



Mapeamento da suscetibilidade a inundações usando os modelos Floresta Aleatória e Máquina Vetor de Suporte com diferentes tipos de kernel

Flood susceptibility mapping using Random Forest and Support Vector Machine models with different kernel types

Francisco Hélder Fernandes do Amaral*¹ ; Edson Luís Piroli¹ ; Viviane Corrêa Santos²

¹Universidade Estadual Paulista, Presidente Prudente, SP, Brasil.

²Universidade do Estado do Pará, Belém, PA, Brasil.

E-mail: edson.piroli@unesp.br (ELP); viviane.santos@uepa.br (VCS)

*Email para correspondência: helter.fernandes@unesp.br

Recebido (Received): 19/06/2023

Aceito (Accepted): 26/06/2024

Resumo: O objetivo deste artigo foi avaliar o potencial preditivo dos modelos Floresta Aleatória (RF) e Máquina de Vetores de Suporte (SVM) com diferentes funções kernel para a previsão espacial da suscetibilidade às inundações. A área de estudo foi a microbacia urbana do rio Castanhal, localizada na Amazônia oriental brasileira. A modelagem foi baseada em um inventário das inundações registradas de 2020 a 2022 e um banco de dados geoespacial com os seguintes fatores condicionantes: altitude, declividade, precipitação, aspecto, índice de potência do fluxo, índice de umidade topográfica, altura acima da drenagem mais próxima (HAND), distância dos canais, perfis de solo e curva número. Foram estimados cinco modelos de suscetibilidade a inundações usando os algoritmos Floresta Aleatória e Máquina de Vetores de Suporte com os quatro tipos de kernel, linear (LN), polinomial (PL), função de base radial (FBR) e sigmoide (SIG). Um conjunto de métricas estatísticas e a área sob a curva (AUC) foram utilizados para validar os modelos. Os resultados da predição AUC para os mapas de suscetibilidade a inundações gerados pelo SVM-FBR, SVM-PL, SVM-LN, e SVM-SIG foram 98,28%, 97,32%, 96,74% e 73,35%, respectivamente. Por outro lado, a taxa de predição AUC do RF foi a maior, 98,94%. Os fatores condicionantes mais influentes foram: altitude, HAND, perfis de solo, distância dos canais e precipitação. Pode-se concluir que tanto o RF quanto o SVM são capazes de gerar modelos de suscetibilidade eficientes e confiáveis. Os mapas de suscetibilidade resultantes podem ser benéficos nas estratégias de mitigação de inundações.

Palavras-chave: Geotecnologias; Risco ambiental; Desastres; Vulnerabilidade social.

Abstract: The objective of this article was to evaluate the predictive potential of the Random Forest (RF) and Support Vector Machine (SVM) models with different kernel functions for the spatial prediction of flood susceptibility. The study area was the urban watershed of the Castanhal River, located in the eastern Brazilian Amazon. The modeling was based on an inventory of floods recorded from 2020 to 2022 and a geospatial database with the following conditioning factors: altitude, slope, precipitation, aspect, flow power index, topographic humidity index, height above drainage plus proximity (HAND), channel distance, soil profiles and curve number. Five flood susceptibility models were estimated using the Random Forest and Support Vector Machine algorithms with four types of kernel, linear (LN), polynomial (PL), radial basis function (FBR) and sigmoid (SIG). A set of statistical metrics and the area under the curve (AUC) were used to validate the models. The AUC prediction results for the flood susceptibility maps generated by SVM-FBR, SVM-PL, SVM-LN, and SVM-SIG were 98.28%, 97.32%, 96.74%, and 73.35%, respectively. On the other hand, the AUC prediction rate of RF was the highest, 98.94%. The most influential conditioning factors were: altitude, HAND, soil profiles, distance from channels and precipitation. It can be concluded that both RF and SVM are capable of generating efficient and reliable susceptibility models. The resulting susceptibility maps can be beneficial in flood mitigation strategies.

Keywords: Geotechnologies; Environmental risk; Disasters; Social vulnerability.

1. Introdução

Acarretadas por chuvas intensas, tempestades tropicais, furacões, marés altas, derretimento de neve, ruptura de barragens ou diques, entre outros. As inundações representam o processo natural de submersão temporária de uma determinada área, geralmente marginal às fontes naturais ou artificiais de drenagem (RAHMATI *et al.*, 2016). Estando presente de forma recorrente em muitas regiões do mundo, incluindo o Brasil, que se encontra entre os dez países mais afetados por esse tipo de desastre (TOMINAGA, *et al.* 2015).

Dados do *Centre for Research on the Epidemiology of Disasters* (CRED) mostram que, entre 2000 e 2019, ocorreram cerca de 3.500 desastres relacionados a inundações em todo o mundo, afetando mais de 2 bilhões de pessoas e causando prejuízos estimados em US\$ 650 bilhões. No Brasil, segundo dados do Ministério do Desenvolvimento Regional, as inundações e enxurradas afetaram mais de 3,7 milhões de pessoas entre 2003 e 2020, causando prejuízos que ultrapassaram os R\$ 20 bilhões.

Além dos prejuízos econômicos, as inundações podem facilitar a propagação de doenças de veiculação hídrica, danificar infraestruturas de transporte e contaminar mananciais superficiais e subterrâneos, os quais são geralmente utilizados de forma direta ou indireta pela população local (DARABI, *et al.* 2019). Em casos extremos podem provocar a perda de vidas humanas, destruição de infraestrutura e de propriedade. No ambiente das cidades, as inundações podem ser agravadas pela extensa impermeabilização do solo, que reflete diretamente na elevação do volume de escoamento superficial, que em episódios de precipitações intensas gera sobrecarga no sistema de drenagem (VOJTEK; VOJTEKOVÁ, 2016). Além disso, estudos como de Cardona *et al.* (2012), destacam que as mudanças climáticas têm contribuído para o aumento da frequência e intensidade desses eventos, tornando o desenvolvimento de técnicas de monotonamente e antecipação dos eventos extremos, cada vez mais importante.

Nesse contexto, a utilização de modelos de suscetibilidade físico-ambiental que buscam tratar das diferentes probabilidades e intensidades com que um dado fenômeno impacta um determinado ambiente (VEDOVELLO; MACEDO, 2007). Tem se mostrado essencial na implementação de estratégias eficazes ligadas ao ordenamento territorial, na gestão do meio ambiente e nas áreas urbanas tem contribuído no monitoramento e predição de cenários físico-espaciais propensos aos impactos de eventos hidroclimáticos intensos (LEE *et al.*, 2017).

Apesar de as modelagens preditivas de determinação dos cenários físico-espaciais de suscetibilidade terem se destacado como um importante instrumento preventivo aos danos causados pelas inundações em diversos contextos e escalas (MASOOD; TAKEUCHI, 2012). Os métodos tradicionais, como técnicas de modelagem física de chuva-vazão e técnicas baseadas em estatísticas matemáticas, mostraram-se ineficazes, pois seguem procedimentos unidimensionais, enquanto a morfologia do rio é dinâmica e instável devido ao alto potencial erosivo (REFSGAARD, 1997). Além disso, esses métodos exigem trabalho de campo, uma faixa temporal extensa de dados e um grande orçamento para coleta de dados, tornando-os de difícil implementação (CUREBAL *et al.*, 2016).

Por conta das desvantagens apresentadas pelos métodos de modelagem tradicional, novas abordagens baseadas em modelos empíricos, gerados a partir de dados amostrais, com base em algoritmos de aprendizagem de máquina e aprendizagem profunda, têm-se popularizado na última década. Alguns desses são as redes neurais artificiais, máquina vetor de suporte, programação genética, regressão polinomial evolutiva, modelos de árvores estocásticas, floresta aleatória e K-vizinhos mais próximos, dentre outros (SELEEM *et al.*, 2022). Tais abordagens têm sido amplamente utilizados na construção de modelagem preditivos da suscetibilidade às inundações, e têm mostrado resultados satisfatórios na predição da magnitude e frequência desses eventos (OH; PRADHAN, 2011).

Apesar de vários modelos baseados em aprendizagem de máquina terem sido aplicados para diferentes contextos geoespaciais ao redor do mundo, a modelagem de suscetibilidade a inundações no ambiente urbano para cenários de microescala ainda se encontra pouco explorada (VOJTEK; VOJTEKOVÁ, 2016; DARABI *et al.*, 2022). Nesse contexto, a pesquisa teve como objetivo, analisar a suscetibilidade a inundações na microbacia hidrográfica urbana do rio Castanhal por meio dos algoritmos Máquina Vetor de Suporte (*Support vector machine* - SVM) e Floresta Aleatória (*Random Forest* - RF), com o propósito de auxiliar na gestão, planejamento e em tomadas de decisões que minimizem os transtornos decorrentes, à nível local.

2. Caracterização da área de estudo

A microbacia hidrográfica do rio Castanhal está situada no perímetro urbano do município de Castanhal, na porção nordeste do estado do Pará, na Amazônia Brasileira. O município é conhecido por sua relevância em prestação de serviços e infraestrutura, ocupando uma posição de influência regional intermediária.

A microbacia encontra-se entre as coordenadas 1°16'47.02" a 1°16'46.61" de latitude Sul e 47°56'52.16" a 47°54'53.18" de longitude Oeste (**Figura 1**). Cobrindo uma área de 8,46km² dos 55,33 km² da área municipal edificada, e abriga uma população de 45.000 habitantes de um total de 203.000 habitantes (IBGE, 2022).

