amostras da área 0 para o *single model*; D) Uso das amostras da área 0 para o *full model*.

Nota-se que mesmo quando considerados os pontos de todas as áreas juntas, os modelos não mostraram relação linear significativa, com valor de 0,01 para R^2 . Isso acontece devido as amostras utilizadas para treinamento se caracterizarem por grande amplitude de valores, fazendo com que a ELV não se distinguisse no processo de modelagem. Quando se consideram vários eventos de deslizamentos, com cicatrizes distantes no espaço e com índices de precipitação diferentes para a ocorrência desse tipo de movimento de massa, não se constata um padrão de valores típicos de altitude para o ponto de ruptura e deslocamento dos materiais deslizados. A unidade geomorfológica da Serra Geral apresenta um decaimento dos valores altimétricos no sentido leste-oeste, fazendo com que a linha de escarpa erosiva e, conseqüentemente, os pontos de ruptura e zonas de transporte de materiais não apresentem uma altitude preferencial para a ocorrência do fenômeno.

Por outro lado, quando utilizadas apenas amostras do Fão, com um intervalo bem marcado e definido de elevações, a relação linear fica presente, mostrando que para pequenas áreas, o treinamento com amostras simples (de apenas uma área) faz com que a ELV seja um atributo de alta importância, atingindo uma média de $R^2 = 0,32$. Entretanto, não é adequado utilizar esse pequeno grupo de amostras, de uma área específica, para realizar generalizações extrapolações para áreas grandes, devido principalmente a significativas mudanças nos valores dos atributos morfométricos, pois quando estes atributos não são tão representativos, acab por prejudicar o desempenho real da RNA (Nefeslioglu *et al.*, 2008).

5.4 CONCLUSÕES

Percebe-se que as áreas amostrais afetam diretamente no resultado final da modelagem da suscetibilidade a deslizamentos por RNA. A distância da coleta de amostras de não ocorrência é um fator determinante para a confecção de modelos com altas taxas de acurácia, resultando em melhores resultados conforme se aumenta essa distância. O uso de múltiplos eventos também se mostra de grande importância, gerando melhores resultados devido a maior diversificação das amostras usadas para treinamento. E, nota-se ainda que existem atributos morfométricos que são fundamentais para o processo de modelagem. Dentre as principais conclusões, destacam-se:

- i) A ELV e o SLP foram os atributos de terreno que mais influenciaram na suscetibilidade a deslizamentos. Dependendo do conjunto amostral utilizado, o valor de importância dos dois atributos pode ser inverso;
- ii) O aumento da distância no processo de amostragem de não ocorrência influencia positivamente a acurácia dos modelos, passando de 0,84 para 0,95 nas amostras totais e de 0,87 para 0,93 nas amostras do Fão;
- iii) O processo de escolha das amostras de treinamento é de vital importância para o resultado final, e a escolha de amostras mais diversificadas, recolhidas em diversas áreas, se mostrou mais eficiente e com maior acurácia no resultado final quando comparada a amostras de regiões específicas e com características morfométricas semelhantes;
- iv) As RNA apresentam uma boa capacidade de extrapolação espacial, desde que os domínios de valores de entrada dos atributos morfométricos sejam respeitados, ou seja, que a extrapolação ocorra para uma área homogênea em relação a ocorrência de deslizamentos, resultando que uma pequena área de evidências possa obter bons resultados para áreas maiores e mais distantes.

5.5 AGRADECIMENTOS

Esta pesquisa foi financiada pela CAPES (processo 88882.438888/2011-01) e pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) Edital 01/2017 - ARD, processo 17/2551-0000894-4.

5.6 REFERÊNCIAS

AGHDAM, I. N.; VARZANDEH, M. H. M.; PRADHAN, B. Landslide susceptibility mapping using an ensemble statistical index (W_i) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). **Environ Earth Sci**, 75, 1-20, 2016.

BAEZA, C.; LANTADA, N.; MOYA, J. Influence of sample and terrain unit on landslide susceptibility assessment at La Pobla de Lillet, Eastern Pyrenees, Spain. **Environ Earth Sci**. 60(1), 155-167, 2010.

