MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2022/04.29.14.46-TDI

MAPEAMENTO DE ÁREAS ALAGÁVEIS NA BACIA AMAZÔNICA UTILIZANDO O CLASSIFICADOR RANDOM FOREST A PARTIR DE DADOS EXTRAÍDOS DO MDE-SRTM

Deborah Lopes Correia Lima

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelo Dr. Camilo Daleles Rennó, aprovada em 25 de abril de 2022.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34T/46Q63PS>

INPE
São José dos Campos
2022

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2022/04.29.14.46-TDI

MAPEAMENTO DE ÁREAS ALAGÁVEIS NA BACIA AMAZÔNICA UTILIZANDO O CLASSIFICADOR RANDOM FOREST A PARTIR DE DADOS EXTRAÍDOS DO MDE-SRTM

Deborah Lopes Correia Lima

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelo Dr. Camilo Daleles Rennó, aprovada em 25 de abril de 2022.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34T/46Q63PS>

INPE
São José dos Campos
2022

Dados Internacionais de Catalogação na Publicação (CIP)

Lima, Deborah Lopes Correia.

L628m Mapeamento de áreas alagáveis na Bacia Amazônica utilizando o classificador Random Forest a partir de dados extraídos do MDE-SRTM / Deborah Lopes Correia Lima. – São José dos Campos : INPE, 2022.

xx + 85 p. ; (sid.inpe.br/mtc-m21d/2022/04.29.14.46-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2022.

Orientador : Dr. Camilo Daleles Rennó.

1. Áreas alagáveis. 2. Random Forest. 3. Modelo digital de elevação. 4. SRTM. 5. Seleção de atributos. I.Título.

CDU 528.8:551.312.2

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

Serviço de Pós-Graduação - SEPGR

DEFESA FINAL DE DISSERTAÇÃO DE DEBORAH LOPES CORREIA LIMA
BANCA Nº 094/2022, REG 183470/2020

No dia 25 de abril de 2022, as 14h, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestra em Sensoriamento Remoto. O trabalho precisa da incorporação das correções sugeridas pela Banca Examinadora e revisão final pelo(s) orientador(es).

Título: "MAPEAMENTO DE ÁREAS ALAGÁVEIS NA BACIA AMAZÔNICA UTILIZANDO O CLASSIFICADOR *RANDOM FOREST* A PARTIR DE DADOS EXTRAÍDOS DO MDE-SRTM"

Membros da banca:

Dra. Evlyn Marcia Leão de Moraes Novo - Presidente - INPE
Dr. Camilo Daleles Rennó - Orientador - INPE
Dr. Thales Sehn Körting - Membro Interno - INPE
Dr. Vitor Souza Martins - Membro Externo - Mississippi State University (MSU)

Documento assinado eletronicamente por **Vitor Souza martins (E)**, **Usuário Externo**, em 10/05/2022, às 10:12 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Thales Sehn Korting**, **Pesquisador**, em 10/05/2022, às 10:37 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Camilo Daleles Rennó**, **Tecnologista**, em 10/05/2022, às 11:48 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Evlyn Marcia Leão de Moraes Novo**, **Pesquisador**, em 11/05/2022, às 08:49 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site http://sei.mctic.gov.br/verifica.html, informando o código verificador **9714644** e o código CRC **D70879EC**.

Referência: Processo nº 01340.002926/2022-65 SEI nº 9714644

“O sonho é o que leva a gente para frente...”

Ariano Suassuna

Dedico a meus pais, Zélia e Joaquim e

a minha irmã, Clara.

Com amor.

AGRADECIMENTOS

Agradeço ao meu orientador, Dr. Camilo Daleles Rennó, pela confiança, suporte durante o desenvolvimento desta pesquisa e incentivos constantes. Agradeço a paciência e todos os ensinamentos transmitidos.

Agradeço aos professores, Dra. Evlyn Márcia Leão de Moraes Novo, Dr. Thales Sehn Körting e Dr. Vitor Souza Martins, pelas contribuições nesta dissertação e por aceitarem o convite para participar da banca examinadora.

Agradeço ao Programa de Pós-Graduação do Instituto Nacional de Pesquisas Espaciais (INPE) e aos professores da PGSER que por meio dos seus ensinamentos contribuíram com meu aprendizado. Também agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, pelo financiamento desta pesquisa.

Agradeço ao grupo de pesquisa ao qual esta pesquisa está vinculada, em especial à Renata Pacheco Quevedo pelo apoio em diferentes momentos deste mestrado. Muito obrigada pelo carinho!

Agradeço aos professores Dr. Adriano Rolim da Paz e Dr. Gustavo Barbosa Lima da Silva pela orientação ainda na graduação e incentivo em cursar o mestrado em Sensoriamento Remoto no INPE. Obrigada por tantos ensinamentos, amizade e por acreditarem no meu potencial.

Agradeço aos amigos, Isadora Haddad, Karolina Gameiro, Maíra Matias, Felipe Sá e Bruno Adorno por tanta parceria durante esse tempo, que nem sempre foi fácil. Obrigada por tornar essa caminhada mais leve e divertida. Em especial, agradeço à Rejane Paulino, minha *roommate* e parceira de estudos na madrugada. Muito obrigada pela amizade e pelas inúmeras conversas sobre esta pesquisa. Foi muito valioso para mim!

Agradeço de maneira muito especial a meus pais Zélia e Joaquim por serem minha base e por me proverem toda a estrutura para eu chegar até aqui. Agradeço também a minha irmã Clara, por ser minha parceira de vida e pelo incentivo constante. Obrigada por fazerem parte dessa conquista!

RESUMO

As áreas alagáveis são uma fonte importante de recursos naturais, desempenham um papel significativo nos ciclos hidrológico e de carbono, e fornecem diversos serviços ecossistêmicos. A bacia Amazônica possui características favoráveis à formação de grandes extensões de áreas alagáveis. O mapeamento acurado dessas áreas é importante para dar suporte ao desenvolvimento de diferentes estudos nesses ambientes. Devido à sua vasta dimensão e à dificuldade de acesso, o sensoriamento remoto se torna uma fonte importante de dados para o seu estudo. Por exemplo, o Modelo Digital de Elevação (MDE) fornece informações morfométricas que podem auxiliar no mapeamento das áreas alagáveis. Nesse sentido, o objetivo desta pesquisa foi propor um método para o mapeamento das áreas alagáveis da bacia Amazônica a partir de atributos morfométricos e hidrológicos extraídos do MDE derivado do *Shuttle Radar Topography Mission* (SRTM). Como área de estudo foram utilizadas cinco subáreas que apresentam diferentes padrões geomorfológicos, distribuídas ao longo da bacia hidrográfica. O classificador *Random Forest* (RF) foi utilizado para gerar o mapeamento, principalmente por oferecer duas abordagens de interesse para esta pesquisa: (i) a quantificação da importância dos atributos por meio da métrica *Mean Decrease Accuracy* (MDA); e (ii) a probabilidade de o *pixel* pertencer a uma determinada classe. O valor do MDA foi utilizado como base do processo de seleção dos atributos. Por outro lado, o valor da probabilidade associada a cada *pixel* foi utilizado para calcular a incerteza da classificação a partir da métrica Entropia de Shannon, o que orientou o processo iterativo de amostragem. A análise do ranqueamento indicou que atributos que em média ocuparam posições mais altas no ranqueamento (considerando múltiplas iterações do RF), apresentaram uma maior estabilidade da posição ocupada comparados aos atributos que ocuparam as menores posições no ranqueamento. A partir do processo de seleção dos atributos, o conjunto inicial de 124 atributos, foi reduzido a apenas 9. Em comparação com o modelo completo (contendo os 124 atributos), o modelo reduzido não apresentou diferenças expressivas em relação à acurácia de mapeamento e às métricas tradicionais de avaliação (e.g. sensibilidade e precisão). A avaliação da incerteza espacialmente distribuída, calculada a partir da entropia de Shannon, indicou que reduzir a quantidade de atributos levou o classificador a convergir para o resultado com um maior nível de certeza. De maneira geral, o mapeamento obtido apresentou coerência com o MDE-SRTM, indicando o potencial da metodologia proposta. A metodologia e os resultados apresentados nesta pesquisa contribuem para um melhor entendimento acerca da: (i) utilização de atributos extraídos do MDE para o processo de classificação das áreas alagáveis; (ii) possibilidade de redução do conjunto de atributos àqueles mais importantes para a classificação; e (iii) do impacto da redução dos atributos sobre o mapeamento.

Palavras-chave: Áreas alagáveis. *Random Forest*. Modelo Digital de Elevação. SRTM. Seleção de atributos. Mapa de incerteza. Entropia de Shannon.

MAPPING WETLANDS IN THE AMAZON BASIN USING RANDOM FOREST CLASSIFIER AND ATTRIBUTES DERIVED FROM DEM-SRTM

ABSTRACT

Wetlands are important natural resources that play a key role in the hydrologic and carbon cycles and provide a wide range of ecosystem services. The Amazon basin has favorable conditions for the formation of wide-ranging wetlands which need to be accurately mapped to support different studies. Since Amazon wetlands present a huge extension and difficult access, remote sensing may be an important source of information. For example, Digital Elevation Model (DEM) may provide information about the topography, which is essential for mapping wetlands. In this sense, the main aim of this research was to develop a method for mapping wetlands in the Amazon basin using morphometric and hydrological attributes extracted from the Shuttle Radar Topography Mission (SRTM) DEM. The study area included five subareas in the watershed with different morphological patterns. The classification and mapping were carried out through the Random Forest (RF) algorithm, which provides: (i) the quantification of attribute importance using the Mean Decrease Accuracy (MDA) metric; and (ii) the probability of a pixel to be classified to a certain class. This pixel probability allowed to estimate the mapping uncertainty based on the Shannon's Entropy, which guided an iterative sampling process. The stability of the attribute's ranking positions was accessed and analyzed. Attributes that, on average, occupied higher ranking positions had higher ranking stability considering multiple RF runs. On the other hand, attributes that on average had lower rank had more unstable ranking. The method for selecting attributes allowed to reduce them from 124 to only 9 attributes. No significant differences were found between the mapping accuracy for the full model (with 124 attributes) and the reduced model (9 attributes). The analysis of the uncertainty mapping, calculated by Shannon's Entropy, indicated that the reduction of attributes provided a classification with higher level of certainty. In general, the obtained mapping was consistent with DEM-SRTM, which indicates the potential of the proposed methodology. The methodology and the results presented in this study contribute to a better understating of (i) the use of DEM attributes for wetland mapping; (ii) the possibility of reducing the dataset to only the most relevant attributes for the classification; and (iii) the impact of the reduction of attributes in the mapping.

Keywords: Wetlands. Random Forest. Digital Elevation Model. SRTM. Variable selection. Uncertainty map. Shannon's Entropy.

LISTA DE FIGURAS

Pág.

LISTA DE TABELAS

Pág.

SUMÁRIO

Pág.

1 INTRODUÇÃO

As áreas alagáveis da bacia Amazônica são áreas de transição entre ecossistemas aquáticos e terrestres, cuja profundidade da coluna d'água apresenta grandes flutuações ao longo do tempo. Essas regiões estão sujeitas ao alagamento causado pelo transbordamento lateral de rios ou lagos, ou ainda pela incidência de precipitação direta ou pela influência das águas subterrâneas (JUNK, 1989; PIEDADE et al., 2013). Esse tipo de ecossistema é considerado valioso devido ao fornecimento de diversos serviços ecossistêmicos, tais como sequestro de carbono, preservação da qualidade da água, formação de habitats de espécies endêmicas e atenuação das cheias dos rios (MITSCH; BERNAL; HERNANDEZ, 2015). A partir da Convenção de Ramsar, realizada em 1971, a importância desses ambientes foi reconhecida e diversas questões foram levantadas a respeito da conservação e uso sustentável dessas áreas. Uma das demandas decorrentes desse tratado internacional é a realização de inventários, ou seja, levantamentos que indiquem onde estão localizadas e qual a extensão das áreas alagáveis (FINLAYSON et al., 2011). Esse tipo de levantamento é de grande importância para tomadores de decisão, ajudando o desenvolvimento de ações para a conservação desses ambientes. Além disso, esses mapeamentos são a base para diversos estudos relacionados, por exemplo a estimativa da emissão de gases do efeito estufa (RICHEY et al., 2002; MITSCH et al., 2013), a avaliação do desflorestamento (RENÓ et al., 2011) e a valoração dos serviços ecossistêmicos (COSTANZA et al., 2014).

Devido à grande extensão e, muitas vezes, à dificuldade de acesso, a obtenção de dados in situ para auxiliar na delimitação das áreas alagáveis se torna limitada. Nesse sentido, o sensoriamento remoto representa uma importante fonte de dados, que torna o mapeamento das áreas alagáveis, a nível de bacia hidrográfica, viável e com menor custo benefício (OZESMI; BAUER, 2002). Diferentes tipos de dados obtidos a partir do sensoriamento remoto têm sido aplicados no estudo das áreas alagáveis (ADAM; MUTANGA; RUGEGE, 2010; GUO et al., 2017). Imagens ópticas têm sido utilizadas de maneira individual (ZHU et al., 2017) ou em conjunto com dados de radar (ROBERTSON; KING;

DAVIES, 2015) para o mapeamento e caracterização das áreas alagáveis. Dados topográficos extraídos de Modelos Digitais de Elevação (MDE) (e.g. declividade e índice topográfico) também têm sido utilizados em combinação com outras informações na realização do mapeamento das áreas alagáveis (BWANGOY et al., 2010; MILLARD; RICHARDSON, 2013; DUBEAU et al., 2017; MARTINS et al., 2020). A utilização da informação da topografia em estudos de mapeamento das áreas alagáveis é algo relevante que pode auxiliar na identificação dessas áreas dada a forte relação entre relevo e hidrologia.

Dentre os estudos realizados para o mapeamento de áreas alagáveis na bacia Amazônica utilizando dados de sensoriamento remoto há o trabalho desenvolvido por Hess et al. (2003) e ampliado por Hess et al. (2015) que foi utilizado como base em diferentes outros estudos, como por exemplo na determinação do fluxo de carbono (RICHEY et al., 2002), na estimativa de emissão do metano (MELACK et al., 2004), no estudo da vulnerabilidade desses ambientes às mudanças climáticas (MELACK; COE, 2013) e na avaliação do desflorestamento (RENÓ et al., 2011). O mapeamento produzido por Hess et al. (2003, 2015) possui resolução espacial de aproximadamente 100 metros e foi feito a partir de um mosaico de imagens *Synthetic Aperture Radar* (SAR) adquiridas pelo *Japanese Earth Resources Sattelite*-1 (JERS-1) durante a época de águas altas e a época de águas baixas, entre os anos de 1995 e 1996. Entretanto, os autores afirmaram que a estimativa de áreas alagáveis resultante do mapeamento pode estar subestimada, visto que situações como inundações ao longo de rios de pequena ordem e durante épocas do ano que não foram representadas pelas imagens utilizadas, não foram consideradas no mapeamento. Além disso, não foi levada em consideração a variabilidade interanual do processo de alagamento das áreas. A partir da análise visual do mapeamento realizado por Hess et al. (2015) em conjunto com o MDE derivado do *Shuttle Radar Topography Mission* (SRTM) notam-se inconformidades entre os produtos em diferentes regiões, o que pode prejudicar estudos em escala local. A atualização deste mapeamento com informações extraídas do MDE-SRTM e a utilização de algoritmos de *machine learning* pode proporcionar dados em escala compatível com estudos locais. Diferentes estudos têm utilizado

métodos de *machine learning* para o mapeamento das áreas alagáveis (BWANGOY et al., 2010; ROBERTSON; KING; DAVIES, 2015; ZHU et al., 2017). Dentre os algoritmos de *machine learning*, o *Random Forest* (RF) tem apresentado um bom desempenho (BERHANE et al., 2018b).