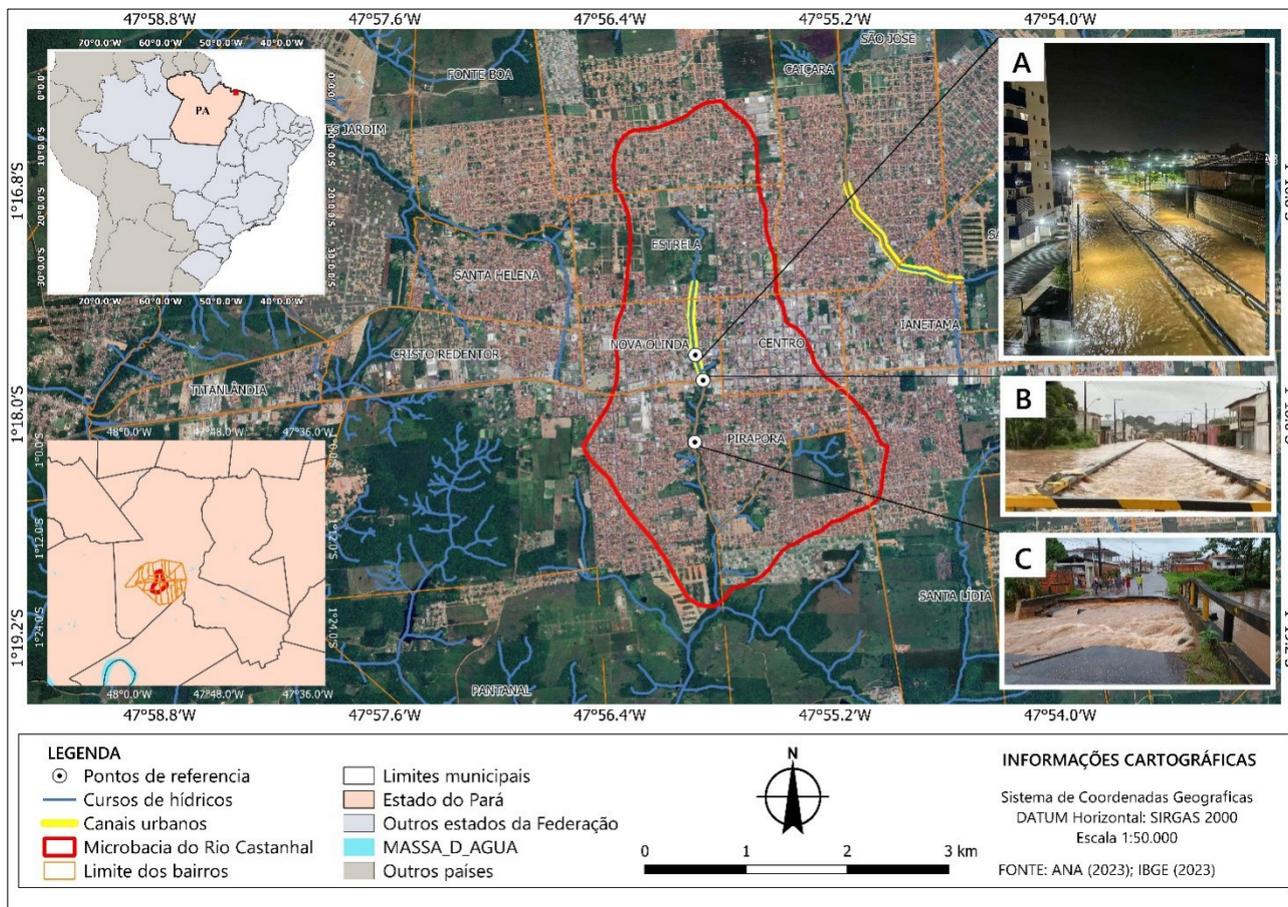


Figura 1: Localização da microbacia do rio Castanhal, estado do Pará, porção oriental da Amazônia, Brasil.

Esta microbacia tem uma topografia plana e uma elevação entre 29 m na jusante e 47 m a montante. O clima é úmido, com precipitação média anual 1.459 mm (2000–2020) e temperatura média de 25 °C (INMET, 2023). A cobertura florestal originária, na microbacia, foi quase completamente substituída pelo assentamento urbano, mas sua constituição florestal remanescente é típica da região amazônica, sendo predominante a floresta aluvionar.

3. Materiais e métodos

A metodologia utilizada neste estudo foi dividida em seis etapas: (I) preparação do inventário de inundação, (II) seleção dos fatores de condicionantes à inundação, (III) modelagem dos mapas de susceptibilidade de inundação usando os algoritmos RF e SVM, (IV) análise da influência dos fatores condicionantes de inundação, (V) análise dos modelos de Suscetibilidade, (VI) avaliação da acurácia preditiva dos modelos. Ressalta-se que toda a metodologia foi executada no software Rstudio.

3.1. Inventário das inundações urbanas

O mapa de inventário de inundações é fundamental para o desenvolvimento dos modelos de susceptibilidade a inundações (RAHMATI *et al.*, 2016). Neste estudo, foi utilizado um total de 400 pontos,

dentre os quais, 200 pontos se referem a condição de inundação e os demais como locais de não inundação. Eles foram gerados aleatoriamente com base no perímetro das manchas de inundação registradas nos eventos de 2020 a 2022, tais registros provem da investigação de campo e dos relatórios disponibilizados pela Defesa Civil Municipal de Castanhal. Uma proporção de 70% - 30% foi usada para dividir aleatoriamente o total do conjunto de dados do inventário de inundação em dois grupos, para treinamento (70% - n=280) e validação (30% - n=120) dos modelos, usando o software Rstudio.

3.2. Fatores que influenciam a inundação urbana

Tendo em vista a inexistência de uma regra universal para a escolha dos fatores condicionantes da suscetibilidade (RAHMATI *et al.*, 2015). Neste estudo, os fatores condicionantes foram selecionados com base na disponibilidade de informações geoespaciais adequadas e na revisão bibliográfica de pesquisas semelhantes (TEHRANY *et al.*, 2015).

Foram selecionados dez parâmetros que influenciam a suscetibilidade a inundações, sendo esses: altitude, declividade, aspecto, índice topográfico de umidade, índice de potência do córrego, modelo da altura acima da drenagem mais próxima (*Height Above the Nearest Drainage - HAND*), distância das infraestruturas de drenagem, grupos hidrológicos do solo (SHG), precipitação e taxa de número curva. Todos estes fatores foram recortados pelo perímetro da microbacia, que por sua vez, foi delimitado seguindo as recomendações de Flügel (1995) que utiliza a análise topográfica e hidrológica em Sistemas de Informações Geográficas (SIG), nesta pesquisa o SIG QGIS versão 3.22, foi utilizado para esse procedimento. com auxílio de um modelo digital de terreno (MDT) de alta resolução espacial.

Vale ressaltar que o Modelo Digital de Terreno (MDT) utilizado neste estudo possui uma resolução espacial de 5×5 metros e foi criado através do interpolador Rede Irregular Triangulada (*Triangulated Irregular Network - TIN*), usando curvas vetoriais planialtimétricas com 1 metro de equidistância vertical (1:10.000) no SIG QGIS. O MDT resultante foi então utilizado na confecção dos seguintes fatores: altitude, declividade, aspecto, índice topográfico de umidade (Topographic Wetness Index - TWI) e índice de potência do córrego (Stream Power Index - SPI), usando ferramentas nativas dos pacotes GDAL e GRASS contidas no SIG QGIS. Além disso, o MDT também foi empregado na confecção do fator HAND, que foi processado na extensão HAND-MODEL (<http://handmodel.ccst.inpe.br/>), desenvolvida pelo Instituto Nacional de Pesquisas Espaciais (INPE) em linguagem C++. Observe a (Figura 2).

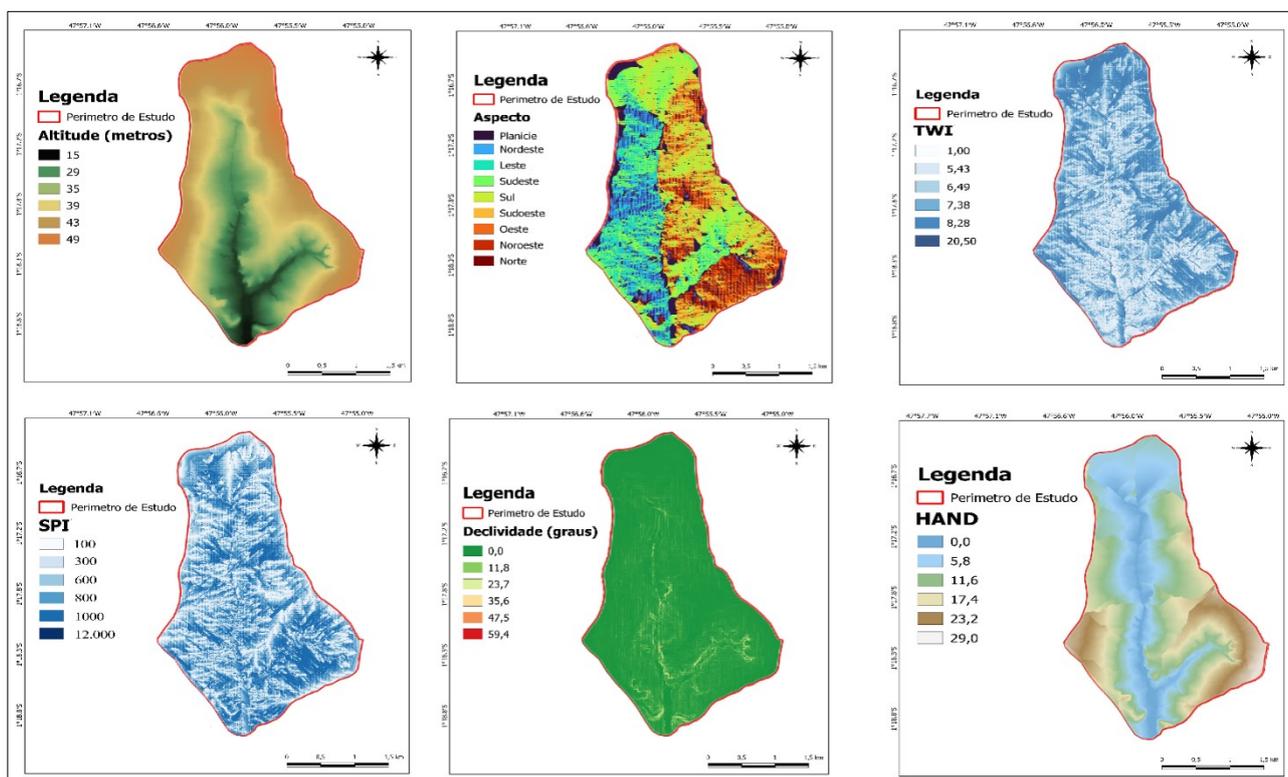


Figura 2: Camadas temáticas de entrada: altitude, aspecto, TWI, SPI, declive, HAND.