BUI, D. T.; TUAN, T. A.; KLEMPE, H.; PRADHAN, B.; REVHAUG, I. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression and logistic model tree. **Landslides**, 13, 361-378, 2016. DOI: 10.1007/s10346-015-0557-6

CANTARINO, I.; CARRION, M. A.; GOERLICH, F.; IBÁÑEZ, V. M. A ROC analysis-based classification method for landslide susceptibility maps. **Landslides** 16, 265–282, 2019. <http://doi.org/10.1007/s10346-018-1063-4>

CHEN, W.; POURGHASEMI, H. R.; KORNEJADY, A.; ZHANG, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. **Geoderma**, 305, 314-327, 2017a.

CHEN, W.; POURGHASEMI, H. R.; PANAHI, M.; KORNEJADY, A.; WANG, J.; XIE, X.; CAO, S. 2017b. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. **Geomorphology** 297, 69-85, 2017b. Doi: 10.1016/j.geomorph.2017.09.007

CHEN, W.; PANAHI, M.; TSANGARATOS, P.; SHAHABI, H.; ILIA, I.; PANAHI, S.; LI, S.; JAAFARI, A.; AHMAD, B. B. Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. **Catena** 172, 212-231, 2019.

GAMEIRO, S.; QUEVEDO, R. P.; OLIVEIRA, G. G.; RUIZ, L. F. C.; GUASSELLI, L. A. Análise e correlação de atributos morfométricos e sua influência nos deslizamentos ocorridos na Bacia do Rio Rolante, RS. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, XIX, 2019 Santos. **Anais...** 17, 2880-2883, 2019.

GÖRUM, T. Landslide recognition and mapping in a mixed forest environment from airborne LiDAR data. **Engineering Geology**, 258, 2019. Doi:10.1016/j.enggeo.2019.105155

HECKMANN, T.; GEGG, K.; BECHT, M. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flow. **Nat. Hazards Earth Syst. Sci.**, 14, 259–278, 2014.

HONG, H.; PRADHAN, B.; XU, C.; BUI, D. T. Spatial Prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. **Catena**, 133, 266-281, 2015.

HONG, H.; TSANGARATOS, P.; ILIA, I.; LIU, J.; ZHU, A-X.; CHEN, W. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. **Science of the Total Environment**, 625, 575-588, 2018a.

HONG, H.; LIU, J.; BUI, D.T.; PRADHAN, B.; ACHARYA, T. D.; PHAM, B. T.; ZHU, A-X.; CHEN, W.; AHMAD, B. B. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). **Catena** 163, 399-413, 2018b. Doi: 10.1016/j.catena.2018.01.005

HONG, H.; MIAO, Y.; LIU, J.; ZUH, A-X. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslides susceptibility mapping. **Catena** 176, 45064, 2019. DOI:10.1016/j.catena.2018.12.035.

JAAFARI, A.; PANAHI, M.; PHAM, B. T.; SHAHABI, H.; BUI, D. T.; REZAIE, F.; LEE, S. Meta optimization of an adaptive neuro-fuzzy inference system with grey Wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. **Catena** 175, 430-445, 2019. <https://doi.org/10.1016/j.catena.2018.12.033>.

JEBUR, M. N.; PRADHAN, B.; TEHRANY, M. S. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. **Remote Sensing of Environment** 152, 150-165, 2014.

KALANTAR, B.; PRADHAN, B.; NAGHIBI, S. A.; MOTEVALLI, A.; MANSOR, S. Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). **Geomatics, Natural Hazards and Risk**, 9 (1), 49-69, 2018. <https://doi.org/10.1080/19475705.2017.1407368>

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

LUCCHESI, L. V.; OLIVEIRA, G. G.; PEDROLLO, O. C. Attribute selection using correlations and principal components for artificial neural networks employment for landslide susceptibility assessment. **Environmental Monitoring and Assessment**, 192:129, 2020. DOI: 10.1007/s10661-019-7968-0

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**, 2 (9), 1 – 17, 2015.