O RF é um classificador baseado em árvores de decisão, e além dos bons resultados de mapeamento, tem se destacado devido à alta velocidade de processamento e à sua capacidade de processar uma grande quantidade de dados (BELGIU; DRAGUT, 2016). Uma das características do RF que o torna vantajoso é a possibilidade de quantificar a importância dos atributos utilizados no modelo. Alguns estudos utilizam a informação de importância dos atributos para realizar uma seleção dos atributos a serem aplicados na classificação (MILLARD; RICHARDSON, 2013; MEYER et al., 2017). A seleção dos atributos pode ser uma etapa importante no processo de classificação, pois permite identificar atributos mais relevantes para a classificação, o que pode auxiliar no entendimento de como eles se relacionam com o fenômeno em estudo. A utilização de um conjunto reduzido de atributos pode reduzir o custo computacional envolvido na obtenção, processamento e armazenamento dos dados. Além disso, a utilização apenas dos atributos mais importantes pode levar a uma melhoria na acurácia e uma redução dos ruídos da classificação em relação à utilização de vários atributos (MILLARD; RICHARDSON, 2015). Dessa forma, o RF combinado com técnicas de seleção de variáveis demonstra ser uma alternativa para o desenvolvimento de metodologias aplicadas ao mapeamento das áreas alagáveis na bacia Amazônica, sendo uma possibilidade para a melhoria do mapeamento realizado por Hess et al. (2015).

1.1 Hipótese

A utilização de atributos extraídos de Modelos Digitais de Elevação como dados de entrada para o algoritmo *Random Forest*, combinada com uma seleção de atributos, pode auxiliar no mapeamento das áreas alagáveis na bacia Amazônica.

1.2 Objetivo geral

O objetivo geral desta pesquisa é desenvolver um método de mapeamento de áreas alagáveis na bacia Amazônica com base em atributos extraídos do MDE-SRTM.

1.3 Objetivos específicos

- Obter o mapeamento de áreas alagáveis de diferentes regiões da bacia Amazônica a partir de atributos extraídos do MDE-SRTM utilizando o classificador RF;
- Propor método de amostragem iterativa utilizando a informação da incerteza de mapeamento de modo a produzir um resultado mais robusto e válido em regiões com diferentes padrões geomorfológicos.
- Propor método de seleção de atributos baseado na métrica de importância *Mean Decrease Accuracy* (MDA);
- Avaliar o impacto da redução de atributos na classificação.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção irá abordar diferentes conceitos necessários para o desenvolvimento desta pesquisa. Inicialmente serão tratados aspectos gerais a respeito das áreas alagáveis na bacia Amazônica. Em seguida será feita uma revisão da utilização de dados de sensoriamento remoto no mapeamento das áreas alagáveis, seguida de uma visão geral sobre o mapeamento tomado como referência no processo de amostragem nesta pesquisa. Esta seção também inclui uma breve descrição sobre o modelo digital de elevação a ser utilizado nesta pesquisa, o classificador *Random Forest* e seleção dos atributos. Por fim, será abordada a avaliação do mapeamento por meio de mapas de incerteza construídos a partir da métrica entropia de Shannon.

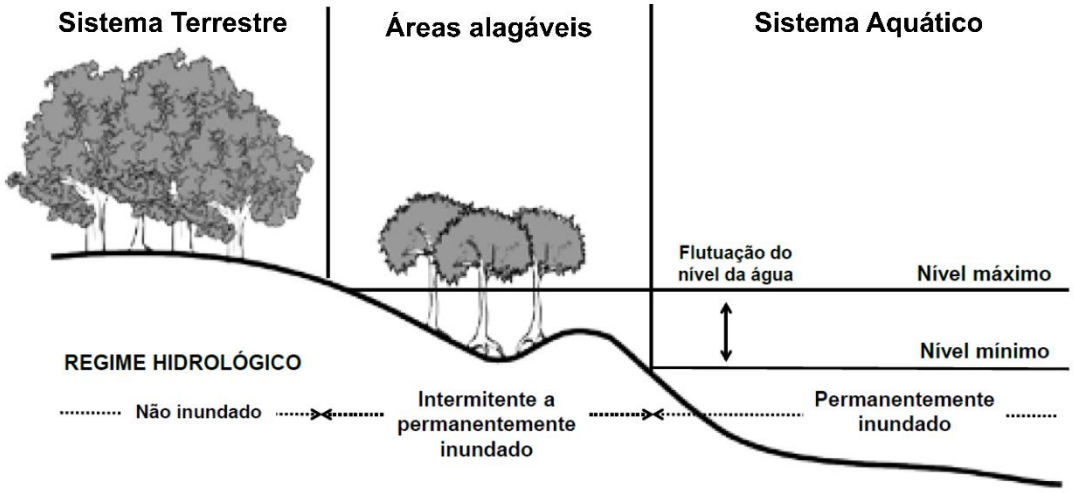
2.1 Áreas alagáveis amazônicas

Áreas úmidas são ecossistemas que ocorrem entre os ambientes aquáticos e terrestres. Essas áreas podem ser naturais ou artificiais, permanentes ou periodicamente inundáveis, além de serem o habitat de diversas espécies de plantas e animais adaptados a sua dinâmica hidrológica (JUNK et al., 2013). De acordo com Piedade et al. (2013), as áreas úmidas podem ser divididas em dois grupos conforme o seu regime hidrológico, sendo eles: áreas que apresentam coluna de água relativamente estável, como os pântanos, e as áreas úmidas com grandes flutuações do nível de água. As áreas alagáveis são ambientes que pertencem a essa segunda categoria (Figura 2.1). Junk et al. (1989) definem as áreas alagáveis como áreas sujeitas ao alagamento causado pelo transbordamento lateral de rios ou lagos ou ainda pela incidência direta de precipitação ou influência das águas subterrâneas. Ainda segundo os autores, a dinâmica única desses ambientes leva os organismos a desenvolverem estratégias adaptativas morfológicas, fenológicas, dentre outras, que propiciam a utilização desses habitats.

A bacia Amazônica possui características geomorfométricas (e.g. extensas áreas planas) que associadas à variação do volume de chuva precipitado ao longo do ano, propiciam a formação de vastas planícies de inundação. As

planícies de inundação são áreas que margeiam os rios e estão sujeitas a alagamentos de acordo com o regime hidrológico da região. Dessa forma, as planícies de inundação são um tipo de áreas alagáveis. As regiões que não sofrem com os processos de inundação, são chamadas de terra firme. As áreas alagáveis ao longo dos rios barrentos, com maior quantidade de nutrientes e pH próximo da neutralidade, são chamadas de várzeas. Já as áreas alagáveis ao longo dos rios de águas pretas e claras, menos férteis e ácidas, denominam-se igapós (JUNK et al., 2011).

Figura 2.1 - Esquema de representação das áreas alagáveis.

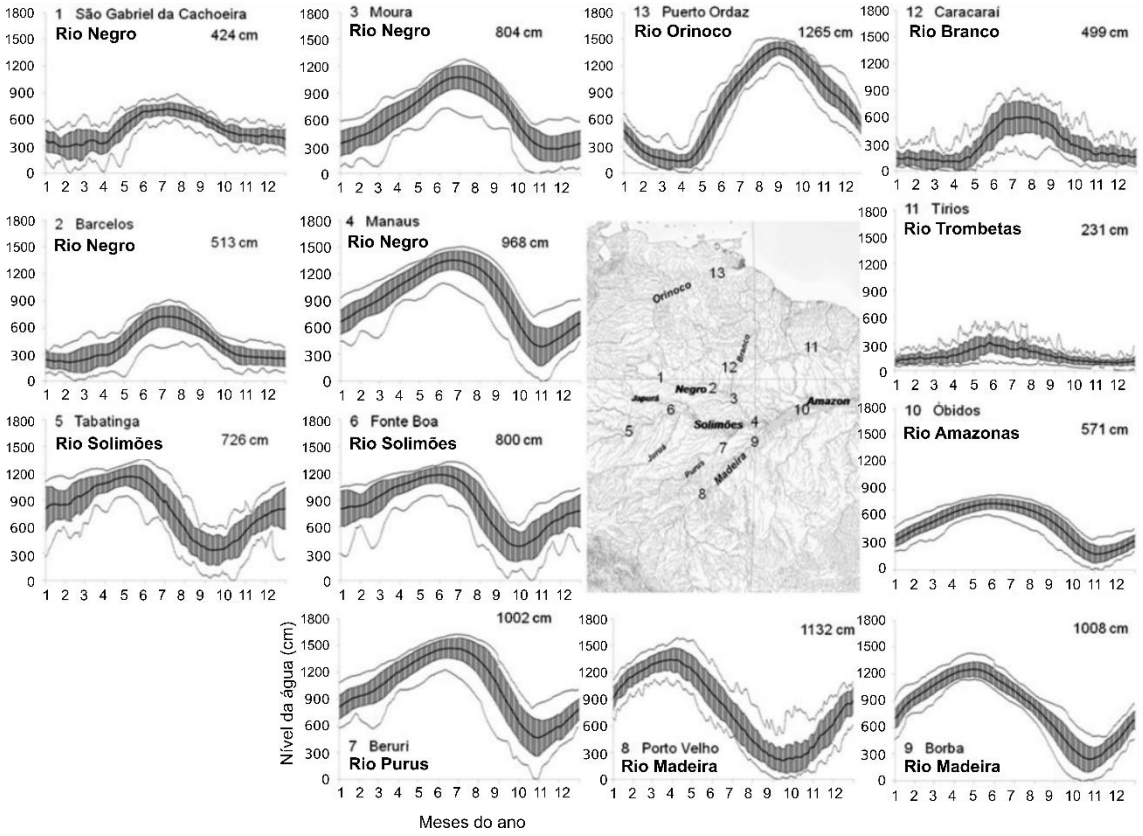


Fonte: Adaptado de Rosa et al. (2015).

A precipitação na bacia Amazônica apresenta uma variabilidade sazonal bem definida. A sazonalidade da precipitação e diferenças na sua distribuição espacial resultam em oscilações do nível da água dos rios, o que define duas fases ao longo do ano, a fase de águas altas (cheia) e outra de águas baixas (conhecida localmente como seca) (PIEDADE et al., 2013). Esse fenômeno de oscilação anual do nível da água é denominado pulso de inundação e determina o grau de conexão entre o rio e sua planície de inundação, sendo esse fenômeno a principal força responsável pela produtividade e interações na biota nas planícies de inundação (JUNK; BAYLEY; SPARKS, 1989). A oscilação do nível d'água não ocorre de maneira igual ao longo dos rios da bacia. A amplitude da

variação do nível d'água, assim como o momento em que o rio atinge o nível máximo variam entre os rios da bacia Amazônica, como mostra a Figura 2.2. Por exemplo, enquanto no Rio Madeira (na estação Porto Velho) o nível máximo é atingido no mês de abril, o Rio Negro (e.g. na estação Barcelos) atinge nível máximo no mês de julho.

Figura 2.2 - Variações do nível da água em 13 estações fluviométricas localizadas nos rios Amazonas, Solimões, Negro, Branco, Trombetas, Purus, Madeira e Orinoco.



Os números distribuídos ao longo da bacia Amazônica indicam a localização geográfica das estações no mapa. Os gráficos indicam durante o ano o nível médio (curva preta), o desvio padrão (área cinza) e máximas e mínimas (curvas cinza) baseado em análises do período de 1983 a 2005. Os números em centímetros indicam a amplitude média (dados: estações 1, 2, 3, 6, 7, 8, 9, 10, 11, 12: Agência Nacional de Águas – ANA; estações 4, 5: Superintendência Estadual de Navegação, Portos e Hidrovias - SNPH; estação 13: Ministerio del Poder Popular para Ciencia y Tecnología, The Environmental Research Observatory (ORE) HYBAM).

Fonte: Junk et al. (2011).

De maneira geral, as áreas alagáveis são classificadas como um dos ecossistemas mais valiosos do mundo (MITSCH; BERNAL; HERNANDEZ, 2015). Além do importante papel ambiental, as áreas alagáveis amazônicas exercem um significativo papel socioeconômico na região. Os habitantes das comunidades ribeirinhas fazem uso dessas áreas, adaptando suas atividades de acordo com o período de águas baixas e águas altas. Entretanto, esses ambientes são constantemente ameaçados pelas atividades antrópicas e pelas mudanças climáticas (KINGSFORD, 2011; CASTELLO et al., 2013). Renó et al. (2011), avaliando o desflorestamento na região Amazônica, evidenciaram que as áreas alagáveis são um dos ecossistemas mais ameaçados da bacia Amazônica. Dessa forma, é evidente a necessidade de desenvolvimento de estudos voltados ao monitoramento e entendimento desses ambientes, possibilitando também a obtenção de informações necessárias para uma utilização mais sustentável dessas regiões.

2.2 Sensoriamento remoto no estudo das áreas alagáveis

A interpretação de imagens aéreas foi o primeiro método de sensoriamento remoto utilizado no mapeamento de áreas alagáveis (MAHDAVI et al., 2018). As imagens aéreas possuem alta resolução espacial, o que permite um mapeamento mais detalhado (OZESMI; BAUER, 2002). Entretanto, a utilização dessas imagens pode demandar mais tempo e recursos financeiros quando comparada à aplicação de imagens de satélite (MAHDAVI et al., 2018). Dentre os produtos obtidos por meio do sensoriamento remoto, os dados de satélite surgem como os mais eficientes e com melhor custo-benefício para o mapeamento e monitoramento das áreas alagáveis, especialmente em áreas extensas. Por serem adquiridas com uma determinada frequência, as imagens de satélite permitem o monitoramento dessas regiões ao longo do tempo. Além disso, possibilitam obter informações a respeito do uso e cobertura do solo nas regiões às margens das áreas alagáveis e sua variação ao longo do tempo, o que pode fornecer informações a respeito do estado de conservação dessas áreas (OZESMI; BAUER, 2002). Entretanto, apesar de ser uma ferramenta útil

para o monitoramento e mapeamento das áreas alagáveis, a utilização dos dados de satélite apresenta alguns desafios devido a determinadas características desses ambientes, como por exemplo: a presença de diferentes tipos de cobertura; a alta variabilidade temporal do processo de alagamento e a ausência de limites bem definidos (GALLANT, 2015).

Diferentes dados de sensoriamento remoto e metodologias têm sido utilizados para o mapeamento das áreas alagáveis. Os dados ópticos são frequentemente aplicados para essa finalidade (ADAM; MUTANGA; RUGEGE, 2010), sendo utilizados por exemplo, para o cálculo de índices espectrais, servindo como dados de entrada para classificadores do tipo *machine learning* (ZHU et al., 2017). Entretanto, o mapeamento dessas áreas utilizando dados ópticos pode ser limitado devido à presença de nuvens e à dificuldade de identificar regiões alagadas abaixo da copa das árvores (HESS et al., 2003). Além disso, a presença de plantas aquáticas flutuantes que pode ser interpretada como cobertura vegetal, assim como a vegetação não totalmente submersa, podem levar a uma classificação incorreta como áreas de terra firme.