No que diz respeito às características relacionais entre os fatores condicionantes elencados, e o processo de inundação, salienta-se que a elevação mante uma relação inversa com as áreas de inundação, ou seja, quanto maior a elevação no terreno, menor a suscetibilidade, e o inverso também é válido (RAHMATI *et al.*, 2016). A declividade, por sua vez, contribui diretamente para a velocidade do escoamento na superfície, e por delinear a percolação vertical do volume hídrico junto as estruturas topográficas, influenciando a magnitude dos processos de inundação (TEHRANY *et al.*, 2015; RAHMATI *et al.*, 2016). O aspecto aparece como fator que descreve o ângulo do escoamento, ou seja, as direções cardeais do escoamento precipitado na superfície do terreno (NACHAPPA. *et al.*, 2020).

O Índice de Umidade Topográfica (TWI), por ser gerado a partir da condição de inclinação e área de contribuição a montante por unidade de largura ortogonal à direção do fluxo, é capaz de delinear a maior ou menor concentração de umidade nos diferentes níveis da topografia (R. BAG *et al.* (2022). A equação (**Equação 1**) de (AMBROISE *et al.* 1996) foi utilizada para calcular o TWI.

$$TWI = L_n \frac{\alpha}{\tan \beta + c} \quad \text{Eq. 1}$$

onde α é a acumulação de fluxo, β é a declividade e C é uma constante (0.01).

O índice de potência do fluxo SPI, é responsável por estimar a poder de transporte de matéria no canal de drenagem partindo do princípio de que o perímetro da bacia e a inclinação são proporcionais à descarga. A formula de (MOORE *et al.*, 1991), utilizada no cálculo do SPI, (**Equação 2**):

$$SPI = As_x \tan \sigma \quad \text{Eq.2}$$

Onde: As é a área de captação específica em metros e σ é o gradiente de inclinação.

Por sua vez, o HAND, concebe uma abordagem de realinhamento topo-hidrológico da altimetria do MDT, de acordo com a estrutura dos cursos de drenagem, determinando o potencial de drenagem gravitacional relativo as estruturas de sub-bacias (RENNÓ *et al.*, 2008). O processamento ocorre em três etapas, que adotam métodos distintos, já consolidados para as duas primeiras etapas e um método original para realização da terceira etapa, responsável pelo rearranjo da topografia do MDE segundo a estrutura HAND (NOBRE *et al.*, 2011).

O fator distância das infraestruturas de drenagem é responsável por determina a zona de espraiamento da inundação (HUBER e RUST, 2016). Para gera-lo foi utilizado o cálculo da distância euclidiana no ARCGIS com base em uma planta planimétrica das estruturas de pontes e canais na escala de 1:10.000, presentes na área da microbacia (**Figura 3**).

A precipitação compreende a principal fonte para as inundações urbanas (TEHRANY *et al.* 2015). Para considerar a espacialização da magnitude anual da precipitação, foram adquiridos dados de precipitação diária no Portal HidroWeb, uma ferramenta integrante do Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH), gerenciado pela Agencia Nacional de Águas e Saneamento Básico (ANA), sendo referentes as estações pluviométricas inseridas nos municípios de Santa Isabel do Pará, Santa Maria do Pará, Terra Alta, Igarapé-Açu, Castanhal e Bujaru, identificadas nos códigos: 148003, 147018, 147017, 147005, 147007, 148017, respectivamente. Posteriormente, esses dados foram rearranjados no software Excel do pacote office da Microsoft, para estimar a quantidade anual precipitada e espacializados no software gratuito QGIS versão 3.22, usando o interpolador da distância inversa ponderada (*Inverse Distance Weighted* - IDW). A quantidade registrada varia de 2.565 mm no leste da área de estudo a 2.568 mm no Oeste (**Figura 3**).

A textura do solo é responsável por controla as dinâmicas de escoamento superficial e inundação (BATJES,1997). Os grupos hidrológicos de solos (HSG) são fundamentais para explicar a capacidade dos diferentes grupos de solo em absorver o volume hídrico precipitado, algo que afeta a capacidade de formação do escoamento superficial, sendo que o potencial de infiltração tende a ser mais elevado nos solos do grupo (A), declinando sua capacidade conforme se desloca para os solos do grupo (D) (MOSAVI *et al.*,2020). Na área da microbacia predominam os solos do grupo D de acordo com a classificação de Sartori (2005), sendo representado pelos Argissolo Vermelho-Amarelo e Argissolo Amarelo, junto as áreas de várzea também encontram-se solos do tipo Neossolo Flúvico pertencente ao grupo C. O mapa de grupos hidrológicos de solos foi estimado com base na capacidade média de infiltração (solos do grupo C e D possuem 2,54 mm e 0,64 mm, respectivamente) descrita por Sartori (2005). Sendo os perfis de solo levantados junto ao banco de dados da Secretaria Municipal de Meio Ambiente de Castanhal (SEMMA), sendo esse compatíveis com a escala 1:10.000 (**Figura 3**).

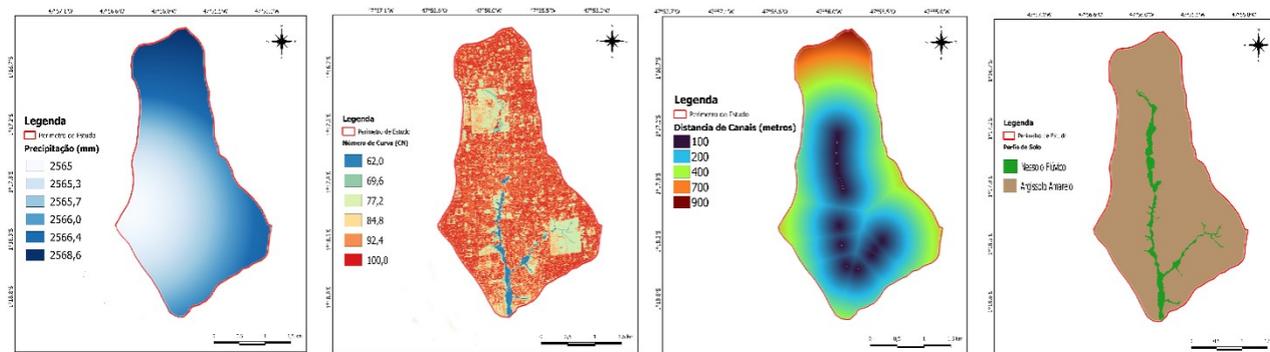


Figura 3: Camadas temáticas de entrada: precipitação, taxa CN, Distância dos canais, grupos hidrológicos de solo.

Por fim, a taxa de número de curva (*curver number*-CN) compreende uma informação importante responsável por descrever a situação de permeabilização ou impermeabilização da superfície do terreno. Este fator deriva da condição de apropriação da terra combinada com o potencial hidrológico de permeabilidade do solo. Para criar a camada CN, foi implementada uma função ArcCN-runof com base na planta planimétrica de uso e cobertura da superfície e o mapa de grupos hidrológicos dos solos. Pode-se observar a faixa de valor CN para a área da microbacia do rio Castanhal (60-100) é relativamente elevada, mostra uma predominância de áreas impermeáveis.

3.3. Modelagem da suscetibilidade a inundações

Neste estudo, o mapa de suscetibilidade a inundações foi criado usando dois modelos de aprendizado de máquina o RF e o SVM. O modelo Máquina vetor de suporte (*Support Vector Machine* - SVM) opera gerando um hiperplano ideal durante o momento de treinamento do modelo que reparte as classes amostrais, seguindo a lógica do menor quantitativo de amostras mal classificadas, convertendo estruturas não lineares em lineares. Os requisitos essenciais na definição dos vetores de suporte são o custo C, Gamma e função kernel. Geralmente o método de definição em grade é utilizado na estruturação dos parâmetros C e Gamma, que geram resultados de previsões confiáveis. O parâmetro de custo C, incorpora um impacto significativo na seleção do vetor de suporte e no desempenho final do SVM. Nesse sentido, a escolha do kernel representa uma etapa fundamental para a geração de um modelo consistente e seguro.

Haja vista, que a função matemática de kernel é responsável pelo método de separação do hiperplano no SVM. Os quatro tipos principais de kernel são: kernel polinomial (PL), kernel sigmoide (SIG), função de base radial (FBR) e kernel linear (LN). Os quatro tipos de kernel foram empregados, neste estudo de caso, para gerar diferentes modelos de suscetibilidade usando o SVM, afim de compreender a inferência destes na confecção dos modelos de suscetibilidade (NACHAPPA, *et al.*, 2020).