NAMI, M. H.; JAAFARI, A.; FALLAH, M.; NABIUNI, S. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. **Int. J. Environ. Sci. Technol.** 15 (2), 373–384, 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**, 40, 47-61, 2015.

OLIVEIRA, G. G.; RUIZ, L. F. C.; GUASSELLI, L. A.; HAETINGER, C. Random Forest and artificial neural networks in landslide susceptibility modeling: a case study of the

Fão River Basin, Southern Brazil. **Natural Hazards**, 99, 1049, 2019a. DOI: 10.1007/s11069-010-03795-x

OLIVEIRA, G. G.; RUIZ, L. F. C.; QUEVEDO, D. M. Redes neurais artificiais e geotecnologias para o mapeamento de áreas suscetíveis a inundações. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, XXIII, 2019 Foz do Iguaçu. **Anais...** p. 10, 2019b.

PANDEY, V. K.; SHARMA, M. C. Probabilistic landslide susceptibility mapping along Tipri to Ghuttu highway corridor, Garhwal Himalaya (India). **Remote Sensing Applications: Society and Environment** 8, 1-11, 2017.

POURGHASEMI, H. R.; PRADHAN, B.; GOKCEOGLU, C. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. **Natural Hazards**, 63, 965-996, 2012. DOI: 10.1007/s11069-012-0217-2.

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

QUEVEDO, R. P.; GUASSELLI, L. A.; OLIVEIRA, G. G.; RUIZ, L. F. C. Modelagem de áreas suscetíveis a deslizamentos: avaliação comparativa de técnicas de amostragem, aprendizado de máquina e modelos digitais de elevação. São Paulo, UNESP, **Geociências**, 38 (3), 781 – 795, 2019.

REGMI, A. D.; DEVKOT, K. C.; YOSHIDA, K.; PRADHAN, B.; POURGHASEMI, H. R.; KUMAMOTO, T.; AKGUN, A. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. **Arab. J. Geosci.** 7, 725-742, 2014.

SACHDEVA, S.; BATHIA, T.; VERMA, A. K. A novel voting ensemble model for spatial prediction of Landslides using GIS. **International Journal of Remote Sensing**, 41 (3), 929-952, 2020. DOI: 10.1080/01431161.2019.1654141.

SAMEEN, M. I.; PRADHAN, B.; LEE, S. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. **Catena** 186, 104249, 2020.

SHIRZADI, A.; SOLAIMANI, K.; ROSHAN, M. H.; KAVIAN, A.; CHAPI, K.; SHAHABI, H.; KEESSTRA, S.; AHMAD, B. B.; BUI, D. T. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. **Catena** 178, 172-188, 2019.

TERMEH, S. V. R.; KORNEJADY, A.; POURGHASEMI, H. R.; KEESSTRA, S. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. **Sci. Total Environ.** 615, 438–451, 2018.

VALENZUELA, P.; DOMÍNGUEZ-CUESTA, M. J.; GARCÍA, M. A. M.; JIMÉNEZ-SÁNCHEZ, M. A spatio-temporal landslide inventory for the NW of Spain: BAPA database. **Geomorphology** 293, 11-23, 2017.

VAN DEN EECKHAUT, M.; HERVÁS, J.; JAEDICKE, C.; MALET, J-P.; MONTANARELLA, L.; NADIM, F. Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. **Landslides** 9, 357-369, 2012. DOI: 10.1007/s10346-011-0299-z

YALCIN, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. **Catena** 72, 1-12, 2008.

YILMAZ, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat – Turkey). **Computers & Geosciences**, 35, 1125-1138, 2009. DOI: 10.1016/j.cageo.2008.08.007.

YOUSSEF, A. M.; POURGHASEMI, H. R.; POURTAGHI, Z. S.; AL-KATHEERI, M. M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. **Landslides** 13 (5), 839–856, 2016.

ZHAO, F.; MENG, X.; ZHANG, Y.; CHEN, G.; SU, X.; YUE, D. Landslide susceptibility mapping of Karakorum Highway combined with the application of SBAS-InSAR technology. **Sensors**, 19, 2685, 2019. DOI: 10.3390/s19122685.