Os dados de radar apresentam algumas vantagens em relação aos dados ópticos e por isso são amplamente empregados no mapeamento das áreas alagáveis (GUO et al., 2017). Os sensores de micro-ondas apresentam algumas vantagens quando comparados aos sensores ópticos, como por exemplo: i) a possibilidade de obter informações a qualquer hora do dia e sob diferentes condições meteorológicas, o que é vantajoso principalmente em áreas com grande cobertura de nuvens e ii) determinados comprimentos de onda na faixa de micro-ondas podem penetrar a vegetação, o que permite obter informações das regiões abaixo das copas das árvores, tornando possível distinguir áreas alagadas e não alagadas (OZESMI; BAUER, 2002). Diferentes metodologias têm sido aplicadas no mapeamento das áreas alagáveis utilizando dados de radar (WHITE et al., 2015). Por exemplo, Chen et al. (2014) utilizaram dados polarimétricos obtidos do *Advanced Land Observing Satellite/ Phased Array L-band Synthetic Aperture Radar* (ALOS/PALSAR) em conjunto com análise baseada em objeto e um algoritmo de árvore de decisão para a classificação dos tipos de cobertura em uma região de áreas alagáveis. As limitações dos dados

de radar envolvem a disponibilidade limitada de dados, a falta de longas séries históricas e a complexidade nas análises (GUO et al., 2017).

Dados ópticos multiespectrais e dados de radar detectam características distintas dos alvos. Nesse sentido, podem ser utilizados de forma complementar para o mapeamento da vegetação e extensão das áreas alagáveis (CAI et al., 2020). Como exemplo, Robertson, King e Davies (2015) utilizando uma combinação de imagens ópticas do WorldView-2 e dados polarimétricos do Radarsat-2, aplicaram a técnica de classificação orientada a objeto para o mapeamento da vegetação presente em uma região de áreas alagáveis. Os autores constataram que a inclusão dos dados de radar levou a uma melhoria da acurácia do mapeamento de algumas classes de vegetação, em relação à acurácia utilizando apenas dados ópticos. Uma alternativa aplicada para o mapeamento das áreas alagáveis é o uso de dados topográficos. Nesse sentido, Millard e Richardson (2013) destacaram o uso de dados extraídos do sensor *Light Detection and Ranging* (LiDAR) para o mapeamento de áreas alagáveis. Os autores, utilizando o classificador RF, observaram que o mapeamento feito com base apenas nos dados LiDAR resultou em uma acurácia global de 71,9%, similar ao valor obtido na classificação usando a combinação dos dados LiDAR e de radar (72,8%). Por outro lado, Dubeau et al. (2017) ressaltaram a importância da combinação de dados ópticos, de radar e topográficos para o mapeamento das áreas alagáveis. Os autores utilizaram o classificador RF e alcançaram uma acurácia global de 94,4% e 92,9% para as estações seca e cheia, respectivamente.

2.3 Mapeamento de áreas alagáveis na bacia Amazônica

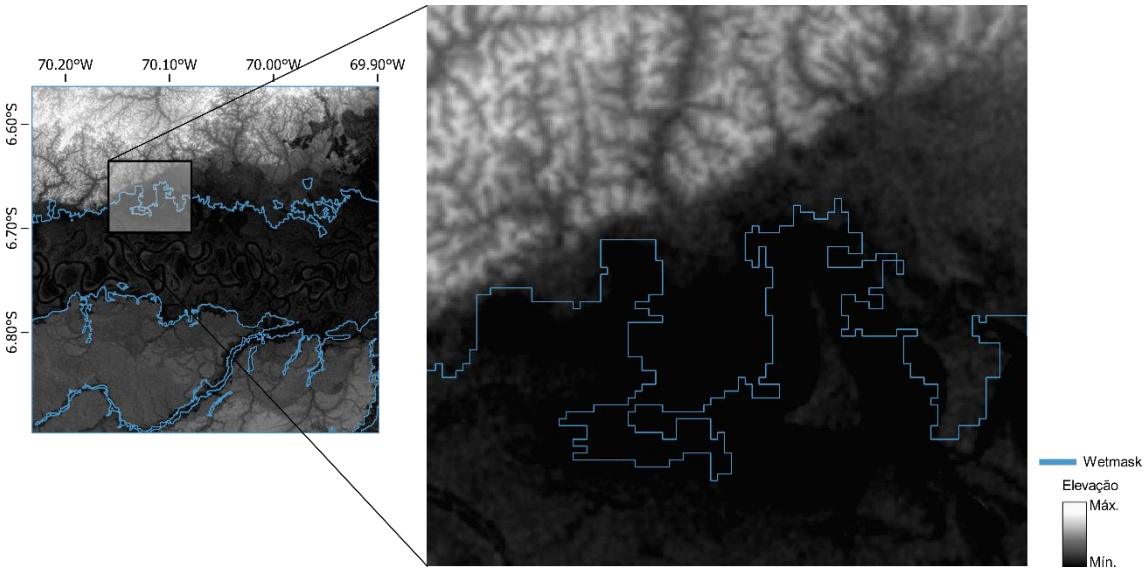
A bacia Amazônica apresenta extensas regiões de áreas alagáveis. O mapeamento desses ambientes é particularmente complexo devido a alguns fatores, como por exemplo: a extensão da bacia; a alta densidade da rede de drenagem; a variação espacial da precipitação e o fato de que as fases de águas altas e águas baixas dos rios ocorrem em diferentes épocas do ano ao longo da bacia (PIEDADE et al., 2013).

Dentre os mapeamentos de áreas alagáveis para a bacia Amazônica há o obtido por Hess et al. (2015). Os autores ampliaram o mapeamento feito por Hess et al. (2003) para toda a região da bacia hidrográfica abaixo de 500 m de altitude (denominada *lowland* pelos autores), o que representa 87% da área total da bacia. Esse mapeamento foi o primeiro para toda a região de *lowland* da bacia Amazônica, em uma resolução espacial considerada moderadamente alta para a época (100 m). Para geração do mapeamento, foram utilizados mosaicos de imagens SAR adquiridas pelo JERS-1 durante a época de cheia e seca entre os anos de 1995 e 1996. Os autores definiram as áreas alagáveis baseados em duas situações, sendo elas: i) as áreas que estavam alagadas no período de aquisição das imagens (durante a cheia e/ou a seca); ii) áreas que não apareceram como alagadas nas imagens, mas que apresentavam geomorfologia relativa à de áreas alagadas e eram próximas ou cercadas por áreas alagáveis. Assim, a máscara de áreas alagáveis foi criada a partir da segmentação das imagens, seguida de uma classificação não supervisionada. A extensão de áreas alagáveis mapeada por Hess et al. (2015) foi validada utilizando dados de sobrevoo da área de estudo e obteve uma acurácia de 93%. A área total estimada como alagável equivale a 14% de toda a área da bacia Amazônica.

Apesar de reconhecida a importância desse mapeamento, Hess et al. (2015) afirmam que a área alagável mapeada pode estar subestimada, visto que as imagens utilizadas não representaram situações como: os alagamentos ao longo de rios de pequena ordem e alagamentos em outras épocas do ano. Como a fase de águas altas (cheia) não ocorre igualmente ao longo dos rios da bacia Amazônica, aumenta-se a possibilidade de o mapeamento gerado não representar corretamente a extensão real das áreas alagáveis. Além disso, para obtenção desse produto não foram consideradas informações a respeito do relevo da região, o que pode contribuir para a presença de inconsistências no mapeamento, visto que os processos hidrológicos de uma bacia hidrográfica são influenciados pela topografia. Quando analisado o mapeamento gerado por Hess et al. (2015) em conjunto com o MDE-SRTM, é possível notar inconformidades entre os dados, em diferentes regiões, como exemplificado na Figura 2.3. Esta figura apresenta um recorte do mapeamento obtido por Hess et al. (2015)

visualizado sobre o MDE-SRTM, onde as áreas mais escuras na imagem possuem uma menor elevação. É possível observar que existem áreas de menor elevação que não estão incluídas na classe de áreas alagáveis, mas que devido à classificação das áreas vizinhas e da análise do MDE era esperado que estivessem incluídas no mapeamento. Isso indica a presença de incoerências nesse dado, apontando uma oportunidade de melhoria no mapeamento das áreas alagáveis para a região.

Figura 2.3 - Exemplo de inconsistência encontrada no mapeamento gerado por Hess et al. (2015) (contorno em linha vermelha) sobre o MDE-SRTM.



Fonte: Produção do autor.

2.4 Extração de atributos e Modelo Digital de Elevação (MDE)

A topografia exerce influência em processos hidrológicos, geomorfológicos e biológicos. Com relação aos processos hidrológicos, a topografia está diretamente relacionada às características do escoamento da água, como por exemplo a direção e a velocidade do fluxo (MOORE; GRAYSON; LADSON, 1991). Dessa forma, a representação da topografia se torna um aspecto importante a ser considerado em estudos hidrológicos. De maneira geral, os avanços no campo da modelagem hidrológica e hidráulica estão diretamente

associados à evolução dos dados topográficos, tanto em relação à sua obtenção quanto disponibilidade (TAVARES DA COSTA; MAZZOLI; BAGLI, 2019). Uma forma de representar a topografia é por meio de um Modelo Digital de Elevação (MDE). A partir do MDE é possível extrair diferentes parâmetros, também chamados de atributos, que podem ser utilizados, por exemplo, para representar um fenômeno relacionado ao escoamento. A ciência envolvida na quantificação da topografia e extração desses atributos é a geomorfometria (PIKE; EVANS; HENGL, 2009). Estudos relacionados ao mapeamento das áreas alagáveis têm considerado a informação da topografia por meio de diferentes atributos, como por exemplo índices topográficos (e.g. índice topográfico de umidade, índice de posição topográfica), declividade e diferentes tipos de curvatura do terreno (BWANGOY et al., 2010; MILLARD; RICHARDSON, 2015; DUBEAU et al., 2017; MARTINS et al., 2020).

O MDE pode ter diferentes formatos, sendo a grade regular retangular a mais comum. Nessa representação, cada célula, ou *pixel*, da grade recebe como atributo um único valor, referente a altimetria média de toda área representada pelo *pixel*. Dentre os diferentes MDEs disponíveis gratuitamente, há os obtidos pelo *Shuttle Radar Topography Mission* (SRTM). O SRTM foi um projeto realizado em conjunto pela *National Aeronautics and Space Administration* (NASA), a *National Imagery and Mapping Agency* (NIMA) e o *German Aerospace Center* (DLR) na Alemanha, em parceria com a Agência Espacial Italiana (ASI). Esse projeto produziu, por meio de interferometria SAR, MDEs com resolução espacial de 1 e 3 arco-segundo (aproximadamente 30 e 90 metros na linha do Equador, respectivamente). A missão ocorreu em fevereiro de 2000 e obteve informações da elevação da superfície para a região compreendida entre 60º de latitude norte e 56º de latitude sul, o que equivale a cerca de 80% da superfície terrestre (VAN ZYL, 2001; FARR et al., 2007).

O MDE-SRTM é um produto disponibilizado gratuitamente e amplamente utilizado, entretanto ele possui algumas limitações. Por se tratar de um MDE obtido por meio de interferometria SAR na banda C, o modelo é afetado pela vegetação. Além disso, esse dado apresenta limitação em sua resolução espacial, erros de aquisição, bem como regiões com ausência de informação

(*voids*) (VALERIANO; ROSSETTI, 2012). Tais limitações podem afetar a modelagem do relevo, bem como os diferentes produtos obtidos a partir do MDE. Nesse sentido, visando corrigir as falhas e limitações apresentadas pelo MDE-SRTM, foi desenvolvido o NASADEM (disponibilizado em fevereiro de 2020), considerado o sucessor do SRTM, com resolução espacial de 1 arco-segundo. O NASADEM é um MDE resultante de uma série de etapas, como o reprocessamento do SRTM, o preenchimento dos *voids* e a combinação de diferentes dados, como por exemplo: o *Advanced Spaceborne Thermal Emission and Reflection Radiometer* (ASTER), *Global Digital Elevation Model* (GDEM), *Ice*, *Cloud*, *and Land Elevation Satellite* (ICESat) e *Geoscience Laser Altimeter System* (GLAS).

2.5 Classificador *Random Forest* (RF)

O *Random Forest* é um algoritmo supervisionado de *machine learning* que pode ser utilizado para o processo de classificação. O RF funciona com base em um conjunto de árvores de decisão, também denominadas modelos. A combinação de diferentes árvores, em geral, leva a melhorias na acurácia da classificação (BREIMAN, 2001). Cada árvore de decisão é construída com base em um subconjunto de amostras, gerado a partir de um processo de reamostragem (sorteio aleatório com reposição) das amostras de treinamento. Esse conjunto de amostras resultantes da reamostragem é denominado amostras *bootstrap* (devido ao nome do método de reamostragem). A escolha de qual atributo irá compor cada nó da árvore ocorre com base em um conjunto menor de atributos pré-selecionados a partir de um sorteio aleatório, a partir das amostras *bootstrap*. A quantidade de atributos sorteada a cada nó é um parâmetro, geralmente denominado *mtry*. Outro parâmetro é a quantidade de árvores que irá formar a "floresta", comumente denominado *ntree*. Ambos os parâmetros devem ser definidos com base em análises prévias.

Como todas as árvores são geradas sem um processo de poda (processo que visa a simplificação das árvores obtidas) e a partir de um sorteio aleatório dos atributos em cada nó, elas tendem a ser diferentes umas das outras, o que

diminui a correlação entre elas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Como as árvores são diferentes, elas podem levar a resultados diferentes de classificação para uma mesma área. A classe final da unidade mínima a ser classificada (e.g. *pixel*) é definida com base no critério de maior quantidade de votos, considerando todas as árvores da floresta. Como a classe atribuída a um *pixel* é resultado de um processo baseado em uma frequência de votos, cada *pixel* terá associado a ele a proporção de vezes em que foi alocado a uma determinada classe. Essa proporção pode ser entendida como um estimador da probabilidade de o *pixel* pertencer a uma dada classe e pode ser usada como um indicador do grau de certeza do classificador em escolher uma determinada classe em detrimento de outras.

Em geral, as amostras coletadas são divididas em amostras de treinamento e amostras de testes. O primeiro conjunto amostral (comumente 70% do total) é usado para a obtenção das árvores que formarão o RF. O segundo conjunto amostral (30%) é utilizado para avaliação independente dos resultados da classificação. No processo de construção de cada árvore, uma parte das amostras de treinamento não são consideradas no conjunto de amostras *bootstrap*, podendo representar cerca de um terço das amostras de treinamento (BREIMAN, 2001). Essas amostras que não foram utilizadas no processo de construção de uma árvore, são denominadas amostras "*out-of-bag*" (OOB). Cada árvore possui o seu conjunto de amostras OOB e o RF utiliza essas amostras para gerar uma medida de erro, denominada erro OOB, que estima internamente a qualidade do modelo. O erro OOB é calculado da seguinte maneira: para cada amostra OOB é feita a predição considerando apenas as árvores em que a amostra não foi utilizada para o seu treinamento. Então, juntam-se as predições e, a partir da comparação entre a classe predita e a classe de referência para cada amostra, obtém-se a proporção de amostras OOB que foram classificadas incorretamente (LIAW; WIENER, 2002).

O classificador RF tem sido utilizado em diferentes aplicações com dados de sensoriamento remoto, tais como no mapeamento da cobertura do solo (GALIANO et al., 2012), da biomassa (KARLSON et al., 2015) e da suscetibilidade a deslizamentos (CATANI et al., 2013). O RF também tem sido

utilizado no mapeamento de áreas alagáveis (MILLARD; RICHARDSON, 2015; DUBEAU et al., 2017; MAHDIANPARI et al., 2017) e tem se destacado devido aos bons resultados e uma maior acurácia da classificação quando comparado com outros métodos de *machine learning*, como por exemplo *decision-tree* e *rule-based* (BERHANE et al., 2018b). Entre as vantagens do RF estão o fato de ser um classificador não-paramétrico, o que significa que para a sua aplicação não são necessárias pressuposições a respeito da distribuição das variáveis explicativas utilizadas; e o fato de permitir o uso de uma grande quantidade de dados, sejam eles categóricos ou numéricos (CATANI et al., 2013).