Por sua vez, o algoritmo de Floresta Aleatória (*Random Forest* - RF) é considerado um dos principais modelos não paramétrico de aprendizado de máquina empregado no mapeamento de suscetibilidade. Sendo um modelo *ensemble*, o RF combina a técnica de *bagging* (*bootstrap aggregating*) com árvores de decisão. O *bagging* é uma técnica que consiste em treinar várias instâncias do mesmo modelo em subconjuntos aleatórios dos dados de treinamento e, em seguida, combinar as previsões desses modelos para obter uma previsão final. Já as árvores de decisão são modelos de aprendizado de máquina que dividem iterativamente os dados em subconjuntos menores com base em critérios de separação, para então prever um valor para a variável de alvo (HO, 1995).

No modelo RF, várias árvores de decisão são treinadas em subconjuntos aleatórios dos dados de treinamento, com cada árvore utilizando um subconjunto diferente de características (variáveis explicativas). Essas árvores são treinadas independentemente umas das outras e produzem previsões individuais para cada registro de entrada. A previsão final é então calculada como a média das previsões de todas as árvores (HO, 1995). Ao combinar as técnicas de *bagging* e árvores de decisão, o modelo Random Forest consegue obter previsões mais precisas e robustas, ao mesmo tempo em que reduz a variância e o sobreajuste (*overfitting*) do modelo.

3.4. Análise da relevância dos fatores de influência da inundação

Antes de se proceder com o processo de predição da suscetibilidade, é recomendado que se verifique a adequabilidade dos fatores condicionantes, em modelos preditivos do tipo regressão, a colinearidade ou multicolinearidade representa a não independência de variáveis ou fatores explicativos para um dado evento (DORMANN, *et al.*, 2013). Sendo recomendado evitar a colinearidade entre os fatores explicativos afim de se gerar modelos não tendenciosos. Neste estudo de caso, a avaliação da existência de multicolinearidade, entre os fatores explicativos selecionados, ocorreu através dos índices do fator de inflação de variância (*Variance Inflation Fator* - VIF) e de tolerância (TOL). Para valores de VIF próximos ou superior a 5 se tem a existência de forte colinearidade. Para o índice de TOL, os valores inferiores a 0,1 indicam a presença de multicolinearidade extrema (HAIR *et al.*, 2013). O VIF (**Equação 3**) e TOL (**Equação 4**): são expressos matematicamente como:

$$VIF = \frac{1}{TOL} \quad \text{Eq.3}$$

$$TOL = 1 - R_v^2 \quad \text{Eq.4}$$

Onde: R_v^2 representa o 'coeficiente de determinação de uma regressão de fatores explicativos em todos os outros fatores explicativos' e $VIF > 5$ e $TOL < 0,1$ indica grave problema de colinearidade

3.5. Avaliação da precisão

Após a modelagem dos cenários de suscetibilidade a inundação usando os algoritmos SVM e RF, foi necessária determinar seu desempenho, para isso foram utilizadas as seguintes métricas estadística (**Equações 5 a 9**): A taxa de predição positivo (TPP), se refere a proporção de áreas inundadas identificados corretamente pelo modelo. taxa de predição negativa (TPN), é a proporção de áreas não inundações identificados corretamente. Sensibilidade, avalia a capacidade do modelo definir assertivamente o número de pixels classificados de forma positiva. A especificidade, mensura a capacidade do modelo em prever o número de pixels classificados corretamente como falsos. A acurácia, descreve a capacidade de classificar corretamente todas as amostras positivas e negativas (YESILNACAR, 2005).

A partir da matriz de confusão foram extraídas as seguintes métricas estatísticas:

$$\text{Taxa de Predição Positiva (TPP)} = \frac{VP}{VP + FP} \quad \text{Eq.5}$$

$$\text{Taxa de Predição Negativa (TPN)} = \frac{VN}{VN + FN} \quad \text{Eq.6}$$

$$\text{sensibilidade} = \frac{VP}{VP + FN} \quad \text{Eq.7}$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad \text{Eq.8}$$

$$\text{Acuracia Global} = \frac{VP + VN}{VP + VN + FP + FN} \quad \text{Eq.9}$$

Onde: número de pixels de inundações identificadas corretamente (verdadeiro positivo – VP), número de pixels de não inundação classificados corretamente (falso positivo - FP), número de Pixels de inundações classificados incorretamente (verdadeiro negativo - VN), número de pixels de não inundação pontuados incorretamente (falso negativo - FN).

Além destas, a característica de operação do receptor ou Curva-ROC (*Receiver Operating Characteristic* - ROC) foi utilizada na estimar a taxa de predição dos modelos, com base nos 30% de dados separados para validação (RAHMATI, *et al.*, 2019). A curva ROC foi avaliada com base na métrica da Área Sob a Curva (*Area Under the Curve* - AUC) responsável por estimar a probabilidade de que o valor de pixels rotulado corretamente ultrapasse o número de pixels estimados incorretamente. O valor numérico de inferência da AUC varia de 0 à 1, sendo quanto mais próximo de 1 o valor for, melhor o desempenho do modelo, enquanto que valores inferiores a 0,5 indicam modelos precários e sem aplicabilidade real (YESILNACAR, 2005).

4. Resultados e Discussões

4.1. Análise de fatores condicionantes

A partir da análise dos índices VIF e TOL (**Tabela 1**), observou-se a inexistência de multicolinearidade alta, entre os fatores, haja vista que o VIF e o TOL para todas as variáveis oscilaram de 1.09 a 3.63 e 0.27 a 0.91, respectivamente, encontrando-se dentro da faixa aceitável.

Tabela 1: Diagnóstico da multicolinearidade entre as variáveis independentes.

Fatores condicionantes	VIF	TOL
Altitude	3.633850	0.2751902
Aspecto	1.092913	0.9149858
Precipitação	1.829386	0.5466314
Número de curva	1.129059	0.8856936
Declividade	1.357023	0.7369070
Distância do canal	1.930404	0.5180262
Modelo HAND	2.786692	0.3588484
Grupos hidrológicos do solo	1.789130	0.5589309
Índice de potência do fluxo	1.126267	0.8878893
Índice de umidade topográfica	1.245940	0.8026072

A análise da relação entre os fatores condicionantes é os pontos de suscetibilidade nas dimensões, de inundação e não inundação, ocorreu usando um quadro de gráficos *box-plot* junto ao teste de Wilcoxon de duas amostras não pareado. O teste de Wilcoxon é utilizado para comparar medianas de duas amostras independentes e pode ser aplicado em situações em que a normalidade dos dados não pode ser pressuposta. Os resultados do teste de Wilcoxon incluem o teste (W), que é a soma dos ranks dos valores positivos ou negativos, dependendo da direção da diferença entre as amostras, e o p-valor, que indica a probabilidade de se obter um resultado tão extremo ou mais extremo que o observado, assumindo-se que as duas amostras são iguais. Se o p-valor for menor que o nível de significância pré-definido (0,05), rejeita-se a hipótese nula de que as duas amostras são iguais e concluir que existe diferença significativa. Caso contrário, não se pode rejeitar a hipótese nula e conclui-se que não há evidência suficiente para afirmar que as duas amostras são diferentes.

Os resultados do teste de Wilcoxon demonstraram que os fatores: elevação, HAND, distância dos canais e grupos hidrológicos, foram os únicos cuja os valores para inundação e não inundação, não se sobrepuseram (**Figura 5**), apresentando P-valor (<0.05), sendo os fatores de maior representatividade. Ressalta-se que no teste (W), esses fatores condicionantes também apresentaram valores mais elevados para inundação do que para não-inundação. Enquanto, os demais fatores: Aspecto, Precipitação, taxa CN, declividade, SPI, TWI, não apresentaram diferenças estatisticamente significante entre inundação e não inundação, mantendo P-valores (> 0.05). observando a (**Figura 2**), pode-se inferir que essa situação pode estar relacionada a condição de não uniformidade espacial das características condicionantes na área da microbacia (KANG *et al.*, 2019).

Considerando o comportamento relacional entre os pontos amostrais de ocorrência de inundação e as características geográficas dos fatores condicionantes representativos, pode-se observar que a ocorrência de inundação se concentra entre as cotas de 26 a 30 metros de altitude, abarcando a porção central da microbacia, onde se encontram as estruturas de canalização e retificação do rio Castanhal, a proximidade em relação a essas estruturas mostra uma predominância das amostras de inundação a uma faixa horizontal de até 50 metros, estando coerente com os achados de estudos anteriores (RAHMATI, *et al.*, 2016; TEHRANY *et al.*, 2015).

Realizando uma correlação espacial entre as amostras de inundação e não inundação com os valores do HAND. Nota-se que a recorrência das amostras de pontos de inundação aparece majoritariamente junto as cotas HAND de 1 a 2,5 metros e de forma pontual junto a faixa de 5 metros da topografia HAND, sendo pertinente a faixa topográfica classificada por Nobre *et al.*, (2011) como zona de suscetibilidade a “*waterlogged* ou alagamentos”. Por sua vez, os pontos amostrais referentes a não inundação se fazem predominantes nos intervalos topográficos 7,6 a 17,4 metros no HAND (**Figura 5**). Ressalta-se que pelo fato de este ser o primeiro estudo que incorpora a topografia HAND como fator condicionante em modelos de suscetibilidade orientados por algoritmos de Machine Learning, existe uma limitação para a comparação

sobre os achados envolvendo esse fator condicionante em específico, contudo, estima-se que estes mesmos resultados descritivos sejam úteis para comparações em pesquisas futuras na mesma temática.