WANG, Y.; FANG, Z.; HONG, H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. **Science of Total Environment** 666, 975-993, 2019.

WANG, Y.; FENG, L.; LI, S.; REN, F.; DU, Q. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. **Catena**, 188, 104425, 2020. DOI: 10.1016/j.catena.2019.104425.

CAPÍTULO 6: CONSIDERAÇÕES FINAIS

Constatou-se que as RNA realmente possuem uma boa capacidade de extrapolação espacial. Contudo, salienta-se que essa capacidade varia muito em função do conjunto de amostras que é utilizado no treinamento da rede. No processo de treinamento, o uso de amostras mais variadas, representativas e de múltiplos eventos de deslizamentos, faz com que a extrapolação espacial tenha melhores resultados do que quando se usa amostras de apenas uma área específica. Ressalta-se também que quanto mais distante o conjunto amostral de não ocorrência estiver, mais fácil será para a RNA fazer a distinção entre áreas suscetíveis e não suscetíveis. Conclui-se que a distância e os valores dos conjuntos amostrais exercem influência direta tanto no processo de modelagem quanto na capacidade de extrapolação das redes, trazendo resultados melhores conforme se aumenta a distância e a variabilidade desse conjunto amostral.

O mapeamento de cicatrizes de deslizamentos ocorridos na Serra Geral no Sul do Brasil resultou em um total de 1210 cicatrizes, sendo 121 no vale do Fão, 53 no vale do Forromeco, 338 em Rolante, 229 em Maquiné e 469 em Timbé do Sul. Quando transformadas em amostras de ocorrência, foram geradas 1638 amostras, mesmo quantidade utilizada para gerar as amostras de não ocorrência, resultando 3276 amostras. Foram gerados seis *buffers* de distância e, em cada um foi gerado o mesmo montante de amostras de não ocorrência, resultando num total de 9828 amostras de não ocorrência.

A análise de correlação e os gráficos de *boxplot* dos atributos de terreno permitiram concluir que os atributos que melhor diferenciam as áreas de ocorrência e as áreas de não ocorrência foram SLP, LSF e VDP. Enquanto os piores para realizar essa distinção foram MBI, RSP, FAC, ASP e ELV. Contrastante a isso, a RNA atribuiu um elevado valor para o atributo ELV, mostrando que mesmo com baixa correlação, a ELV ainda é um dos atributos principais para análise da suscetibilidade a deslizamentos. Ressalta-se ainda que atributos como as curvaturas do terreno, índices de umidade e índices topográficos apresentaram ainda valores de relevância para nas modelagens.

O processo de amostragem é um fator de extrema relevância para a obtenção de resultados com altas acurácias. Amostras de áreas mais restritivas, com poucas variações em seus atributos morfométricos se mostraram mais eficientes do que amostras que eram mais distribuídas entre os atributos, evidenciando que os valores das amostras interferem muito na capacidade de generalização e extrapolação dos modelos. Ressalta-se também que quanto mais distintas e bem

distribuídas forem as amostras, melhor será a acurácia do modelo e mais limpa e clara será a visualização dos mapas.

Os *buffers* de distância para amostras de não ocorrência também mostraram ter uma grande influência nos resultados. Quanto mais longe as amostras de não ocorrência foram coletadas, mais representativo e maior foi a acurácia do modelo.

As acurácias obtidas variaram entre 0,759 e 0,954, demonstrando a grande capacidade das RNA em modelar áreas suscetíveis. As acurácias considerando áreas de treinamento isoladas atingiram valores elevados, em áreas com características bem marcadas e atributos com valores mais específicos e distintos do seu entorno.

A maior acurácia foi atingida utilizando para treinamento uma mistura das amostras de todas as áreas com cicatrizes, revelando que a maior variação desses atributos foi mais competente para diferenciar as áreas com maiores e menores valores de suscetibilidade.