Além disso, o RF permite quantificar a importância de cada atributo (ou variável) presente no modelo, o que indica a relevância de cada atributo na classificação. A informação da importância pode ser utilizada para selecionar os atributos mais relevantes. Entretanto, a escolha de quais atributos são mais importantes baseada unicamente na informação da importância pode não ser uma escolha simples, uma vez que os valores de importância podem ser muito próximos, o que dificulta a decisão. Dessa forma, o valor da importância pode ser utilizado para ranquear os atributos e servir como base para métodos de seleção. A seleção dos atributos, também conhecida como seleção de variáveis, é o processo de reduzir o conjunto total de atributos àqueles mais relevantes, com base em um determinado critério de seleção (CAI et al., 2018). Selecionar um subconjunto de atributos importantes para predição do modelo pode facilitar a compreensão da relação entre os atributos e o fenômeno em estudo. O processo de seleção dos atributos é importante quando se trabalha com uma grande quantidade de dados, pois pode-se levar a uma redução no custo computacional associado ao processamento e armazenamento os dados. Há estudos em diferentes áreas que, utilizando grandes bancos de dados, compararam métodos que utilizam o RF para selecionar os atributos (DEGENHARDT; SEIFERT; SZYMCZAK, 2019; SPEISER et al., 2019). Na classificação de imagens de sensoriamento remoto, alguns trabalhos também têm aplicado o RF em combinação com o processo de seleção dos atributos (CORCORAN; KNIGHT; GALLANT, 2013; MILLARD; RICHARDSON, 2013, 2015; MEYER et al., 2017). Entretanto, o uso do RF com o propósito de se fazer uma seleção de atributos

não é muito comum no processo de classificação, o que indica oportunidades para o desenvolvimento de estudos na área.

2.6 Mapas de incerteza - entropia de Shannon

O desempenho de uma classificação é comumente avaliado a partir da construção de uma matriz de confusão, por meio da qual são calculados índices de acurácia para cada uma das classes (e.g. acurácia do produtor e do usuário), bem como índices globais, como a acurácia global. Entretanto, essa abordagem apresenta algumas limitações, como por exemplo a necessidade de um conjunto amostral representativo da verdade de campo e a incapacidade de informar sobre onde os erros ocorrem no mapeamento gerado (GONÇALVES et al., 2009; BROWN; FOODY; ATKINSON, 2008).

Uma forma de avaliar a distribuição espacial dos erros na classificação é por meio da elaboração de mapas de incerteza, construídos a partir de métricas que quantificam a incerteza associada à decisão do classificador em determinar a que classe um determinado *pixel* pertence. Dentre as diferentes medidas de incerteza, a entropia de Shannon é a mais utilizada (GONÇALVES et al., 2009). Nesse caso, a entropia indica a incerteza em uma dada distribuição de probabilidade. Como o processo de seleção de uma classe entre as várias existentes tem atrelado a ele a noção de incerteza, é possível utilizar a informação da probabilidade de um *pixel* pertencer a uma dada classe para o cálculo da entropia de Shannon, a ele associada.

A entropia de Shannon (H), pode ser expressa por:

$$H = -\sum_{i=1}^{N} p_i \times log_2(p_i) \tag{2.1}$$

onde: p_i representa a probabilidade associada a cada classe *i*; e *N* o número total de classes. Os valores de *H* variam de 0 a $\log_2 (N)$.

A análise da Equação 2.1 indica que a entropia assume o valor máximo quando a probabilidade de ocorrência de todas as classes for igual e assume valor mínimo (zero) quando existe certeza sobre qual classe o *pixel* pertence. Dessa forma, quanto maior o valor da entropia de Shannon, maior a incerteza atrelada à classificação, o que significa incerteza, por parte do classificador, na atribuição das classes selecionadas aos *pixels* da imagem.

3 MATERIAIS E MÉTODOS

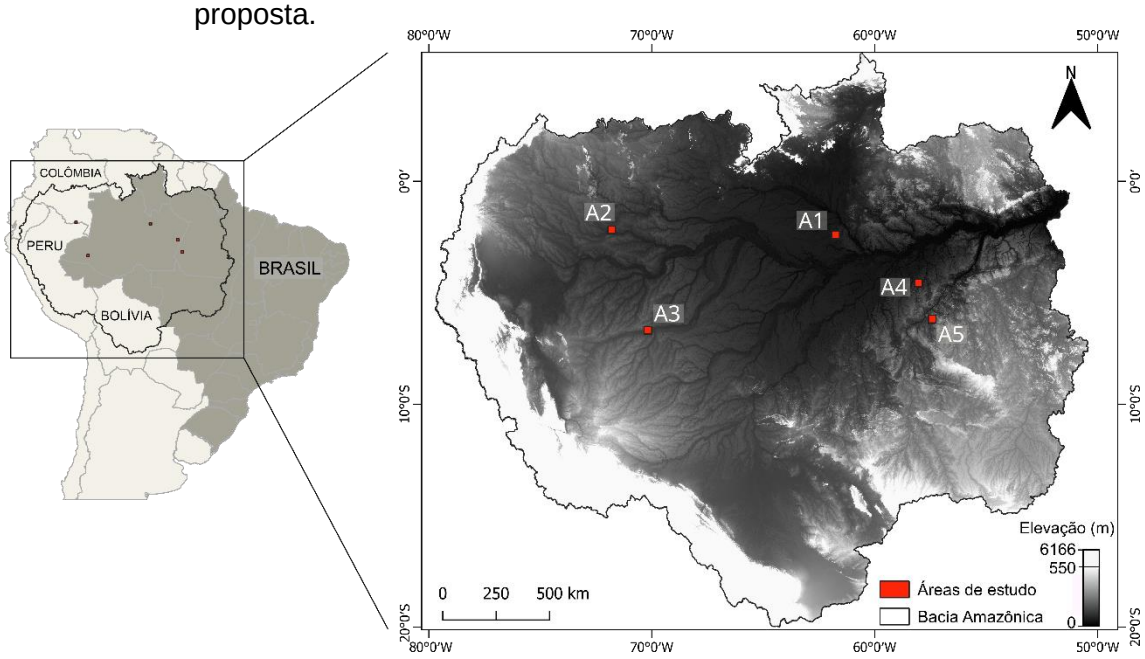
Nesta seção, será apresentada a metodologia utilizada para o desenvolvimento desta pesquisa. Inicialmente será apresentada a área de estudo, seguida dos materiais utilizados. Posteriormente, será feita uma descrição dos atributos utilizados na classificação e do processo de amostragem. Também será discutido como o classificador RF foi utilizado para gerar o mapeamento, assim como a metodologia adotada na seleção dos atributos. Por fim, serão descritos os métodos utilizados para avaliar o resultado da classificação.

3.1 Área de estudo

A área de estudo desta pesquisa compreende a bacia Amazônica. A região Amazônica é caracterizada por uma elevada precipitação, que se distribui de forma desigual espacial e temporalmente e apresenta aspectos topográficos que, quando combinados, propiciam a ocorrência de regiões alagáveis (JUNK et al., 2011), o que torna essa região uma área de interesse para esta pesquisa.

Devido sua grande extensão, foram selecionadas cinco áreas localizadas ao longo da bacia hidrográfica (Figura 3.1) para o desenvolvimento da metodologia, cujas características estão contidas na Tabela 3.1. Todas as áreas possuem tamanho de 1200 x 1200 *pixels* (aproximadamente 1300 km²) e foram escolhidas por serem regiões com diferentes padrões geomorfológicos, que resultam em diferentes feições de áreas alagáveis (diferentes tamanhos da planície de inundação). É importante destacar que em todas as áreas escolhidas foram identificadas inconsistências entre o mapeamento gerado por Hess et al. (2015) e o MDE-SRTM e que estas áreas apresentavam pequeno impacto antrópico, o que poderia limitar a extração dos atributos a partir do MDE.

Figura 3.1 - Localização das áreas selecionadas para aplicação da metodologia proposta.



MDE NASADEM para a bacia Amazônica. A escala da elevação está saturada em 550m.

Fonte: Produção do autor.

Tabela 3.1 - Características das áreas selecionadas para o desenvolvimento da pesquisa.

Área	Retângulo Envolvente	Variação altimétrica (m)*
1	2° 36' 33"S, 61° 48' 29"W a 2° 16' 33", 61° 28' 29"W	7 a 87
2	2° 23' 48"S, 71° 51' 25"W a 2° 03' 48" S, 71° 31' 25"W	15 a 175
3	6° 53' 48"S, 70° 13' 55"W a 6º 33' 48" S, 69° 53' 55"W	91 a 237
4	4° 46' 18"S, 58° 6' 25"W a 4° 26' 18"S, 57° 46' 25"W	5 a 181
5	6° 23' 48" S, 57° 28' 55"W a 6° 3' 48", 57° 8' 55"W	66 a 338

*A variação altimétrica é dada pelo valor de elevação mínima e máxima no MDE-SRTM.

Fonte: Produção do autor.

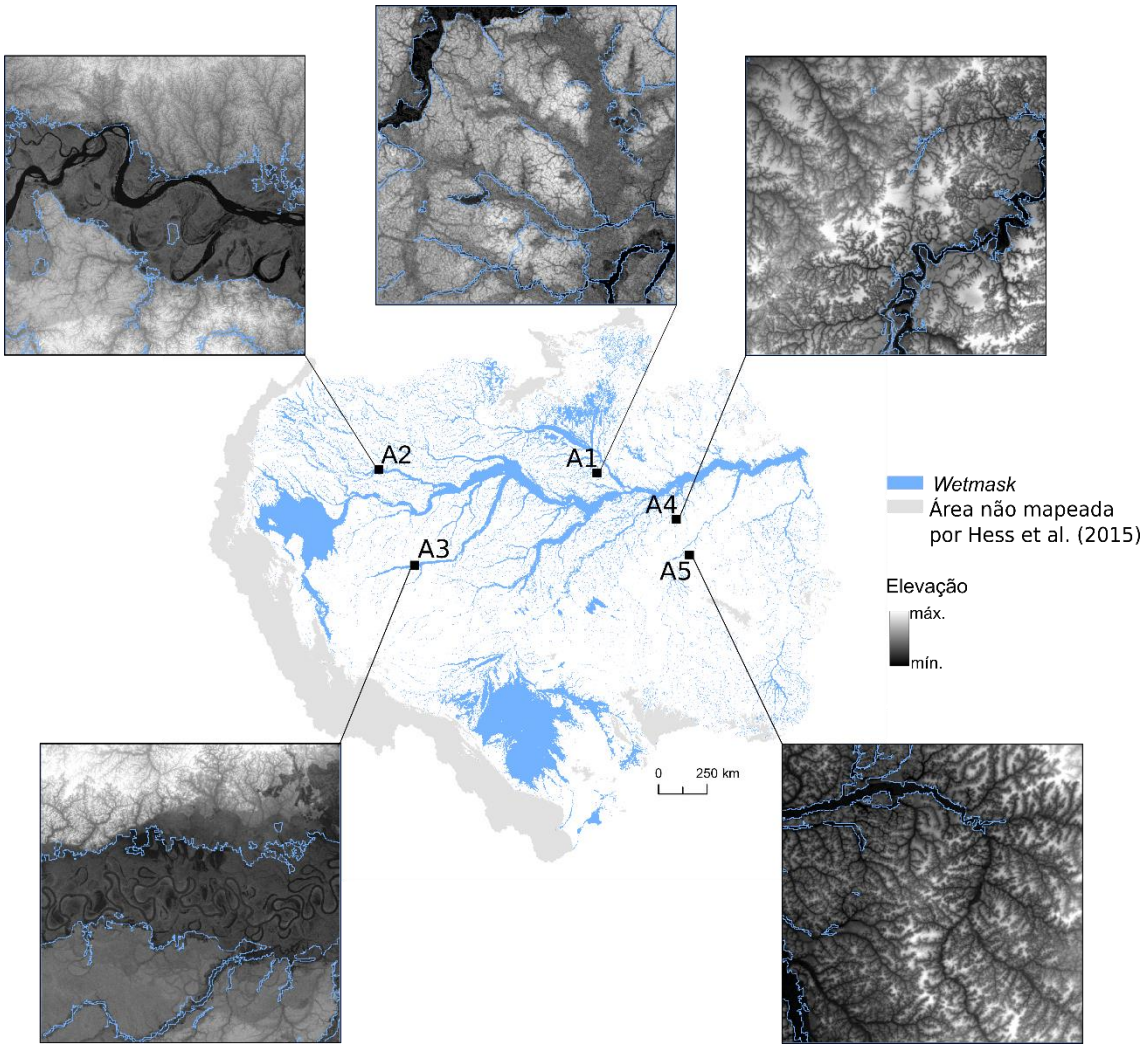
3.2 Materiais

O modelo digital de elevação utilizado nesta pesquisa foi o NASADEM, um produto gerado a partir do reprocessamento do MDE proveniente da missão SRTM, com resolução espacial de 1 arco-segundo (aproximadamente 30 metros na linha do Equador), disponível em: https://search.earthdata.nasa.gov/search.

A representação adequada da rede de drenagem é muito importante na definição das áreas alagáveis. No entanto, a representação cartográfica da hidrografia existente para a região Amazônica é falha, sendo composto por documentos cartográficos que representam o nível da copa das árvores e não as feições planialtimétricas no nível do solo (CORREIA, 2011). A fim de minimizar esse problema, Banon e Novo (2018) geraram uma rede de drenagem para toda a bacia Amazônica a partir da classificação por árvore de decisão de atributos extraídos do MDE-SRTM. Nesta pesquisa, foi utilizado um recorte desse dado de acordo com as áreas de interesse. Também foi utilizada a máscara de água derivada do SRTM, denominada SRTM *Water Body Data* (SWBD) e disponibilizada pelo *United States Geological Survey* (USGS). O dado SWBD representa corpos d'água visíveis na época da passagem do SRTM, considerando uma resolução espacial de 1 arco-segundo (aproximadamente 30 metros) (SLATER et al., 2006).

O mapeamento de áreas alagáveis elaborado por Hess et al. (2015) (disponível em: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1284) foi utilizado como referência na identificação de áreas alagáveis, sendo identificado neste trabalho como *wetmask*. É importante salientar que a *wetmask* orientou a coleta de amostra, não sendo considerado como verdade absoluta. O recorte da *wetmask* para cada uma das áreas de estudo selecionadas pode ser visualizada em conjunto com o MDE-NASADEM na Figura 3.2.

Figura 3.2- Contorno da *wetmask* para áreas de estudo sob o MDE-NASADEM.