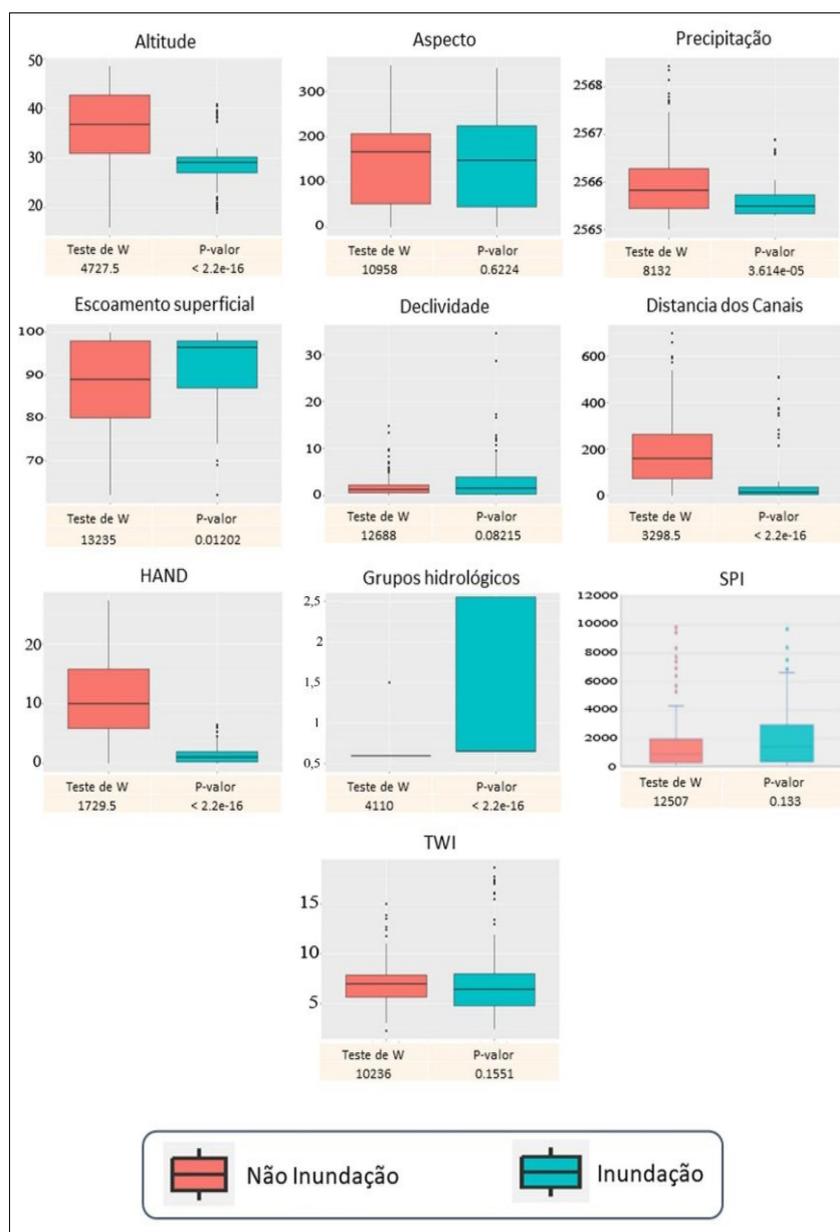


Figura 5. Estatísticas *box-plot* dos fatores condicionantes para ocorrência de inundação e amostras de não inundação

A análise da relação entre as amostras de inundação e os grupos hidrológicos de solo mostra que os Neossolos Flúvicos, pertencentes ao grupo hidrológico D, são predominantes. Segundo Sartori (2005), esses solos têm uma baixa permeabilidade de 0,64 mm e estão intimamente associados às áreas de várzeas, que são moldadas pela dinâmica hidrológica natural.

Assim como apontado por Darabi *et al.* (2022), que observa uma crescente antropização das estruturas de várzeas devido ao adensamento de habitações ocupadas por populações de baixa renda, e, portanto, com baixa capacidade de resiliência. Na microbacia do rio Castanhal também se verifica na região a jusante uma lógica similar, enquanto, que na porção a montante, a ocupação é voltada principalmente para o desenvolvimento de áreas comerciais, que embora tenha maior poder de resiliência, também apresenta maior potencial de prejuízos econômicos em cenários de inundações.

No que diz respeito a representatividade que os fatores condicionantes demonstraram sobre os modelos, destaca-se que os valores de importância dos fatores para os modelos SVM foram adquiridos usando a função “*varImp()*” do pacote “*caret*”, enquanto que a importância para o modelo RF foi carregada usando a

função “*varImpPlot()*” do pacote “*randomForest*”, no RStudio. Posteriormente, esses dados foram padronizados pela variável de maior importância em cada modelo. Na (Figura 6) são apresentadas as variáveis de maior importância para cada modelo. No modelo SVM-PL e para o modelo RF o fator de maior significância foi o HAND, para o modelo SVM-LN foi o Grupos hidrológicos do solo, no SVM-FBR foi o SPI e no SVM-SIG o fator Precipitação foi o que mais influenciou.

Ao analisar o ranking das médias dos valores de importância de cada fator sobre cada um dos modelos SVM, pode-se constatar que os cinco fatores de maiores relevâncias foram, Grupos hidrológicos do solo, HAND, precipitação, distância do canal, e altitude, respectivamente. Com exceção da precipitação, o nível de influência dos demais fatores demonstra estar correlacionado a configuração morfológica bem definida das áreas de várzea na microbacia, que também concentram depósitos de Nossolo Flúvicos e são alvo de processo de artificialização intensa (Figura 6). Cenário semelhante também é encontrado no trabalho de Seleem *et al.* (2022), onde os fatores elevação, precipitação e distância do canal de drenagem demonstram significância aparente nos modelos SVM e RF para cenários de inundação em áreas urbanas, devido a sua capacidade em determinam como a água se comporta em eventos de chuva intensa, influenciando a ocorrência e extensão das inundações.

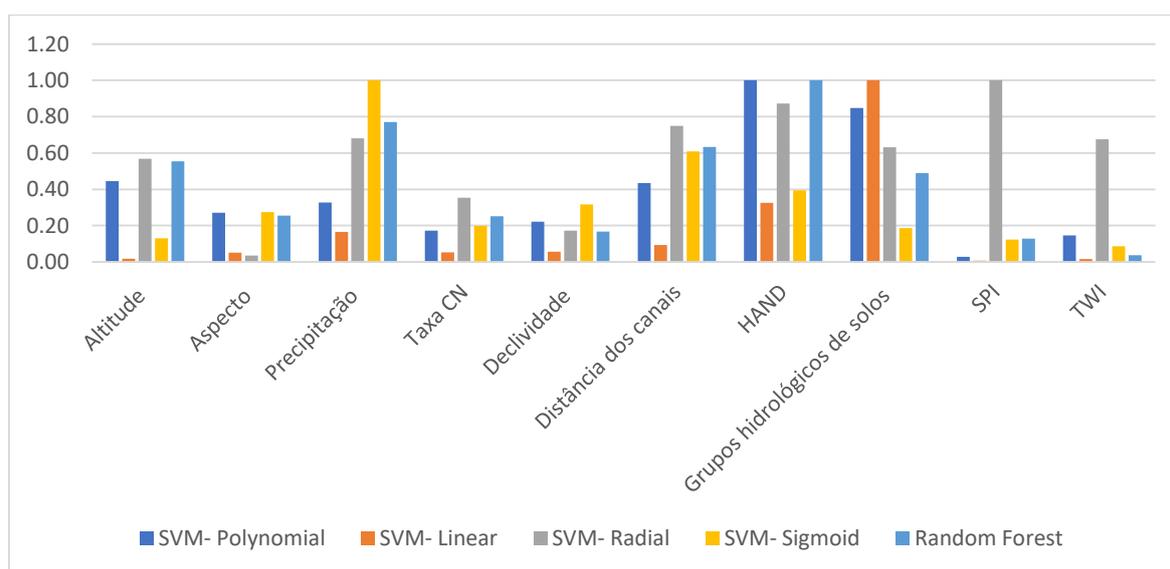


Figura 6: Nível da importância dos fatores de influência para os modelos gerados.

Os fatores condicionantes SPI e TWI foram considerados de grande importância para o modelo SVM-FBR, mas de baixa relevância para todos os outros modelos. A precipitação, apesar de não apresentar amplitudes de valores mínimo e máximo significativas no espaço, influenciou intensamente os modelos SVM-SIG, SVM-FBR e Random Forest. Em contraste, os fatores aspecto, taxa CN e declividade tiveram baixa significância em todos os modelos. Isso ocorre porque, em cenários de microescala, a variação espacial das características topográficas e do uso da terra é geralmente mínima, (como observado na Figura 2, onde o valor de declividade similar a oito graus compõe mais de 95% da área da microbacia, estando presente tanto em regiões amostradas como de inundação, quanto nas áreas de não-inundação), tornando difícil distinguir entre os atributos que descrevem os fatores condicionantes e as amostras de treinamento para um cenário de modelagem bilinear de presença ou ausência do fenômeno. Na microbacia, essa situação é aparente no teste de Wilcoxon que apontou apenas quatro fatores condicionantes como apresentando diferenciação significativa entre as dimensões de inundação e não-inundação.