A capacidade de generalização e extrapolação das RNA para grandes áreas se mostrou eficiente para a região da Serra Geral no sul do Brasil. Entretanto, cabe ressaltar que o processo de amostragem e treinamento tem influência direta nesses elementos das RNA. Conforme as amostras são mais amplas e representativas, essas capacidades acabam por aumentar e se tornam muito importantes, permitindo que áreas maiores sejam analisadas.

Os resultados permitem concluir que as RNA foram eficazes para mapear áreas suscetíveis a deslizamentos, com acurácias maiores que 0,95. Este estudo se mostra de grande relevância, auxiliando a explicar a influência que as áreas amostrais e as distâncias de onde são feitas as amostras de não ocorrência tem no processo de modelagem como um todo, modificando atributos, na capacidade de generalização e extrapolação, no tempo de processamento, na visualização e nas acurácias dos modelos.

CAPÍTULO 7: REFERÊNCIAS

ADA, M.; SAN, B. T. Comparison of Machine-learning techniques for landslide susceptibility mapping using two-level Random sampling (2LRS) in Alakir catchment área, Antalya, Turkey. **Natural Hazards**, 90, 237–263, 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** 318, 101-111, 2018.

AL-NAJJAR, H. A. H.; KALANTAR, B.; PRADHAN, B.; SAEIDI, V. Conditioning factor determination for mapping and prediction of landslide susceptibility using machine learning algorithms. **Proc. SPIE 11156, Earth Resources and Environmental Remote Sensing/GIS Applications X**, 111560K, 2019. doi: 10.1117/12.2532687.

ARAÚJO, T. B. A. **Redes neurais incorporadas a métodos numéricos para solução de problemas de engenharia estrutural**. Dissertação (Mestrado em Engenharia Civil), Centro de Tecnologia, Universidade Federal de Alagoas, Maceió, p. 91, 2015.

AZAMBUJA, E.; ZENI, C.A.; BRESSANI, A.; NANNI, A.S. **Avaliação dos fluxos de detritos em Alto Feliz no Rio Grande do Sul**. In: 3ª Conferência Brasileira de Estabilidade de Encostas (3º COBRAE). ABMS/ABGE, Anais, 17-24, 2001.

BALLABIO, C.; STERLACCHINI S. Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy. **Mathematical Geosciences**, 44 (1), 47–70, 2012.

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

BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados: novos sistemas sensores métodos inovadores**. São Paulo: Oficina de Textos, 2005.

BRAGAGNOLO, L.; SILVA, R. V. D.; GRZYBOWSKI, J. M. V. Artificial neural network ensembles applied to the mapping of landslide susceptibility. **Catena** 184, 104240, 2020.

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

BUI, D. T.; TUAN, T. A.; KLEMPE, H.; PRADHAN, B.; REVHAUG, I. Spatial prediction models for shallow landslides hazards: a comparative assessment of the efficacy of

support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. **Landslides**, 13, 361-378, 2016. Doi: 10.1007/s10346-015-0557-6

CHEN, W.; POURGHASEMI, H. R.; PANAH, M.; KORNEJADY, A.; WANG, J.; XIE, X.; CAO, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. **Geomorphology** 297, 69-85, 2017a. Doi: 10.1016/j.geomorph.2017.09.007

CHEN, W.; XIE, X.; WANG, J.; PRADHAN, B.; HONG, H.; BUI, D. T.; DUAN, Z.; MA, J. A comparative study of logistic model tree, Random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. **Catena** 151, 147–160, 2017b.

CHEN, W.; REZA, H.; KORNEJADY, A.; ZHANG, N. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* 305, 314–327, 2017c. <https://doi.org/10.1016/j.geoderma.2017.06.020>.

CHERKASSKY, V.; MULIER, F. **Learning from Data: Concepts, Theory, and Methods**. 2nd edition, Wiley-IEEE Press, 2007.

CONFORTI, M.; PASCALE, S.; ROBUSTELLI, G.; SDAO, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena* 113, 236–250, 2014.

GÓMEZ, H.; KAVZOGLU, T. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. **Engineering Geology**, 78 (1-2), 11–27, 2005.