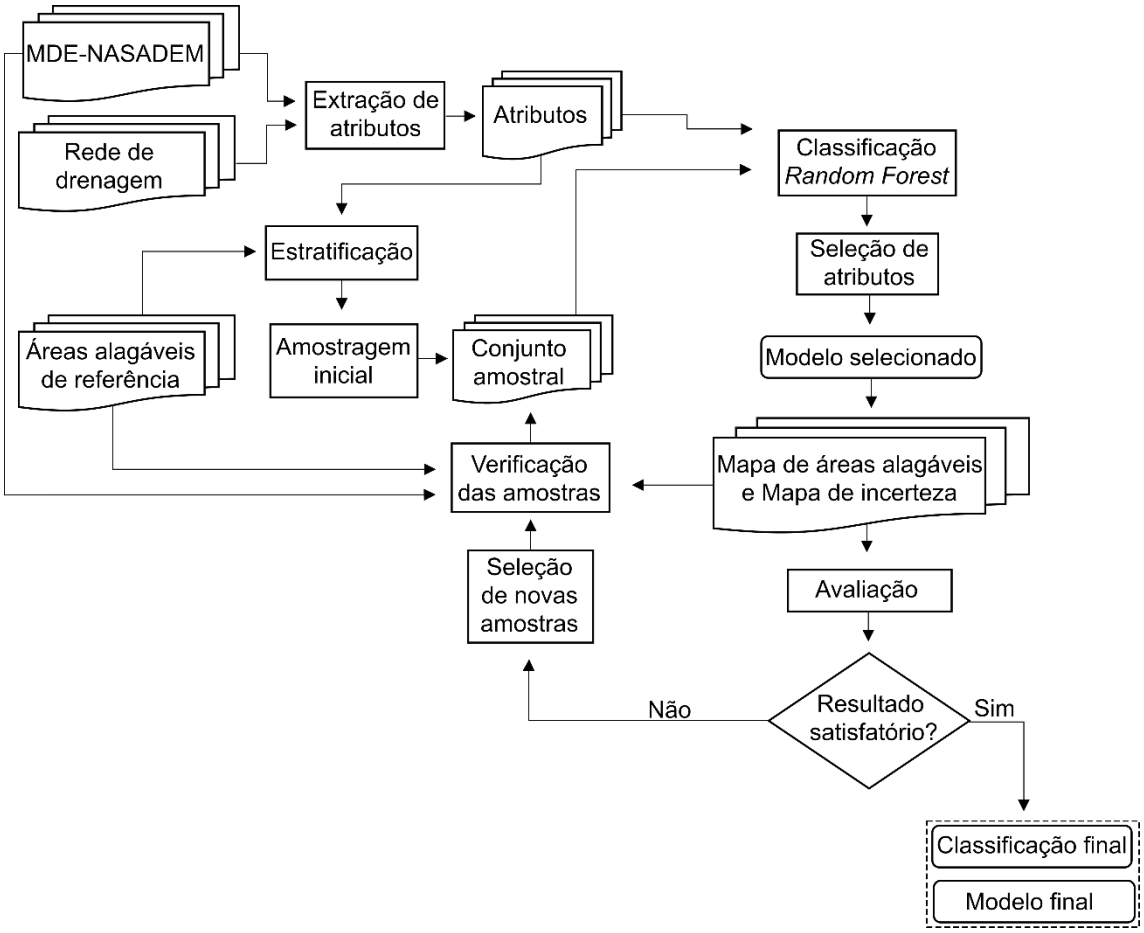


Fonte: Produção do autor.

3.3 Métodos

O fluxograma contendo as etapas seguidas nesta pesquisa é apresentado na Figura 3.3. De maneira geral, a metodologia compreende cinco etapas principais que serão explicadas a seguir: (i) extração dos atributos; (ii) amostragem; (iii) classificação utilizando o algoritmo *Random Forest*; (iv) seleção dos atributos e (v) avaliação da classificação.

Figura 3.3 - Fluxograma geral da metodologia.



Fonte: Produção do autor.

3.3.1 Extração de atributos

Nesta pesquisa, foram utilizados diversos atributos com potencial relação com o fenômeno do alagamento. A escolha dos atributos foi embasada em estudos prévios que utilizaram diferentes informações extraídas a partir de MDEs como dados de entrada para classificadores no mapeamento de áreas alagáveis (BWANGOY et al., 2010; DUBEAU et al., 2017; BERHANE et al., 2018b; MARTINS et al., 2020) e na extração de rede de drenagem (BANON, 2013; BANON et al., 2013). Os atributos calculados podem ser divididos em duas categorias: (i) morfométricos, que são extraídos diretamente do MDE; e (ii) hidrológicos, que para sua obtenção é utilizada a informação da direção de fluxo e rede de drenagem, além do MDE. Os atributos morfométricos descrevem características locais do relevo e geralmente são calculados a partir de uma

janela (*kernel*) de 3x3 *pixels*. Nesse trabalho, esses atributos foram obtidos considerando-se diferentes contextos de vizinhança espacial (tamanhos de janelas), assim como, aplicando-se técnicas de filtragem sobre o MDE e/ou sobre os atributos originais (janela 3x3) de modo a incorporar a informação contextual espacial mais ampla. Considerando os atributos morfométricos e hidrológicos, foram calculados ao todo 124 atributos.

3.3.1.1 Atributos morfométricos

Os atributos morfométricos utilizados foram: Declividade, Curvaturas, *Topographic Position Index* (TPI) e *Topographic Ruggedness Index* (TRI). Uma breve descrição desses atributos é apresentada a seguir:

a) **Declividade:** corresponde ao ângulo entre o plano horizontal e o plano tangente à superfície. A análise desse parâmetro permite determinar a distribuição das inclinações da superfície do terreno. Esse atributo influencia o processo de escoamento superficial da água, assim como a infiltração. Com relação ao fenômeno de alagamento estudado, é esperado que áreas mais planas tenham maior potencial ao alagamento. Nesta pesquisa, para aumentar a representatividade espacial desse atributo, a declividade foi representada de diferentes maneiras, calculada a partir de 3 diferentes métodos: (i) a partir do MDE modificado utilizando um filtro Gaussiano com janelas que variaram de 3x3 até 17x17 *pixels*; (ii) variando o tamanho da janela utilizada para gerar a declividade em 5x5, 7x7 e 9x9 *pixels*; (iii) aplicando um filtro Gaussiano na informação da declividade, obtida a partir do MDE original, considerando uma janela de 3x3 até 29x29 *pixels*.

b) **Curvaturas:** são parâmetros obtidos a partir da derivada de segunda ordem da altitude e descrevem características da forma do terreno. Há diferentes tipos de curvaturas a depender da orientação do plano de interseção em que a derivada foi calculada (Wood, 1996). Para o desenvolvimento da metodologia, foram consideradas as curvaturas: horizontal, vertical, máxima, mínima, média e longitudinal. Foram testadas

diferentes representações desses atributos, de acordo com a modificação do MDE a partir da aplicação do filtro Gaussiano com diferentes tamanhos de janelas. Foram testadas curvaturas obtidas a partir do MDE modificado aplicando-se um filtro 3x3 até um filtro 17x17 *pixels*.

c) **TPI:** é um atributo que representa a comparação do valor da elevação do terreno em um dado *pixel* com a elevação do terreno ao seu entorno. De acordo com Weiss (2001), o valor do TPI para um *pixel* do MDE é calculado a partir da elevação do *pixel* central subtraída pela elevação média de seus *pixels* vizinhos, como expresso em:

$$TPI = h - x_h(r) \tag{3.1}$$

onde, *h* é o valor da elevação em um dado *pixel*; x_h é a média da elevação nos *pixels* vizinhos, em que a vizinhança é determinada a partir de um raio *r*. Valores negativos do TPI indicam pontos com elevação menor que seu entorno; em contrapartida, valores positivos, indicam pontos com elevação maior do que sua vizinhança. Nesta pesquisa, foram testadas seis diferentes formas de representar o TPI, em que foi variado o tamanho do raio *r*, considerando uma janela 3x3, 5x5, 7x7, 9x9, 11x11 e 13x13 *pixels* utilizando o MDE original.

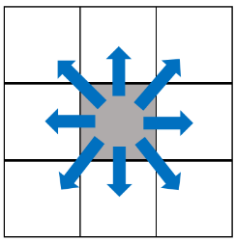
d) **TRI:** expressa a diferença de elevação entre células imediatamente vizinhas do MDE. De maneira geral, o valor do TRI para um determinado *pixel* é calculado a partir do somatório das diferenças de elevação entre o *pixel* central e seus vizinhos de acordo um determinado tamanho de janela (para mais detalhes, ver Riley, Degloria e Elliot, 1999). Esse atributo está relacionado às variações na elevação da área, possibilitando identificar regiões mais ou menos heterogêneas. A determinação deste atributo também depende do tamanho da janela a ser considerada. No caso, conforme feito no TPI, o TRI foi representado de seis diferentes maneiras, a partir do cálculo considerando diferentes tamanhos de janela, desde 3x3 até 13x13 *pixels* utilizando o MDE original.

3.3.1.2 Pré-processamento do MDE

Os atributos do tipo hidrológicos são baseados nas direções de fluxo e na rede de drenagem. Dessa maneira, para a extração desses atributos faz-se necessário realizar um pré-processamento no MDE a fim de obter a informação do caminho percorrido pelo fluxo.

As direções de fluxo podem ser representadas por meio de uma grade regular, conhecida por *Local Drain Direction* (LDD), em que a informação armazenada em cada célula é uma codificação que identifica a direção do fluxo da célula correspondente no MDE (JENSON, 1991). Há diferentes métodos para determinar as direções de escoamento da água a partir de um MDE (BUARQUE et al., 2009). A maior parte dos métodos trabalha com direção de fluxo única, ou seja, cada célula drena a água para uma de suas 8 células vizinhas (Figura 3.4).

Figura 3.4 - Oito possíveis direções de fluxo.



Fonte: Produção do autor.

Geralmente, a determinação da célula a receber o escoamento se dá por meio do critério de maior declividade. Esses aspectos são a base dos algoritmos chamados D8 (*determinist c eight-neighbors*), como o de Jenson e Domingue (1988). A diferença entre os diversos métodos está na forma pela qual os algoritmos tratam as regiões planas, as áreas de depressões (mínimos locais) e situações em que não existe uma declividade máxima única (BUARQUE et al., 2009).

O TerraHidro (ROSIM et al., 2003), *software* utilizado para obtenção das direções de fluxo e outros atributos, trata o problema de áreas planas por meio da abordagem de “cavar” o MDE. Para a resolução do problema de mínimos locais,

o *software* utiliza o método PFS (*Priority First Search*) que realiza correções no MDE de forma a simular o comportamento físico do escoamento superficial da água (JARDIM, 2017). Dessa maneira, por meio desses dois processos, o valor de elevação de alguns *pixels* do MDE é modificado de forma a garantir um fluxo contínuo. Assim, o MDE se torna hidrologicamente consistente para aplicação do algoritmo de determinação das direções de fluxo.

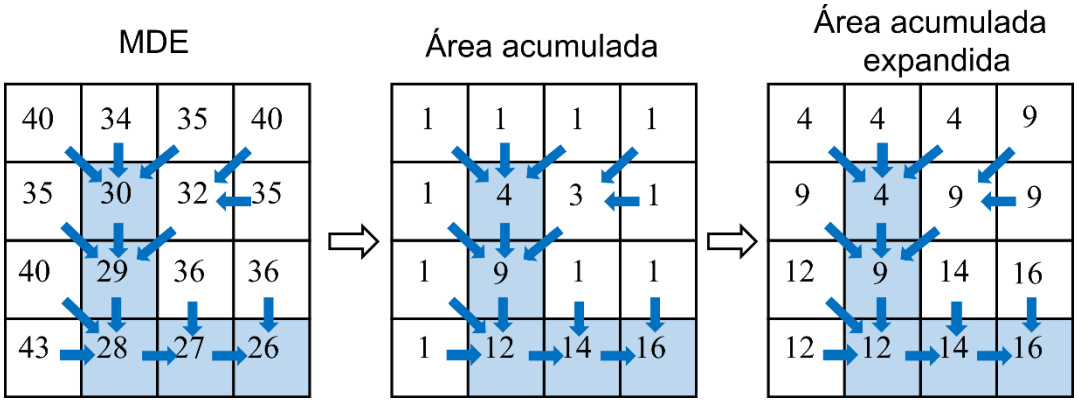
A partir da grade de direções de fluxo, pode-se obter uma grade regular que representa o fluxo acumulado, ou seja, toda a área a montante que drena água para a área representada por um dado *pixel*. Nesse sentido, é importante destacar que a obtenção do dado de área acumulada depende de uma correção de continuidade feita no MDE, a partir da obtenção das direções de fluxo. Assim, o fluxo acumulado é representado pela contagem de *pixels* que contribuem com o escoamento para o *pixel* em análise (JENSON, 1991).

3.3.1.3 Atributos hidrológicos

A partir do MDE, da grade de direção de fluxo, da informação de área acumulada e da rede de drenagem foram extraídos os atributos hidrológicos. Foram utilizados os atributos: Área acumulada expandida, *Height Above the Nearest Drainage* (HAND), Distância relativa à drenagem mais próxima (DIRDMP), Declividade relativa à drenagem mais próxima (DERDMP) e *Topographical Wetness index* (TWI). Segue uma breve descrição desses atributos:

a) **Área acumulada expandida:** esse atributo foi inicialmente proposto por Banon et al. (2019) e consiste em expandir a informação de área acumulada de um *pixel* pertencente à rede de drenagem para todos os pontos conectados a ele, como ilustrado na Figura 3.5. Dessa forma, quanto maior o valor desse atributo em um dado *pixel*, maior a evidência de que esse *pixel* está conectado a um trecho de rio onde é esperada uma vazão maior, pois o escoamento que chega a ele provém de uma grande porção de área. Esse atributo permite identificar áreas relacionadas às calhas principais dos grandes rios, regiões essas que possuem uma maior propensão ao alagamento.

Figura 3.5 - Esquema da obtenção do atributo área acumulada expandida.



As setas em azul representam as direções de fluxo obtidas a partir do MDE e as células em azul representam a rede de drenagem usada como referência para o cálculo do atributo. A área acumulada está representada em número de células.

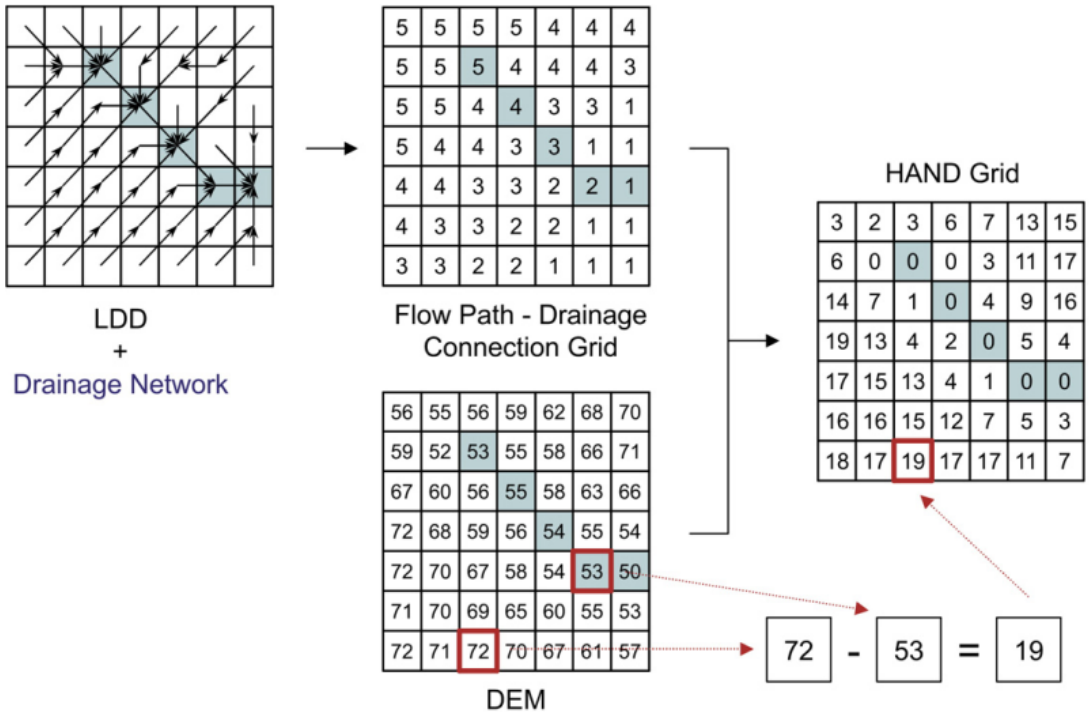
Fonte: Produção do autor.

b) ***Height Above the Nearest Drainage*** **(HAND):** é um algoritmo descritor do terreno que normaliza o MDE utilizando a rede de drenagem como referência. O algoritmo calcula a diferença entre a elevação de um *pixel* e a elevação do *pixel* pertencente ao canal de drenagem mais próximo, seguindo o caminho de fluxo (RENNÓ et al., 2008). O HAND determina uma distância vertical à drenagem, assim, de maneira geral, espera-se que áreas com menor valor de HAND sejam mais propensas ao alagamento; por outro lado, quanto maior o valor desse atributo, mais provável a área em questão pertencer ao ambiente não alagável. Nesse sentido, é um atributo com potencial de diferenciação entre uma área alagável e não alagável.