4.2. Mapeamento de suscetibilidade a inundações

A cartografia da suscetibilidade a inundação, foram gerados no ambiente do software QGIS 3.22, usando a distribuição igual, por ser o esquema de classificação que melhor se ajustou para representatividade dos modelos. As classes de suscetibilidade a inundação variaram de muito baixa a muito alta e demonstraram que os modelos derivados do RF, SVM-FBR, SVM-POL, SVM-LN apresentam padrões de espacialização das classes de suscetibilidades “muito alta” e “alta” relativamente próximos. Contudo, o SVM-SIG apresenta dimensões de áreas suscetíveis disformes em comparação aos observados nos demais modelos, (Figura 7).

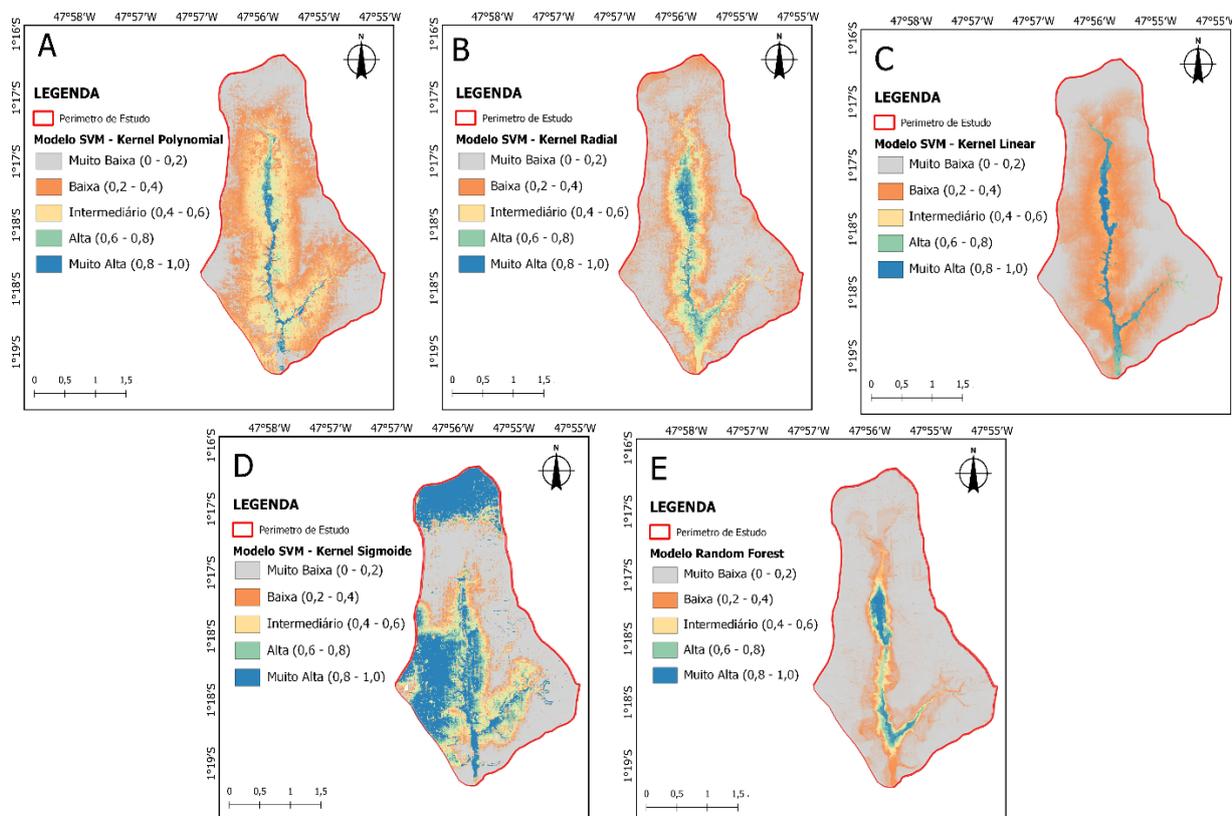


Figura 7: Mapas de suscetibilidade a inundações, gerado pelos modelos A = SVM-PL; B = SVM-FBR; C = SVM-LN; D = SVM-SIG; E = RF.

O mapa gerado pelo modelo RF mostrou a classe de “muito baixa” suscetibilidade recobrando a maior área ($7,56 \text{ km}^2$) e a classe de suscetibilidade “muito alta” cobrindo a menor área ($0,25 \text{ km}^2$) (**Figura 7**). No mapa do modelo SVM-SIG, a classe de suscetibilidade “muito alta” cobriu a maior área ($2,50 \text{ km}^2$) e a classe de suscetibilidade “muito baixa” cobriu uma área de ($3,98 \text{ km}^2$). De acordo com os modelos SVM-LN, SVM-FBR e SVM-PL, respectivamente, $0,38 \text{ km}^2$, $0,34 \text{ km}^2$ e $0,26 \text{ km}^2$ da área total da microbacia foram colocados na classe de suscetibilidade “muito alta”, e $7,12 \text{ km}^2$, $6,22 \text{ km}^2$ e $4,68 \text{ km}^2$ foram encontrados na classe de suscetibilidade “muito baixa”. A soma das áreas nas classes de suscetibilidade “muito alta” e “alta” são iguais a $0,41 \text{ km}^2$, $0,41 \text{ km}^2$, $0,69 \text{ km}^2$, $3,13 \text{ km}^2$ e $0,40 \text{ km}^2$ para os modelos SVM-LN, SVM-PL, SVM-FBR, SVM-SIG e RF, respectivamente (**Tabela 2**).

Tabela 2: volume de áreas (km^2) classificadas como suscetíveis a inundações por modelo.

Classes de suscetibilidade a inundações	SVM-LN	SVM-PL	SVM-FBR	SVM-SIG	RF
Muito baixa	7,128	4,690	6,223	3,985	7,566
baixa	0,911	2,324	0,896	0,648	0,353
Intermediaria	0,004	1,035	0,647	0,693	0,134
Alta	0,035	0,152	0,352	0,630	0,150
Muito alta	0,383	0,260	0,342	2,505	0,259

Na **Figura 7** pode-se observar também uma concentração das manchas de “muito alta” e “alta” suscetibilidade restrita aos bairros Nova Olinda, Saudade e Pirapora, na maioria dos modelos gerados, expondo um trecho pouco significativo de suscetibilidade elevada no bairro Cariri. Ficando evidente nestes modelos, que a presença de infraestruturas urbanas (sistema de macrodrenagem) junto aos sistemas de drenagem natural proporciona condições que favorecem processo de inundações, algo igualmente observado em Darabi *et al.* (2020), que teve a distancia do canal como principal fator condicionantes nos modelos de suscetibilidade a inundações gerados para analisar a área urbana da cidade de Amol, no Irã.

A **Tabela 3** que indica o número de ocorrências de inundação em cada classe de suscetibilidade mostra que o modelo derivado do SVM-SIG identificou o maior número de locais de inundação (189 pontos dos 200 locais de inundação catalogados aleatoriamente através do inventário de áreas atingidas entre (2010-2020) na classe de suscetibilidade “muito alta”, seguido por SVM-LN, SVM-PL, RF e SVM-FBR com 181, 179, 178 e 175 pontos de localização de inundação, respectivamente. Vale destacar que o SVM-LN identificou nove pontos de inundação na classe “baixa”. Isso porque a área na classe de suscetibilidade “muito alta”, “alta” e “intermediária” se concentram junto a área de concentração dos grupos hidrológicos associados ao Nossolo Flúvico e argissolo vermelho-amarelo. A análise constatou que o modelo SVM-SIG superou os demais modelos em termos de identificação do local de inundação, contudo, deve-se ressaltar que isso ocorreu devido a uma forte superestimação das áreas tidas com suscetibilidade “muito alta” e “alta”. (**Figura 7**).

Tabela 3: Número de ocorrências de inundação dentro de cada classe de susceptibilidade de inundação para cada modelo.

Classes de suscetibilidade a inundação	SVM-LN	SVM-PL	SVM-FBR	SVM-SIG	RF
Muito baixa	0	0	0	0	0
baixa	9	0	0	0	0
Intermediária	1	11	11	2	7
Alta	9	10	14	9	16
Muito alta	181	179	175	189	178

4.3. Avaliação de desempenho do modelo

Os resultados das métricas de validação mostram que o maior valor para TPP, ou seja, o modelo que representou o melhor ajuste mediante as amostras de ocorrência de inundação foi o SVM-FBR (0,964), seguido pelo RF (0,951) e SVM-PL (0,906). Todavia, os modelos SVM-LN (0,791) e SVM-SIG (0,661) demonstraram valores de baixa representatividade para TPP. Um padrão distinto é verificado na TPN, que avalia a proporção de acerto do modelo para as amostras de não-inundação. Neste índice, o modelo gerado pelo algoritmo RF demonstrou o melhor resultado (0,951), seguido pelo SVM-PL (0,964), SVM-LN (0,937), SVM-FBR (0,920) e SVM-SIG (0,655), respectivamente.

Analisando o índice de acurácia, responsável por pontuar a capacidade geral dos modelos quanto ao ajuste correto perante as amostras de inundação e não-inundação, tem-se que a melhor acurácia foi produzida pelo RF (AC = 0,966), seguida pelos SVM-FBR (AC = 0,9417), SVM-PL (AC = 0,933), SVM-LN (AC = 0,933) e SVM-SIG (AC = 0,658).

Em síntese, os modelos RF, SVM-PL e SVM-FBR, foram os que melhor representaram as áreas com inundação e de não-inundação, o modelo gerado pelo SVM-LN apesar de demonstrar uma TPP inferior aos do RF, SVM-FBR e SVM-PL, demonstrou uma TPN superior ao do SVM-FBR e uma acurácia adequada. Contudo, o SVM-SIG apresentou valores de baixa correspondência em todos os índices, reforçando a baixa adequabilidade desse modelo, para prever mapear a suscetibilidade a inundação no contexto de microescala, com dados de alta resolução-espacial.