GUIDICINI, G.; NIEBLE, C. M. **Estabilidade de taludes naturais e de escavação**. São Paulo. Edgard Blücher; Ed. da Universidade de São Paulo, p. 194, 1984.

HAYKIN, S. **Neural Networks - A Comprehensive Foundation**. Prentice-Hall, China, 3rd edition, 2007.

HARMENDANI, P.; ZANETTI, M. Aprendizado de máquina por reforço aplicado no jogo de carta uno. **Revista de Sistemas e Computação**, Salvador, 9 (2), 245-251, jul./dez. 2019

HÖFIG, P.; GIASSON, E.; VENDRAME, P. R. S. Mapeamento digital de solos com base na extrapolação de mapas entre áreas fisiograficamente semelhantes. **Pesquisa Agropecuária Brasileira**, Brasília, 49 (12), 958-966, dez. 2014.

HONG, H.; MIAO, Y.; LIU, J.; ZHU, A-X. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *Catena* 176, 45–64, 2019.

HUANG, F.; YIN, K.; HUANG, J.; GUI, L.; WANG, P. Landslide susceptibility mapping based on self-organizing-map network and extreme learning Machine. **Engineering Geology**, 223, 11-22, 2017

IPT – Instituto de Pesquisas Tecnológicas do Estado de São Paulo S. A. **Mapeamento de Riscos em Encostas e Margem de Rios**. Celso Santos Carvalho, Eduardo Soares de Macedo e Agostinho Tadashi Ogura, organizadores – Brasília: Ministério das Cidades; Instituto de Pesquisas Tecnológicas – IPT, p. 176, 2007.

JOACHIMS, T. Learning to classify texts using support vector machines: methods, theory and algorithms. Kluwer Academic Publishers, 2002.

KALANTAR, B.; PRADHAN, B.; NAGHIBI, S. A.; MOTEVALLI, A.; MANSOR, S. Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). **Geomatics, Natural Hazards and Risk**, 9 (1), 49-69, 2018.

KASABOV, N.K. **Foundation of neural networks, fuzzy systems and knowledge engineering**. 2nded. Massachusetts: The MIT Press, 1998.

KIM, K. I.; JUNG, K.; PARK, S. H.; KIM, H. J. Support vector machines for texture classification. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 24(11), 1542–1550, 2002.

KUMAR, D.; THAKUR, M.; DUBEY, S. C.; SHUKLA, P. D. Landslide Susceptibility Mapping & Prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India. **Geomorphology**, 295, 115-125, 2017

LEE, S., RYU, J.-H., WON, J.-S., AND PARK, H.-J. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. **Engineering Geology**, 71 (3), 289–302, 2004.

LEE, S.; RYU, J. KIM, L. Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. **Landslides**, 4, 327–338, 2007.

LOPES, E. S. S. **Modelagem espacial dinâmica aplicada ao estudo de deslizamentos em uma região da Serra do Mar paulista, na escala de 1:10.000**. Tese (Doutorado em Geociências e Meio Ambiente), Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, p. 302, 2006.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma introdução às *Support Vector Machines*. **Revista de Informática Teórica e Aplicada**. 14(2), 43-67, 2007

LOSASSO, L.; SDAO, F. The artificial neural network for the rockfall susceptibility assessment. A case study in Basilicata (Southern Italy). **Geomatics, Natural Hazards and Risk**, 9 (1), 737-759, 2018.

LUCCHESI, L. V.; OLIVEIRA, G. G.; PEDROLLO, O. C. Attribute selection using correlations and principal components for artificial neural networks employment for landslide susceptibility assessment. **Environmental Monitoring and Assessment**, 192, 129, 2020. DOI: 10.1007/s10661-019-7968-0

MARCELINO, E. V. Desastres Naturais e Geotecnologias: Conceitos Básicos. **Caderno Didático** nº 1. INPE/CRS, Santa Maria, 2008.