A Figura 3.6 exemplifica as operações realizadas para obtenção do que é chamado de MDE-HAND. Com base no MDE, na grade das direções de fluxo e na grade que representa a rede de drenagem, o algoritmo identifica o valor de elevação no MDE para um dado *pixel* e, seguindo o caminho de fluxo (a grade LDD), identifica o *pixel* pertencente à rede de drenagem mais próxima. Desse modo, obtém-se o valor de elevação para o *pixel* pertencente à rede de drenagem. No resultado da grade do HAND, os

pixels da superfície recebem como atributo o resultado da diferença dos valores de elevação encontrados no MDE, enquanto os *pixels* pertencentes à rede de drenagem recebem como atributo o valor zero.

Figura 3.6 - Processos para obtenção do HAND.



Fonte: Rennó et al. (2008).

Foram obtidas diferentes grades do descritor HAND, de acordo com as diferentes ordens de Strahler da rede de drenagem considerada. Dessa forma, foi calculado para cada área de estudo o HAND considerando as diferentes ordens da rede de drenagem. O HAND calculado utilizando a rede de drenagem com ordem maior ou igual a um, foi denominado HAND (Ord ≥ 1) e assim sucessivamente, até o HAND (Ord ≥ 7), que considerou apenas os trechos de rios com ordem igual ou superior a sete. Além disso, o HAND também foi calculado considerando como a drenagem a máscara de água SWBD, sendo denominado HAND SWBD. Dessa forma, no total, foram consideradas oito formas de representação do atributo HAND.

c) **Distância relativa à drenagem mais próxima (DIRDMP):** esse atributo foi calculado utilizando o método da distância euclidiana. Assim, foi considerada a menor distância em linha reta, entre o *pixel* analisado e o *pixel* pertencente à rede de drenagem. Assim como para a determinação do HAND, foram consideradas as distâncias relativas à rede de drenagem de acordo com diferentes ordens de Strahler dos rios. Sendo assim, foram gerados diferentes planos de informação do DIRDMP considerando inicialmente os trechos de rio de ordem maior ou igual a um, depois os trechos de rio de ordem maior ou igual a dois, até a drenagem menos ramificada considerada, com os rios de ordem maior ou igual a sete.

d) **Declividade relativa à drenagem mais próxima (DERDMP):** esse atributo foi definido como a declividade do terreno ao longo do caminho de fluxo entre um determinado ponto e a rede de drenagem mais próxima. No caso, o DERDMP foi obtido a partir da seguinte relação entre o HAND e o DIRDMP:

$$\mathrm{DERDMP} = \frac{HAND}{\mathrm{DIRDMP}} \tag{3.2}$$

Por simplificação, a distância à drenagem considerada foi a distância euclidiana. Foram obtidos sete diferentes planos de informação do DERDMP, de acordo com a menor ordem da rede de drenagem considerada nos atributos HAND e DIRDMP.

e) **TWI:** esse atributo expressa o potencial de uma área em acumular água. O índice foi desenvolvido por Beven e Kirkby (1979) e foi calculado de acordo com:

$$TWI = ln\left(\frac{\frac{A}{L}}{tan\beta}\right) \tag{3.3}$$

onde, β é o valor da declividade do terreno; A é o valor de área acumulada (em metros quadrados) e L é a largura do fluxo (em metros). No caso de a direção do fluxo ser no sentindo perpendicular, a largura do fluxo é considerada a resolução espacial. Por outro lado, quando a direção do fluxo é na diagonal, a largura do fluxo é dada pela raiz de dois ($\sqrt{2}$) multiplicada pela resolução espacial. O valor desse atributo será maior quanto maior for o valor de área acumulada para um dado *pixel*. Nesse sentido, o TWI possibilita identificar áreas com a tendência de acumular água.

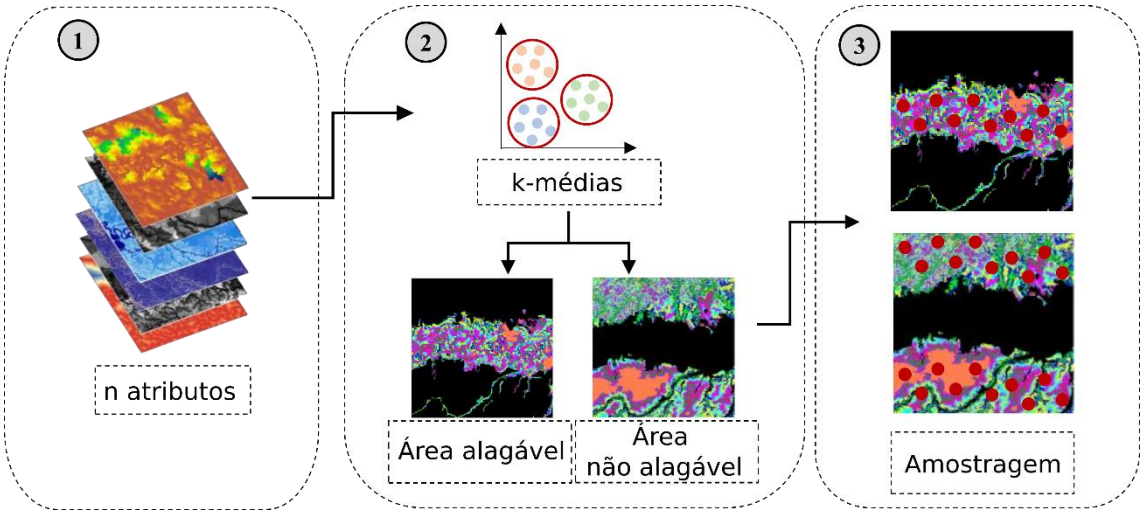
3.3.2 Amostragem

Para realizar a etapa da classificação, faz-se necessário fornecer amostras ao classificador. A condição ideal para um conjunto amostral é aquela em que as amostras são representativas das características da população (FOODY, 2002). Nesta pesquisa, a obtenção inicial das amostras, foi feita com base na *wetmask* (HESS et al., 2015). Com o objetivo de obter um conjunto amostral representativo considerando todos os atributos analisados, foi feito um processo de estratificação por meio da classificação não supervisionada k-médias (MACQUEEN, 1967). O k-médias permitiu delimitar regiões com padrões semelhantes de acordo com os atributos utilizados, para então realizar a amostragem com base nessas regiões. As amostras foram distribuídas de forma a caracterizar as duas classes em estudo: áreas alagáveis e áreas não alagáveis.

Como os atributos considerados possuem grandezas diferentes, houve a necessidade de se realizar a normalização deles antes de inicializar o método k-médias. Dessa forma, todos os atributos foram normalizados entre 0 e 1, utilizando os valores máximos e mínimos para cada área. É importante destacar que a normalização dos atributos foi feita apenas para o processo do k-médias. Todas as etapas seguintes desta pesquisa, utilizaram os atributos com seus valores originais. Após a normalização, o processo inicial de obtenção das amostras seguiu as seguintes etapas: (1) todos os n atributos foram empilhados,

formando o chamado *stack*; (2) o *stack* foi submetido ao classificador k-médias, que agrupou os *pixels* em 10 classes, em que os *pixels* pertencentes à mesma classe apresentam padrões reconhecido pelo algoritmo como semelhantes, e (3) foi feita a amostragem aleatória estratificada com base nas classes definidas pelo k-médias, distribuindo-se proporcionalmente as amostras nas 10 classes obtidas. O processo de classificação utilizando o k-médias foi feito duas vezes. Primeiramente foi feita uma classificação só para a região de área alagável e em seguida, foi feita outra classificação para as áreas de terra firme, ambas as áreas consideradas de acordo com a *wetmask*. A Figura 3.7 exemplifica o fluxo de atividades para esta etapa.

Figura 3.7 - Etapas para a realização da amostragem inicial.



Fonte: Produção do autor.

Como a *wetmask*, referência utilizada para realizar o processo de amostragem, apresenta inconsistências em algumas regiões, após o processo de amostragem inicial, as amostras passaram por uma inspeção visual, com o objetivo de eliminar pontos amostrais que, a partir da análise em conjunto com o MDE, foram identificados como pontos classificados erroneamente pela *wetmask*. Inicialmente, foram feitos testes utilizando um elevado número de amostras, mas esse procedimento não se mostrou uma boa alternativa, pois resultou em um gasto de tempo considerável para a revisão de todos os pontos, tornando assim

inviável a análise visual para cada uma das amostras. Dessa maneira, optou-se por coletar um conjunto amostral pequeno, o que possibilitou a inspeção visual de todos os pontos selecionados, de forma a garantir que as amostras representassem corretamente as classes em estudo. Nesta etapa, todos os pontos em que não havia certeza de qual classe atribuir foram descartados.

É importante compreender que as etapas que compõe este trabalho formam um processo iterativo. Dessa maneira, o passo inicial foi dado com base em um conjunto amostral, mas após a primeira rodada do algoritmo, pontos poderiam ser removidos e outros acrescentados. Essa modificação se deve ao fato de que a cada iteração foram identificadas regiões que precisavam ser melhor representadas, então foram coletadas mais amostras. A decisão pela exclusão de amostras ou inclusão de novas amostras foi feita com base nos resultados da classificação. Esses critérios serão abordados no Item 3.3.4.

3.3.3 Classificação *Random Forest*

O algoritmo RF foi implementado utilizando o pacote "*randomForest*" (LIAW; WIENER, 2002) no *software* R. Nesta implementação do RF, faz-se necessário determinar dois parâmetros: o número de árvores (*ntree*) e a quantidade de atributos sorteados a cada nó da árvore (*mtry*). Para determinar o *ntree*, foi analisada a variação do erro OOB considerando diferentes números de árvores, baseado no modelo completo contendo 124 variáveis. Com base em testes preliminares, notou-se que o erro OOB reduziu consideravelmente depois das 50 árvores, estabilizando a partir de 1000 árvores. Assim, optou-se por trabalhar com um valor de 1000 árvores. Para classificações utilizando o algoritmo RF, o valor padrão utilizado para o parâmetro *mtry* é $\sqrt{\mathrm{p}}$, sendo p o número de variáveis. Entretanto, isso deve ser avaliado caso a caso, sendo recomendado então determinar o valor ótimo para cada estudo (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Em testes preliminares, foi fixado o valor de *ntree* em 1000 e então rodou-se o algoritmo RF, variando o valor do *mtry* e analisando o erro OOB para cada um dos modelos. Com base nesses testes, foi escolhido o valor de 22 para o *mtry*, devido ao menor erro encontrado. Do total de amostras

coletadas, 70% foram utilizadas para treinar o modelo e 30% para validar. Optou-se por não amostrar pontos na Área 5 com o objetivo de analisar a resposta do modelo em uma área em que não foram coletados pontos de treinamento.

3.3.3.1 Importância dos atributos

O *Random Forest* é um algoritmo que permite gerar a importância dos atributos presentes no modelo. Ao quantificar essa importância, é possível compreender a relevância de cada atributo na predição, auxiliando assim a interpretação do resultado. Nesta pesquisa, foi utilizada a métrica *Mean Decrease Accuracy* (MDA), também conhecida como medida de permutação, para determinar a importância dos atributos. O MDA foi aplicado, pois é uma medida de importância frequentemente utilizada (BELGIU; DRAGUT, 2016), inclusive no contexto de seleção de atributos por meio do RF (MILLARD; RICHARDSON, 2013; FOX et al., 2017), e está diretamente relacionada com a acurácia da classificação. Como exposto em Genuer, Poggi e Tuleau-Malot (2010), o MDA é calculado da seguinte maneira: para o conjunto de amostras OOB (as amostras que não foram usadas para construção da árvore) associado a cada árvore t é calculado o erro OOB ($errOOB_t$). Então, os valores da variável X^j nas amostras OOB são permutados aleatoriamente e em seguida, é calculado novamente o erro OOB para cada árvore, denominado como $\widehat{errOOB_t^j}$. Por fim, o MDA é obtido de acordo com a expressão:

$$\mathrm{MDA}^j = \frac{1}{ntree} \sum_{t=1}^{ntree} \left(\widehat{errOOB_t^j} - errOOB_t \right) \tag{3.4}$$

onde, *ntree* é a quantidade de árvores utilizadas no modelo RF. A partir da Equação 3.4, pode-se notar que as variáveis que, ao sofrerem permutações aleatórias de seus valores, resultaram em um alto erro de classificação (erro OOB), terão um maior valor na métrica de importância. Por outro lado, quando o erro de classificação após a permutação aleatória dos valores de uma

determinada variável não apresentou valor muito diferente daquele antes da alteração, mostra que a perturbação dos valores da variável não foi muito significante, indicando assim que essa variável não foi tão importante para a classificação.

Como a obtenção da métrica MDA depende de um processo aleatório, foi analisada a variação dos valores do MDA para cada um dos atributos, ao longo de 1000 iterações do RF. Para isso, foram gerados 1000 modelos RF, fixando os valores de *ntree* e *mtry* em 1000 e 22, respectivamente. Assim, para cada atributo determinou-se o valor do MDA em cada um dos 1000 modelos. Além disso, o valor do MDA foi utilizado para ranquear os atributos presentes no modelo, de acordo com a importância. Então, também foi analisada a estabilidade do ranqueamento, verificando quais as posições ocupadas pelos atributos nos 1000 modelos gerados.

3.3.3.2 Seleção dos atributos

A utilização de uma grande quantidade de atributos pode estar associada a um alto custo computacional relacionado à extração, armazenamento e processamento de uma grande quantidade de dados. Além disso, quando analisada a importância relativa a cada atributo, é comum notar que dentre os vários atributos utilizados, apenas uma pequena quantidade exerce uma influência considerável na resposta do modelo (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Nesta pesquisa, o objetivo foi, além de gerar uma classificação das áreas alagáveis com um valor satisfatório de acurácia, selecionar um conjunto mínimo de atributos para compor um modelo reduzido, suficientemente robusto para realizar a predição. Além disso, pensando na utilização da metodologia proposta para ampliação do mapeamento para toda a região da bacia Amazônica, buscou-se reduzir a quantidade de atributos utilizada uma vez que o processo de extração de alguns atributos tem um alto custo computacional, além de que a obtenção dos dados para essa grande área, demandaria também um alto custo de processamento e armazenamento dos dados.

Para a seleção dos atributos nesta pesquisa, foi aplicado um método de seleção baseado na quantificação da importância dos atributos a partir da métrica MDA, proveniente do *Random Forest*. Inicialmente, a análise da estabilidade do ranqueamento das variáveis (obtido a partir da métrica MDA) foi utilizada para realizar uma pré-seleção dos atributos morfométricos que foram representados a partir de diferentes filtros ou foram calculados considerando janelas de diferentes tamanhos. Esta pré-seleção não foi realizada sobre os atributos hidrológicos calculados considerando diferentes ordens da rede de drenagem, pois acredita-se que esses atributos representem informações contextuais complementares. Sendo assim, essa pré-seleção foi feita de modo que apenas uma representação de cada um dos atributos morfométricos fosse escolhida. A escolha ocorreu com base na maior estabilidade no ranqueamento e na maior importância, ou seja, foi escolhido o atributo cuja posição variou menos no ranqueamento e estava associado a uma maior importância dentre os demais.