A capacidade de predição dos cinco mapas de suscetibilidade gerados pelos modelos SVM e RF avaliados pelo método AUC (**Figura 8**), mostrou que o modelo SVM-SIG apresentou o menor valor AUC (AUC = 73%), enquanto que todos os demais modelos alcançaram valores de previsão AUC superiores 95%. Tehrany *et al.* (2014) que desenvolveu um estudo similar para Terengganu, na Malásia, associa esses resultados a menor eficiência do kernel sigmoide em separar com precisão as áreas prováveis de inundação, a partir de dados condicionantes que apresentam características espaciais muito próximas ou similares para as condições de inundação e não-inundação. Apesar deste método demonstra-se menos eficiente em comparação aos demais métodos SVM e RF, testados nessa pesquisa. Seu valor de AUC ainda é aceitável (TEHRANY *et al.*, 2015). Usando o sistema de categorização apresentado em Yesilnacar (2005), o modelo gerado pelo SVM-SIG seria classificado com nível de inferência “mediano” (AUC entre 0,6 - 0,7 ou 60% - 70%), enquanto que os demais receberiam a classificação “Excelente” (AUC entre 0,9 - 1 ou 90% - 100%).

Os resultados da estatística AUC também indicam que o modelo RF teve a maior precisão preditiva (AUC = 98,94%), seguido pelos modelos SVM-FBR (AUC = 98,28%), SVM-LN (AUC = 97%), SVM-PL (AUC = 96%) e SVM-SIG (AUC = 73%). Choubin *et al.* (2018) relaciona esse comportamento no resultado ao fato que o RF e um modelo ensemble, portanto, lida melhor com problemas complexos que os modelos não ensemble. Tal colocação é reforçada nas contribuições de Costache *et al.* (2021) que destacam a capacidade

do algoritmo Random Forest em lidar com conjuntos de dados de alta dimensionalidade, possuir baixa suscetibilidade a *overfitting*, além de possuir elevado potencial em lidar com interações complexas entre variáveis.

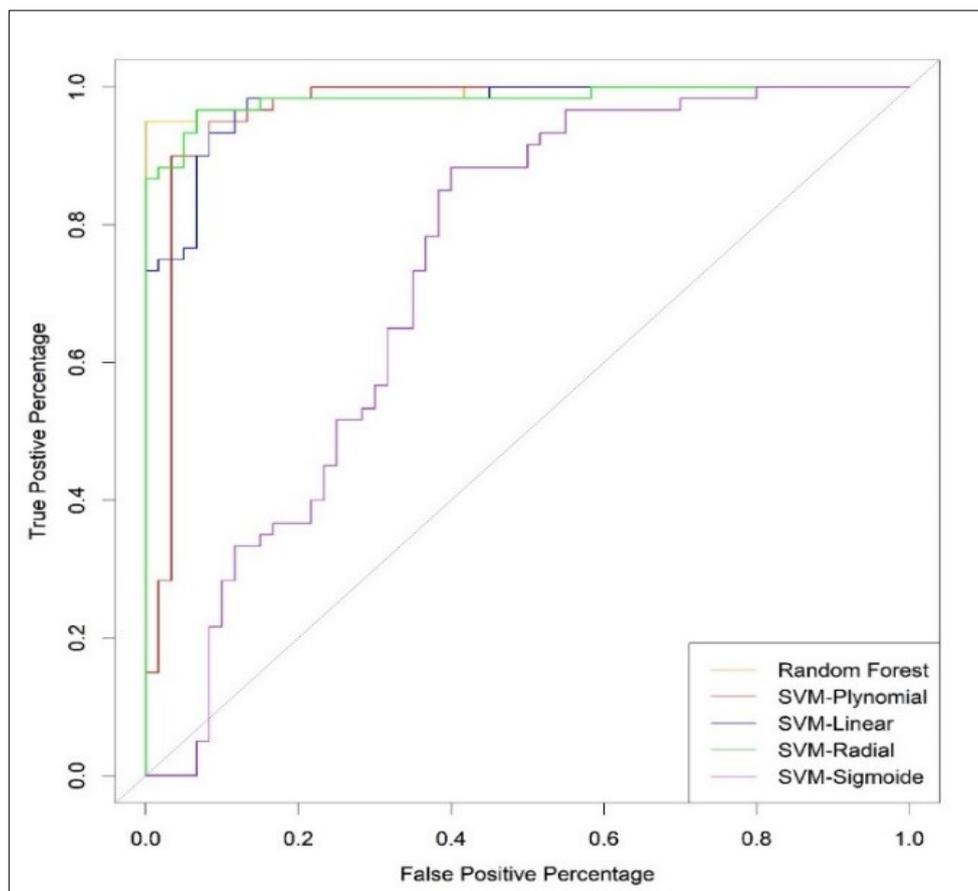


Figura 8: Curva AUC referente a Taxa de Previsão para cada um dos modelos gerados

Neste estudo, a taxa de previsão para todos os modelos (**Figura 8**). Demonstra que tanto o modelo ensemble Random Forest, quanto os modelos SVM derivados dos kernel's, polynomial, radial e linear são adequados na previsão da suscetibilidade a inundações. Apesar do SVM gerado pelo kernel sigmoide apresentar menor eficiência e potencial em superestimar as áreas com elevada suscetibilidade, fator também observado no trabalho de Tehrany *et al.* (2015), onde sua previsão estatística ainda é aceitável (YESILNACAR, 2005).

5. Conclusões

O mapeamento da suscetibilidade às inundações é uma ferramenta não estrutural válida de auxílio a políticas e estratégias na gestão de eventos extremos com potencial para se transformar em desastres, especialmente em áreas urbanas (DARABI *et al.*, 2020). Esses mapas demonstram ser recursos valiosos para a implementação de medidas de mitigação, aumentando a conscientização pública, melhorando a resposta a emergências e facilitando a recuperação pós-desastre (SELEEM *et al.*, 2022). Nesse contexto, a presente pesquisa demonstra elevado potencial de auxiliar os estudos hidrológicos locais na região Amazônica, permitindo detectar com mais precisão as áreas de maior predisposição a serem atingidas por inundações, bem como mensurando os fatores físicos-ambientais que mais estão relacionados.

Dentre os achados desta pesquisa, destaca-se que, conforme demonstrado por Choubin *et al.* (2018), o algoritmo ensemble RF obteve um valor de acurácia AUC superior aos modelos SVM. No entanto, o modelo SVM derivado do kernel radial também alcançou um nível de "excelência" na precisão da AUC, com um valor de referência muito próximo ao produzido pelo RF. De forma semelhante ao observado nos estudos de Tehrany *et al.* (2014, 2015), esta pesquisa destaca que a escolha do kernel impacta diretamente o desempenho geral do modelo SVM. Entre os kernels avaliados, os kernels radial, linear e polynomial se mostraram os mais adequados, enquanto o kernel sigmoide não foi considerado apto para gerar modelos de suscetibilidade.

Destaca-se ainda que, ao contrário de algumas pesquisas que apresentaram a declividade, o uso da terra e fatores correlatos, como a Curva CN, como extremamente relevantes na descrição da suscetibilidade a inundações em modelos de Machine Learning (CHOUBIN *et al.* 2018; DARABI *et al.* 2021; COSTACHE *et al.* 2021), este estudo apontou que fatores relacionados ao clima, infraestrutura urbana e caracterização da umidade no gradiente topográfico são descritores mais importantes. O HAND (Height Above Nearest Drainage) destacou-se como um achado significativo devido ao seu alto nível de significância em todos os modelos e à sua capacidade de delinear espacialmente as áreas de suscetibilidade ao ponto de permitir uma relação mais direta entre os espaços mais suscetíveis e os menos suscetíveis.

Contudo, apesar dos bons resultados alcançados neste estudo, é importante destacar algumas limitações. Primeiramente, a impossibilidade de mensurar a magnitude da inundação devido à inexistência de dados hidráulicos, como profundidade e velocidade da inundação, entre outros. Além disso, a escolha dos fatores condicionantes empregados pode não ser a mais adequada, sendo necessária uma futura reavaliação com a inclusão de novos fatores para uma compreensão mais precisa das variáveis que desencadeiam as inundações urbanas. Também é necessário replicar a metodologia em outros contextos urbanos, especialmente na região Amazônica, e incorporar outros algoritmos para validar com maior precisão a eficácia dos algoritmos propostos.

Por fim, recomendamos aos responsáveis pelo planejamento urbano e pela gestão de recursos hídricos em Castanhal que utilizem os recursos produzidos neste estudo para identificar zonas com alta propensão a inundações. De forma complementar, destacamos algumas ponderações importantes para mitigar a ocorrência de novos eventos de inundação na microbacia. Entre esses estão a necessidade de readequar as estruturas de canais, intensificar as atividades de dragagem, e utilizar soluções baseadas na natureza, como a restauração de áreas úmidas e a construção de lagoas de amortecimento (DARABI *et al.*, 2015). Acreditamos que essas abordagens combinadas podem diminuir consideravelmente a recorrência de futuros episódios de inundação na área da microbacia do rio Castanhal.

Notas:

Este trabalho faz parte da dissertação de mestrado do primeiro autor.