MATSUNAGA, V.Y. **Curso de redes neurais utilizando o MATLAB**. Belém, 2012. Apostila. Disponível em: <http://paginas.uepa.br/engprod/index.php?option=com_rokdownloads&view=file&task=download&id=132%3Aapostilaredesneurais&Itemid=67>. Acesso em: 20 dez. 2018

MITCHELL, T. **Machine Learning**. McGraw Hill, 1997.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos Sobre Aprendizado de Máquina. Sistemas Inteligentes Fundamentos e Aplicações**. 1 ed. Barueri-SP: Manole Ltda, 89-114, 2003.

NEFESLIOGLU, H. A.; GOKCEOGLU, C.; SONMEZ, H. Na assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. **Engineering Geology**, 97, 171-191, 2009.

NETO, A. B.; BONINI, C. S. B. Redes Neurais Artificiais: Apresentação e utilização do algoritmo perceptron em biosistemas. **Revista Brasileira de Engenharia de Biosistemas - BioEng**, Tupã, 4 (2), 87-95, 2010.

NOBLE, W. S. **Support vector machine applications in computational biology**. In B. Schölkopf, K. Tsuda, and J.-P. Vert, editors, *Kernel Methods in computational biology*, pages 71–92. MIT Press, 2004.

NOLA, I. T. S. **Avaliação de dados geológico-geotécnicos prévios para elaboração de carta de eventos perigosos de deslizamentos gravitacionais por meio de redes neurais artificiais e probabilidade**. Dissertação (Mestrado) - Programa de Pós-Graduação e Área de Concentração em Geotecnia -- Escola de Engenharia de São Carlos da Universidade de São Paulo, p. 351, 2015.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R.; BRAVO, J. M. Simulações hidrológicas com diferentes proporções de área controlada na bacia hidrográfica. RBRH – **Revista Brasileira de Recursos Hídricos** 18 (3), 193-204, 2013.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. O desempenho das Redes Neurais Artificiais (RNAs) para simulação hidrológica mensal. RBRH – **Revista Brasileira de Recursos Hídricos** 19 (2), 251-265, 2014.

OLIVEIRA, G. G.; FLORES, T.; BRESOLIN JR. N. Análise do Evento Extremo Hidrológico Ocorrido na Bacia hidrográfica do Rio Forqueta em Janeiro de 2010. **Revista Caderno Pedagógico**, Lajeado, 14 (1), 181-192, 2017.

OLIVEIRA, G. G.; RUIZ, L. F. C.; GUASSELLI, L. A.; HAETINGUER, C. Random forest and artificial neural networks in landslide susceptibility modeling: a case study of the Fão River Basin, Southern Brazil. **Nat Hazards** 99, 1049, 2019. <https://doi.org/10.1007/s11069-019-03795-x>.

PAIXÃO, M. A.; KOBIYAMA, M.; MICHEL, G. P.; SILVA, J. S. **Aprender hidrologia para prevenção de desastres naturais**. 6º Congresso Brasileiro de Extensão Universitária, Universidade Federal do Pará, Belém, 2014.

PARK. S.; CHOI. C.; KIM, B.; KIM, J. Landslide susceptibility mapping using Frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje área. Korea. **Environment Earth Science** 68, 1443–1464, 2013

PELLERIN, J.; DUARTE, G. M.; SCHEIBE, L. F.; MENDONÇA, M.; BUSS, M. D.; MONTEIRO, M. A. Timbé do Sul - Jacinto Machado: Avaliação preliminar da extensão da catástrofe de 23-24/12/95. **Revista Geosul** n°23, 1º semestre de 1997

PINTO, R. C.; PASSOS, E.; CANEPARO, S. C. Classificação dos deslizamentos ocorridos em março de 2011 na Serra da Prata, Estado do Paraná, **Geoingá: Revista do Programa de Pós-Graduação em Geografia**, Maringá, 4 (1), 3-27, 2012.

PONTIL, M. & VERRI, A. Support vector machines for 3-D object recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 20 (6), 637–646, 1998.