Posteriormente, considerando um modelo com os atributos morfométricos pré-selecionados e todos os atributos hidrológicos, os parâmetros do RF foram novamente otimizados. Optou-se por se trabalhar com o mesmo número de árvores (*ntree* = 1000), pois esse valor já estava associado a uma estabilização do erro OOB. O parâmetro *mtry* foi ajustado, devido à redução no número de atributos utilizados. Então, com o objetivo de realizar a seleção dentre os atributos morfométricos pré-selecionados e os atributos hidrológicos, foi implementado o seguinte procedimento:

1. Ranqueamento das variáveis em função da métrica de importância MDA. Para obter um ranqueamento estável, foi feita a média do MDA para cada atributo em 1000 iterações do RF.
2. Aplicação da técnica *stepwise backward*, em que as variáveis foram retiradas uma a uma, seguindo a ordem crescente de importância. Ou seja, gerou-se o RF com todos os p atributos, depois com p-1, até um RF com um único atributo (sendo esse atributo considerado mais o importante no ranqueamento inicial). A

cada passo foi calculada a acurácia (com base nas amostras de validação) e o erro OOB referente a cada modelo.

3. Todos os p-1 modelos gerados foram comparados a partir de um critério de avaliação. Optou-se por escolher o modelo com maior acurácia e menor erro OOB. Em caso de empate entre os modelos, foi escolhido o modelo mais simples, ou seja, aquele composto por menos atributos. Então, os atributos presentes no modelo escolhido foram os atributos selecionados.

Essa metodologia iterativa de seleção dos atributos baseada na informação da importância das variáveis também foi aplicada em outros trabalhos. Por exemplo, na área de bioinformática, Díaz-Uriarte e Alvarez de Andrés (2006) a utilizaram em problemas de seleção de gene. Já Genuer, Poggi e Tuleau-Malot (2010) utilizaram essa metodologia em problemas mais gerais relacionados à utilização de grandes volumes de dados. Ambos os trabalhos utilizaram como critério para escolha do melhor modelo apenas o erro OOB. Em relação a problemas de mapeamento de áreas alagáveis utilizando imagens de sensoriamento remoto, Millard e Richardson (2013) também basearam a seleção dos atributos na técnica semelhante ao *stepwise backward*.

Devido ao caráter iterativo desta pesquisa, toda vez que um novo atributo era considerado ou as amostras eram modificadas, todo o processo de seleção de atributos era refeito.

3.3.4 Avaliação

3.3.4.1 Avaliação da classificação durante o processo iterativo

Inicialmente, a avaliação da classificação gerada durante o processo iterativo de construção dos modelos foi feita por meio da visualização do mapa de incerteza calculado a partir da Entropia de Shannon. Ao analisar a entropia espacializada, foi dedicada uma maior atenção às regiões que apresentaram agrupamentos de valores elevados de entropia. Essas regiões representam áreas em que as

regras de decisão não foram capazes de identificar com alto grau de certeza se as mesmas eram ou não áreas alagáveis. Nesses casos, foram amostrados mais pontos nessas regiões com o intuito de melhorar a representatividade dessas áreas. Quando foram identificadas amostras com um elevado valor de entropia e o analista não tinha certeza em relação à sua classificação, elas foram removidas do conjunto amostral.

O processo de classificação, coleta e exclusão de amostras foi repetido até o momento em que a modificação do conjunto amostral não resultou em mudanças expressivas no mapa de áreas alagáveis obtido nem na distribuição espacial da entropia. Neste caso, foi determinado o modelo final, com os atributos selecionados e o mapa final de áreas alagáveis, juntamente com o mapa de incerteza da classificação.

3.3.4.2 Avaliação final dos três modelos gerados

Ao final do processo de redução do modelo original, foi feita a comparação dos 3 modelos gerados, sendo eles: (i) o modelo com todos os 124 atributos; (ii) o modelo com os atributos morfométricos pré-selecionados e os atributos hidrológicos, e (iii) o modelo reduzido resultante da seleção dos atributos. A comparação entre os modelos foi feita a partir da análise visual dos mapeamentos gerados, assim como do mapa de incerteza. Além disso, a partir da matriz de confusão, obtida utilizando as amostras de validação, foram calculadas a acurácia global e os índices sensibilidade, precisão e *F1 score:*

$$\text{acurácia global} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}} \tag{3.5}$$

$$\text{sensibilidade} = \frac{\text{VP}}{\text{VP} + \text{FN}} \tag{3.6}$$

$$\text{precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}} \tag{3.7}$$

$$\text{F1 score} = 2 \times \frac{\text{precisão} \times \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \tag{3.8}$$

onde VP representa os *pixels* verdadeiro positivo (classificados corretamente como áreas alagáveis); VN indica os verdadeiros negativos (classificados corretamente como áreas não alagáveis, no caso terra firme); FP são os falsos positivos (classificados incorretamente como áreas alagáveis) e FN são os falsos negativos (classificados incorretamente como terra firme).

Outra métrica utilizada para avaliar e comparar o desempenho dos modelos foi o índice *Area Under the Curve* (AUC), obtido a partir da curva ROC (*Receiver Operating Characteristics*). A curva ROC é obtida traçando um gráfico da taxa de verdadeiros positivos *versus* taxa de falsos positivos. A área abaixo da curva ROC é denominada AUC e varia de 0 a 1. Quanto mais próximo de 1 for o valor da métrica, melhor o desempenho do modelo.

Para avaliar se a métrica acurácia global variou de maneira estatisticamente significativa entre os modelos comparados, foi calculado o teste de McNemar (FOODY, 2004). Esse é um teste não paramétrico que compara se as proporções analisadas entre dois grupos são iguais estatisticamente. Nesse teste, a hipótese nula considerada é a de que as proporções são iguais. Dessa forma, caso a hipótese nula seja aceita considerando um determinado nível de significância, as proporções não diferem estatisticamente entre si. O teste de McNemar é feito com base em informações obtidas a partir da matriz de confusão de cada um dos modelos comparados.

Por fim, foi feita uma comparação entre a *wetmask* e a classificação obtida a partir do modelo selecionado reduzido. A partir da sobreposição entre os mapas, foram analisadas as áreas em concordância e discordância entre as classificações.

4 RESULTADOS E DISCUSSÃO

Nesta seção, serão apresentados os resultados obtidos e as discussões feitas em relação a eles. Inicialmente, será abordado o conjunto amostral utilizado, seguido da análise da métrica MDA usada para ranquear os atributos. Depois, serão expostos os resultados obtidos a partir do processo de seleção dos atributos e a análise do efeito da redução dos atributos no mapeamento. Por fim, serão apresentadas as comparações entre a classificação final obtida e a *wetmask*.

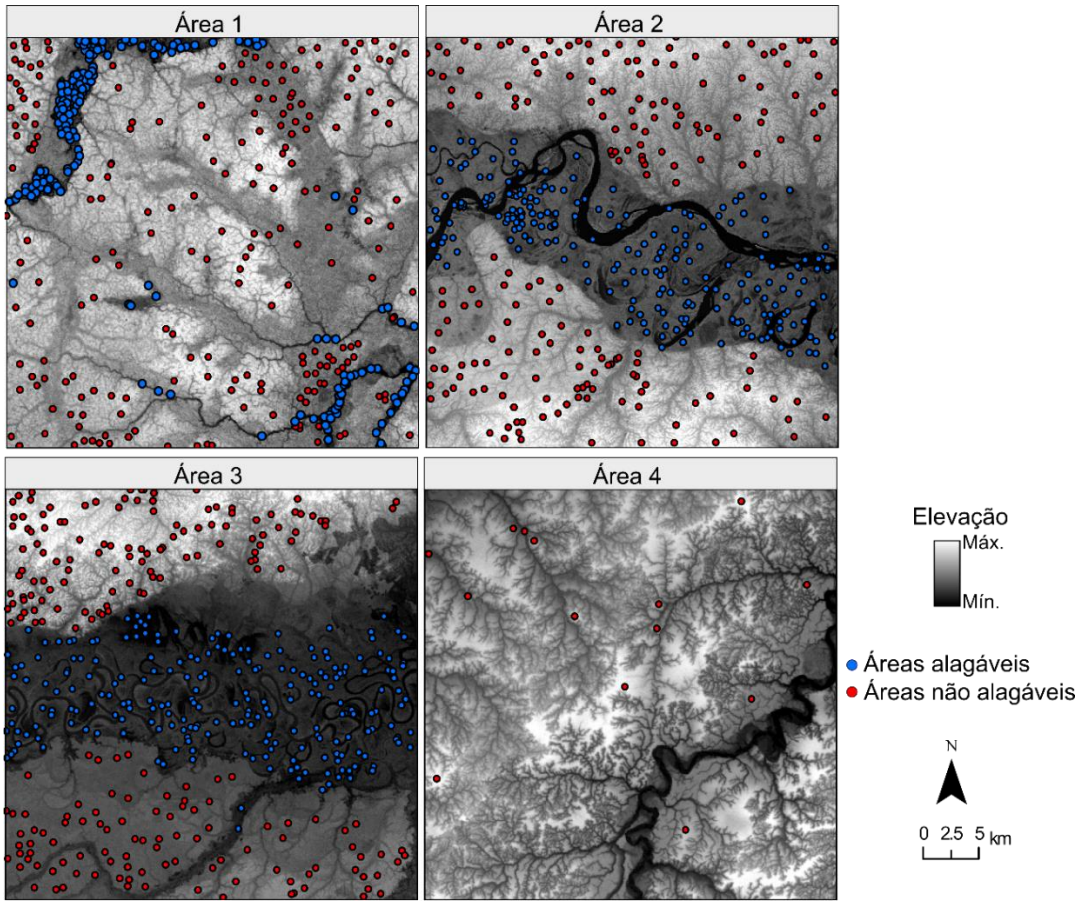
4.1 Amostragem

Durante o processo de amostragem, foram consideradas apenas as Áreas 1, 2, 3 e 4 (Figura 3.2). Optou-se por não amostrar pontos na Área 5 a fim de avaliar a aplicação do modelo em uma área que não tivesse sido utilizada na fase de treinamento. Dessa forma, inicialmente foi amostrado um conjunto de pontos na Área 3, escolhida por possuir uma grande área alagável de acordo com a *wetmask*, sendo assim, uma área mais representativa. Essa primeira amostragem foi utilizada para gerar a classificação RF utilizando-se todos os 124 atributos. Em seguida, foram avaliados os produtos gerados da classificação (o mapa de áreas alagáveis e o mapa de entropia) para as Áreas 1, 2, 3 e 4. A partir dessa avaliação preliminar, notou-se que o modelo gerado não conseguiu representar adequadamente as feições presentes nas áreas não amostradas. Então, foram acrescidos pontos amostrais provenientes da Área 2, gerando-se novamente a classificação RF e avaliando-se os resultados da classificação para todas as áreas analisadas. Também houve a necessidade de acrescentar pontos amostrais das Áreas 1 e 4 para que fosse obtida uma classificação considerada satisfatória para todas as áreas analisadas. A distribuição dos pontos nessa amostragem inicial para as Áreas 1, 2 e 3 foi feita com base nos agrupamentos gerados a partir do k-médias, como descrito na metodologia (Seção 3.3.2). Por outro lado, a Área 4 contribuiu com poucos pontos amostrais da classe não alagável apenas para corrigir problemas de superestimação da área alagável. É importante lembrar, que o processo de classificação foi composto por diversas

iterações. A cada iteração, foram avaliados o mapa de áreas alagáveis e o mapa de incerteza, sendo o conjunto amostral modificado à medida que fosse constatada a necessidade de se amostrar pontos em outras regiões (mal representadas ou com alta incerteza) ou de se remover determinadas amostras cuja classe não pudesse ser confirmada por uma inspeção visual.

O processo iterativo de classificação e modificação do conjunto amostral seguiu até o momento em que não foram identificadas mudanças expressivas no mapeamento de áreas alagáveis e no mapa de entropia, em relação à classificação obtida na iteração anterior. Por fim, na última iteração foram utilizadas 1256 amostras, sendo 630 da classe de área alagável e 626 da classe de área não alagável distribuídas ao longo de quatro das cinco áreas de estudo, como apresentado na Figura 4.1. Do total de amostras, 70% foi utilizado para treinamento e 30% para validação do modelo.

Figura 4.1 - Distribuição espacial dos pontos amostrados sobre o MDE - NASADEM.

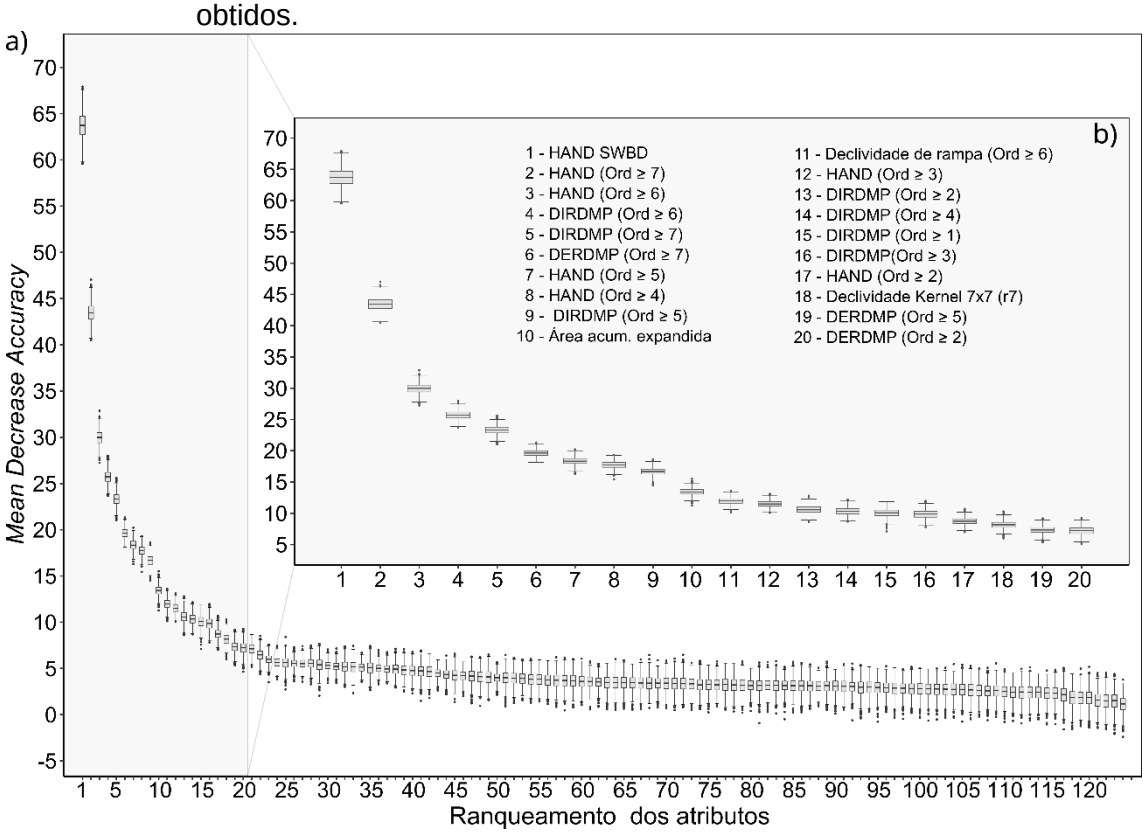


Fonte: Produção do autor.