Referencias

AMBROISE, B.; BEVEN, K.; FREER, J. Toward a Generalization of the TOPMODEL Concepts: Topographic Indices of Hydrological Similarity. **Water Resources Research**, v.32, 2135–2145, 1996. DOI:10.1029/95wr03716

BAG, R.; MONDAL, I.; DEHBOZORGI, M.; BANK, S.P.; DAS, D.N.; BANDYOPADHYAY, J.; PHAM, Q.B.; AL-QURAIISHI, AYAD M.F.; NGUYEN, X.C.; Modelling and mapping of soil erosion susceptibility using machine learning in a tropical hot sub-humid environment. **Journal of Cleaner Production**. v.364, 132-428, 2022. DOI: 10.1016/j.jclepro.2022.132428

BATJES, N. H. A world dataset of derived soil properties by FAO-UNESCO soil unit for global modelling. **Soil Use and Management**, v.13, 9–16, 1997. DOI:10.1111/j.1475-2743.1997.tb00550.x

CARDONA, O. D.; VAN AALST, M. K.; BIRKMANN, J.; FORDHAM, M.; MCGREGOR, G.; PEREZ, R.; PULWARTY, R. S. Determinants of risk: exposure and vulnerability. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. **Cambridge University Press**. v.1, 65-108, 2012.

CHOUBIN, B.; MORADI, E.; GOLSHAN, M.; ADAMOWSKI J.; SAJEDI-HOSSEINI, F.; MOSAVI, A. An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. **Sci Total Environ**. v.651, 2087-2096, 2018. DOI: 10.1016/j.scitotenv.2018.10.064

COSTACHE, R.; ARABAMERI, A.; ELKHRACHY, I.; GHORBANZADEH, O.; PHAM, Q.B. Detection of areas prone to flood risk using state-of-the-art machine learning models. **Geomatics, Natural Hazards and Risk**. v.12, 1488-1507, 2021. DOI: 10.1080/19475705.2021.1920480

- CUREBAL, I.; EFE, R.; OZDEMIR, H.; SOYKAN, A.; SÖNMEZ, S. GIS-based approach for flood analysis: case study of Keçidere flash flood event (Turkey). **Geocarto Int.** v.31, 355–366, 2016. DOI: 10.1080/10106049.2015.1047411
- DARABI, H.; RAHMATI, O.; NAGHIBI, S.A.; MOHAMMADI, F.; AHMADISHARAF, E.; KALANTARI, Z.; HAGHIGHI, A.T.; SOLEIMANPOUR, S.M.; TIEFENBACHER, J.P.; BUI, D.T. Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood. **Geocarto International**, v.37, 5716-5741, 2022. DOI: 10.1080/10106049.2021.1920629
- DARABI, H.; HAGHIGHI, A.T.; MOHAMMADI, M.A.; RASHIDPOUR, M.; ZIEGLER, A.D.; HEKMATZADEH, A.A.; KLØVE, B. Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. **Hydrology Research**. v.51, 127–142, 2020. DOI:10.2166/nh.2019.090
- DORMANN, C. F.; ELITH, J.; BACHER, S.; CARRÉ, G. C. G.; GARCÍA MÁRQUEZ, J. R.; GRUBER, B.; LAFOURCADE, B.; LEITAO, P. J.; MÜNKEMÜLLER, T.; MCCLEAN, C. J.; OSBORNE, P. E.; RENEKING, B.; SCHRÖDER, B.; SKIDMORE, A. K.; ZURELL, D.; LAUTENBACH, S. Collinearity. a review of methods to deal with it and a simulation study evaluating their performance: open access. **Ecography**. v.36, 27-46, 2013. DOI:10.1111/j.1600-0587.2012.07348.x
- FLÜGEL, W.A. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. **Hydrological Processes**. v.9, 423–436, 1995. DOI:10.1002/hyp.3360090313
- GUDIYANGADA, N. T.; TAVAKKOLI, P. S.; GHOLAMNIA, K.; GHORBANZADEH, O.; RAHMATI, O.; BLASCHKE, T. Flood Susceptibility Mapping with Machine Learning, Multi-Criteria Decision Analysis and Ensemble Using Dempster Shafer Theory, **Journal of Hydrology**. v. 590, 2020, DOI: 10.1016/j.jhydrol.2020.125275
- HAIR, J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E. **Multivariate Data Analysis: Pearson New International Edition**. 7.ed. Pearson Education Limited; 2013.
- HO, T. K. Random decision forests, **Proceedings of 3rd International Conference on Document Analysis and Recognition**, Montreal, QC, Canada, v.1, 278-282, 1995. DOI: 10.1109/ICDAR.1995.598994.
- HUBER, S.; RUST, C. Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM). **Stata J.** v.16, 416–423, 2016.
- LEE, S.; KIM, J.C.; JUNG, H.S.; LEE, M.J.; LEE, S. Predição espacial da suscetibilidade a inundações usando modelos de floresta aleatória e árvores impulsionadas na cidade metropolitana de Seul, Coreia. **Geomática, Perigos e Riscos Naturais**. v.8, 1185-1203, 2017. DOI: 10.1080/19475705.2017.1308971
- MASOOD, M.; TAKEUCHI, K. Assessment of flood hazard, vulnerability and risk of mid-eastern Dhaka using DEM and 1D hydrodynamic model. **Nat. Hazards**. V.6, 1757–770, 2012. DOI: 10.1007/s11069-011-0060-x
- MOSAVI, A.; SAJEDI-HOSSEINI, F.; CHOUBIN, B.; TAROMIDEH, F.; HI, G.; DINEVA, A.A. Susceptibility mapping of soil water erosion using machine learning models. **Water**. v.12, 1-17, 2020, DOI:10.3390/w12071995
- NOBRE, A. D.; CUARTAS, L. A.; HODNETT, M.; RENNO, C. D.; RODRIGUES, G.; SILVEIRA, A.; WATERLOO, M.; SALESKA, S. Height Above the Nearest Drainage – a hydrologically relevant new terrain model. **Journal of Hydrology**. v.404, 13–29, 2011. DOI:10.1016/j.jhydrol.2011.03.051
- OH, H. J.; PRADHAN, B. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. **Computers & Geosciences**. v.37, 1264-1276, 2011. DOI:10.1016/j.cageo.2010.10.012

- RAHMATI, O. *et al.* PMT: new analytical framework for automated evaluation of geo-environmental modelling approaches. **Sci. Total Environ.** 664, 296–311, 2019. DOI: 10.1016/j.scitotenv.2019.02.017
- RAHMATI, O.; ZEINIVAND, H.; BESHARAT, M. Flood hazard zoning in Yasooj region, Iran, 533 using GIS and multi-criteria decision analysis. **Geomat. Nat. Haz. Risk.** v.7, 1000-1017, 2016. DOI: 10.1080/19475705.2015.1045043
- RAHMATI, O.; POURGHASEMI, H. R.; ZEINIVAND, H. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. **Geocarto International**, v.31, 42-70, 2015. doi:10.1080/10106049.2015.1041559
- REFSGAARD, J.C. Parameterisation, calibration and validation of distributed hydrological models. **Journal of Hydrology**, v. 198, p. 69-97, 1997. DOI: 10.1016/s0022-1694(96)03329-2.
- RENNÓ, C. D.; NOBRE, A. D.; CUARTAS, L. A.; SOARES, J. V.; HODNETT, M. G.; TOMASELLA, J.; WATERLOO, M.J. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. **Remote Sensing of Environment**, v.112, 3469-3481, 2008. DOI: 10.1016/j.rse.2008.03.018
- SARTORI, A.; NETO, F. L.; GENOVEZ, A, M. Classificação Hidrológica de Solos Brasileiros para a Estimativa da Chuva Excedente com o Método do Serviço de Conservação do Solo dos Estados Unidos Parte 1: Classificação. **RBRH – Revista Brasileira de Recursos Hídricos.** v.10, p. 05-18, 2005.
- SELEEM, O.; AYZEL, G.; DE SOUZA, A.C.T.; BRONSTERT, A.; HEISTERMANN, M. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. **Geomatics, Natural Hazards and Risk**, v.13, 1640-1662, 2022. DOI: 10.1080/19475705.2022.2097131
- TEHRANY, M. S.; PRADHAN, B.; MANSOR, S.; AHMAD, N. Flood susceptibility assessment using GIS-based support vector machine model with diferent kernel types. **Catena.** v.125, 91–101, 2015. Doi: 10.1016/j.catena.2014.10.017
- TEHRANY, M. S.; PRADHAN, B.; JEBUR, M. N. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. **Journal of Hydrology**, v.512, 332–343, 2014. DOI: 10.1016/j.jhydrol.2014.03.008
- TOMINAGA, L. K; SANTORO, J; AMARAL, R. do (Orgs). **Desastres naturais: conhecer para prevenir.** 3. ed. São Paulo: Instituto Geológico, 2015. 196 p. il. color. ISBN 978-85-87235-09-1.
- VEDOVELLO, R. ;MACEDO, E. S. DE. Deslizamentos de encostas. In: Santos, Rosely Ferreira dos. (org.). **Vulnerabilidade ambiental: desastres naturais ou fenômenos induzidos?** Brasília, Ministério do Meio Ambiente – MMA, pp. 75-94. 2007.
- VOJTEK. M.; VOJTEKOVÁ J. Flood hazard and flood risk assessment at the local spatial scale: a case study. **Geomatics, Natural Hazards and Risk.** v.7, 1973–1992, 2016. DOI: 10.1080/19475705.2016.1166874
- YESILNACAR, E.K. **The application of computational intelligence to landslide susceptibility mapping in Turkey.** Parkville, Victoria: University of Melbourne; 2005.



Este artigo é distribuído nos termos e condições do *Creative Commons Attributions/Atribuição-NãoComercial-CompartilhaIgual (CC BY-NC-SA)*.