POURGHASEMI, H. R.; RAHMATI, O. Prediction of the landslide susceptibility: Which algorithm, which precision? **Catena**, 162, 177 – 192, 2017

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

QUEVEDO, R. P.; GUASSELLI, L. A.; OLIVEIRA, G. G.; RUIZ, L. F. C. Modelagem de áreas suscetíveis a deslizamentos: avaliação comparativa de técnicas de amostragem, aprendizado de máquina e modelos digitais de elevação. São Paulo, UNESP, **Geociências**, 38 (3), 781 – 795, 2019.

RENNÓ, C.D. **Sistema de análise e simulação hidrológica aplicado a bacias hidrográficas**. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, p. 119, 2003.

RIFFEL, E. S. **Metodologias para mapeamento de suscetibilidade a deslizamentos**. Tese (Doutorado), Instituto de Geociências, Programa de Pós-Graduação em Geografia da Universidade Federal do Rio Grande do Sul, p. 125, 2017.

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

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**, 3 edition, p. 1151, 2010.

SAHIN, E. K.; COLKESEN, I.; KAVZOGLU, T. A Comparative Assessment of Canonical Correlation Forest, Random Forest, Rotation Forest and Logistic Regression Methods for Landslide Susceptibility Mapping. **Geocarto International**, 33 (12), 341-363, 2018.

SCHÖLKOPF, B.; GUYON, I.; WESTON, J. **Statistical learning and kernel methods in bioinformatics**. In P. Frasconi and R. Shamir, editors, Artificial Intelligence and Heuristic Methods in Bioinformatics, 1–21. IOS Press, 2003.

SILVA, I.N.; SPATTI, D.; FLAUZINO, R. **Redes neurais artificiais para engenharia e ciências aplicadas: curso prático**. São Paulo: Artliber, 2010.

SOUZA, M. C. C. **Uma análise do algoritmo K-Means como introdução ao aprendizado de máquinas**. Trabalho de conclusão de curso (Graduação), curso de licenciatura em matemática, Universidade Federal do Tocantins, p. 75, 2019.

THOMÉ, A.C.G. **Redes neurais – uma ferramenta para KDD e data mining**. Rio de Janeiro, 2003. Apostila. Disponível em: <http://equipe.nce.ufrj.br/thome/grad/nn/mat_didatico/apostila_kdd_mbi.pdf>.

TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. **Desastres Naturais: Conhecer para Prevenir**. São Paulo: Instituto Geológico, 2009.

VANACÔR, R. N.; ROLIM, S. B. A. Mapeamento da susceptibilidade a deslizamentos usando técnicas de estatística bivariada e sistema de informações geográficas na região nordeste do Rio Grande do Sul. **Revista Brasileira de Geomorfologia**, 13 (1), 15-28, 2012.

VON ZUBEN, F. J. **Máquinas de Vetores Suporte**. Departamento de Computação Aplicada, Faculdade de Engenharia Elétrica e de Computação, Unicamp, 2013.

XIAO, C-C.; TIAN, Y.; SHI, W-Z.; GUO, Q-H.; WU, L. A new method of pseudo absence data generation in landslide susceptibility mapping with a case study of Shenzhen.

Science China, **Technological Sciences** 53 (1), 75-84, 2010. <https://doi.org/10.1007/s11431-010-3219-x>

YILMAZ, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from kat landslides (Tokat-Turkey). **Computers and Geosciences**, 35, 1125-1138, 2009.

ZHAO, F.; MENG, X.; ZHANG, Y.; CHEN, G.; SU, X.; YUE, D. Landslide susceptibility mapping of Karakorum Highway combined with the application of SBAS-InSAR technology. **Sensors**, 19, 2685, 2019. DOI: 10.3390/s19122685.

ZHOU, C.; YIN, K.; CAO, Y.; AHMED, B.; LI, Y.; CATANI, F.; POURGHASEMI, H. R. Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. **Computers and Geosciences**, 112, 23-37, 2018.

WANG, Y.; FANG, Z.; HONG, H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. **Science of the Total Environment** 666, 975–993, 2019.

WANG, Y.; FENG, L.; LI, S.; REN, F.; DU, Q. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. **Catena** 188, 104425, 2020. <https://doi.org/10.1016/j.catena.2019.104425>