4.2 Análise da estabilidade da métrica de importância *Mean Decrease Accuracy* (MDA)

A métrica MDA foi utilizada como uma medida da importância dos atributos e serviu como base para ranqueá-los em cada modelo RF avaliado. Inicialmente foi analisado o comportamento do valor do MDA para todos os 124 atributos ao se rodar diversas vezes um modelo RF, fixando os parâmetros *mtry* e *ntree* nos valores ótimos definidos (22 e 1000, respectivamente). Assim, foram obtidos 1000 modelos RF e observado, para cada atributo, o valor do MDA, em cada um dos modelos, como mostrado na Figura 4.2. Nessa figura, o ranqueamento dos atributos exibido no eixo x foi definido a partir do valor médio do MDA ao longo dos 1000 modelos obtidos.

Figura 4.2 - Comportamento dos valores de MDA baseado nos 1000 modelos RF obtidos.



a) Boxplots da métrica MDA para os 124 atributos. b) Destaque para os vinte atributos mais importantes.

Fonte: Produção do autor.

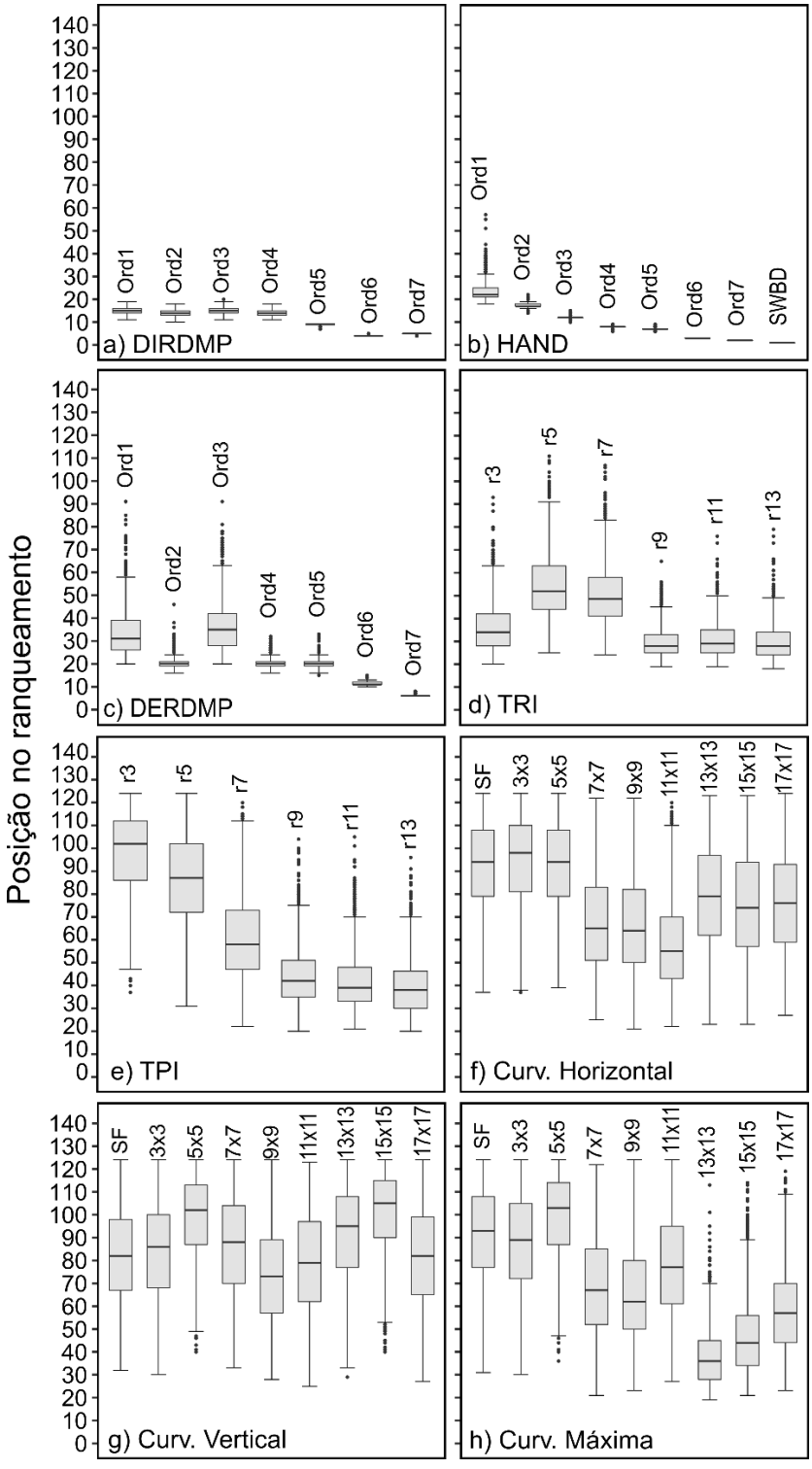
A partir da Figura 4.2, nota-se que há uma diferença considerável na magnitude do MDA dos dois primeiros atributos (HAND SWBD e HAND Ord ≥7) em relação aos demais, indicando que em todas as rodadas, o valor da importância atribuído a eles foi sempre alto, quando comparado com o valor atribuído aos outros atributos. Nesse caso, diante da configuração adotada nesta análise, ou seja, do conjunto amostral utilizado e dos atributos testados, esses dois primeiros atributos exerceram grande influência na classificação e possuem um peso maior na importância dos atributos, quando comparados com os demais. O atributo posicionado em primeiro lugar no ranqueamento (HAND SWBD) apresentou a maior variação nos valores do MDA, variando de 60 a 68. Essa grande variação dos valores de MDA também pode ser observada em atributos menos importantes, não havendo, portanto, relação entre a variação e a importância em si do atributo analisado.

Ainda sobre a Figura 4.2, quando analisada a posição de um *boxplot* em relação aos outros, nota-se que os seis primeiros colocados, ou seja, os seis atributos considerados mais importantes, apresentaram as maiores diferenças relacionadas à posição dos *boxplots*. Por outro lado, a partir da 13ª posição, fica evidente que vários *boxplots* estão sobrepostos, indicando que houve muita semelhança entre as distribuições dos valores de MDA para esses atributos. O fato de haver muita sobreposição entre os *boxplots*, indica que esses atributos apresentaram aproximadamente a mesma importância para as diferentes classificações. Assim, pode-se concluir que, para os diferentes modelos RF obtidos, esses atributos podem trocar de posições no ranqueamento entre si, ou seja, não se pode afirmar que um atributo tem maior ou menor importância que outro.

Para verificar como a variação do MDA se refletiu no ranqueamento dos atributos, foi avaliada a estabilidade do ranqueamento. A Figura 4.3 apresenta os *boxplots* referentes à distribuição das posições ocupadas por cada um dos atributos nos 1000 modelos RF obtidos. Para facilitar a visualização e análise, os atributos foram separados por grupos. Um grupo foi formado por atributos que, de maneira geral, expressam uma mesma informação diferindo na forma com que essa informação foi representada. Por exemplo, o grupo “HAND” foi

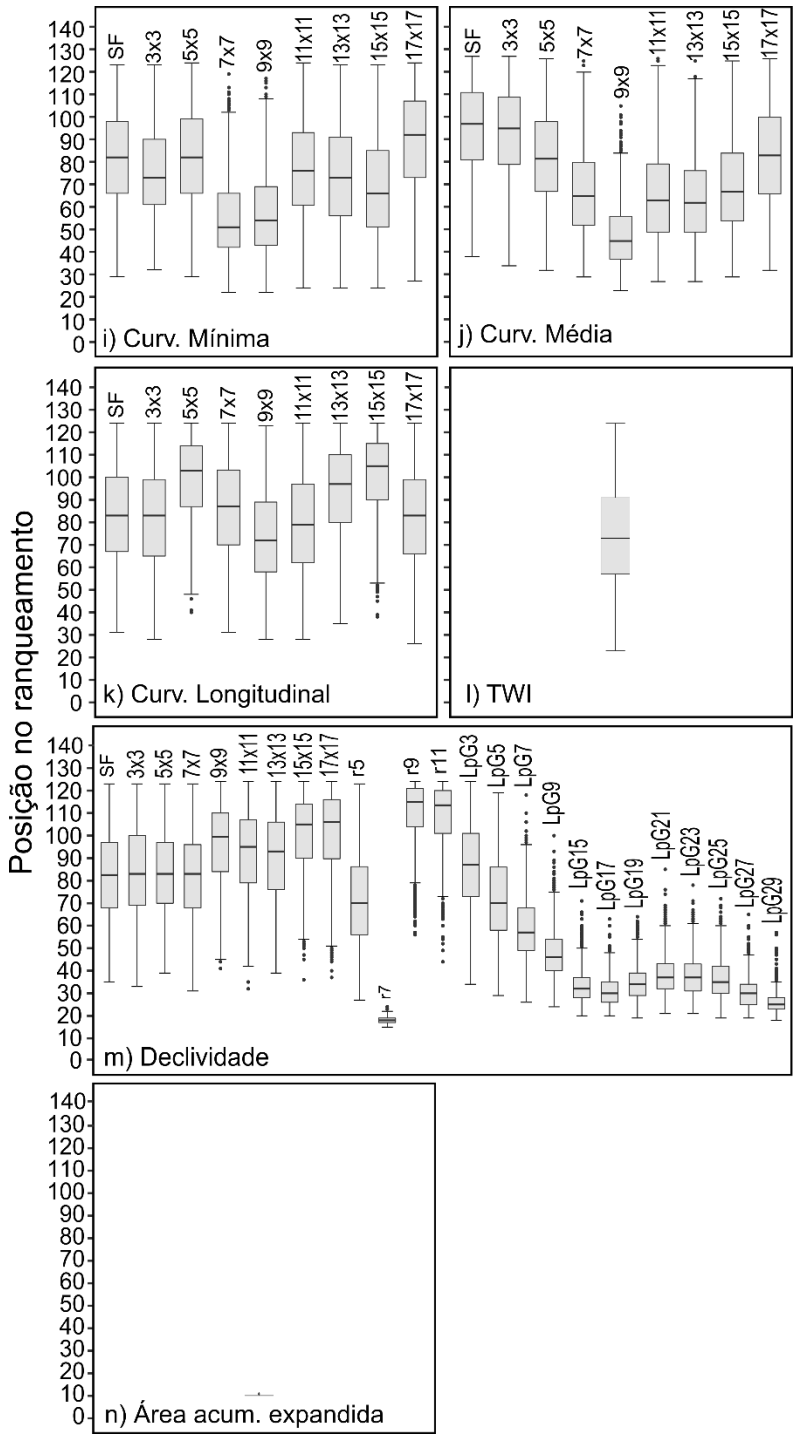
formado por 8 atributos, sendo eles, as diferentes maneiras com que o HAND foi representado neste estudo, como descrito na Seção 3.3.1.3.

Figura 4.3 - *Boxplots* da posição ocupada pelos atributos no ranqueamento em 1000 modelos RF obtidos.



continua

Figura 4.3 – Conclusão.



a) até n) indicam os grupos dos atributos; Ord*n* refere-se à representação da rede de drenagem (*n* representa a menor ordem considerada); r*n* indica que o cálculo do atributo foi feito a partir de janelas (*n* representa o tamanho da janela); SF indica sem aplicação de filtro; 'n'x'n' indica o tamanho da janela utilizada para filtrar o MDE antes de calcular o atributo; LpGn representa o filtro Gaussiano (n indica o tamanho da janela).

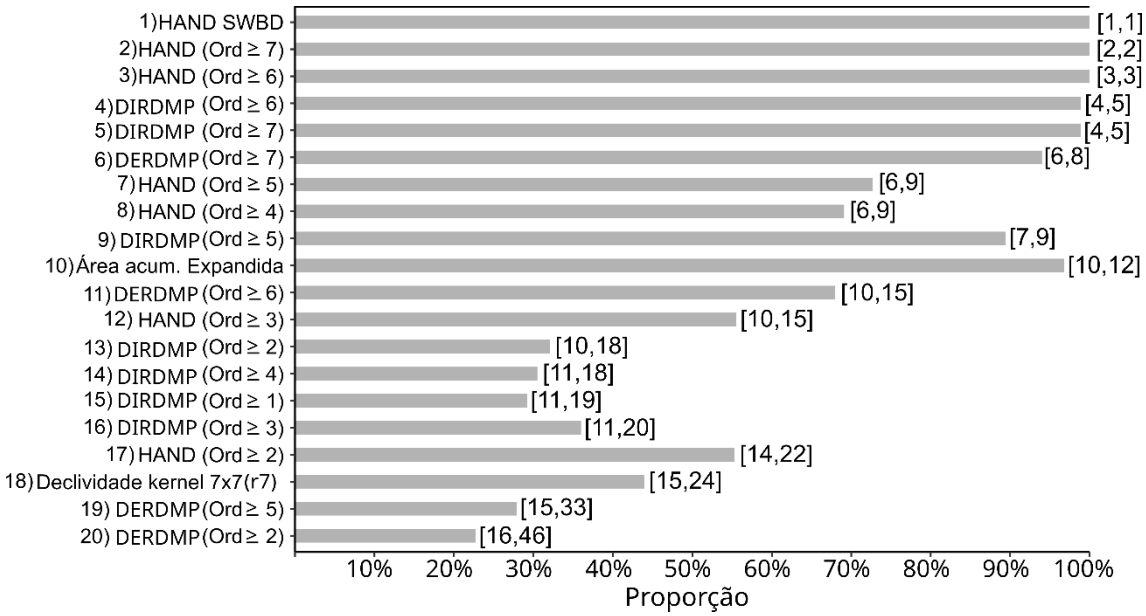
Fonte: Produção do autor.

Os atributos morfométricos (grupos *d, e, f, g, h, i, j, k* e *m* identificados na Figura 4.3) apresentaram uma maior variação quando comparados aos atributos hidrológicos (grupos *a, b, e* e *n* identificados na Figura 4.3), com exceção do atributo TWI (grupo *l* identificado na Figura 4.3), um atributo hidrológico que apresentou expressiva variação nas posições ocupadas. De maneira geral, os atributos que em média ocuparam posições associadas a uma maior importância, apresentaram uma baixa variação na posição ocupada no ranqueamento ao longo dos 1000 modelos. Em contrapartida, os atributos que em geral ocuparam posições de menor importância, apresentaram uma alta variação na posição ocupada. Por exemplo, analisando o grupo de declividade, tem-se que o atributo da declividade gerada a partir de uma janela 5x5 *pixels* (r5), apresentou uma amplitude de 96 unidades, ocupando desde a 27ª posição até a 124ª. Essa alta variação da posição ocupada demonstra a instabilidade do ranqueamento para os atributos que não foram considerados tão importantes, confirmando que a partir de uma determinada posição no ranqueamento, esses atributos passaram a competir entre si pelas posições, o que era esperado, de acordo com o que foi exposto anteriormente na Figura 4.2.

A instabilidade observada no ranqueamento, associada principalmente aos atributos que em média ocuparam posições menos relevantes, indica que ordenar a importância dos atributos a partir do MDA pode induzir à atribuição de uma falsa importância ao atributo. Dessa forma, um atributo pode ser considerado importante quando na verdade, pode não ter tanta relevância para a classificação. A instabilidade da medida MDA constatada neste estudo pode estar associada à alta correlação entre os atributos. Gregorutti, Michel e Saint-Pierre (2017), por meio de simulações, constaram que a medida MDA é sensível à correlação entre os atributos. Assim, a presença de atributos altamente correlacionados aumenta a instabilidade da métrica MDA. Como o ranqueamento gerado a partir do MDA foi usado para uma seleção de variáveis, como apresentado a seguir (Seção 4.3), a estratégia adotada aqui para lidar com a instabilidade da métrica MDA, e por consequência, do ranqueamento, foi gerar um ranqueamento baseado em múltiplas rodadas do RF, como feito em outros estudos (GENUER; POGGI; TULEAU-MALOT, 2010; FOX et al., 2017).

A Figura 4.4 apresenta o ranqueamento dos vinte atributos mais importantes, acompanhados da proporção de vezes que eles foram ranqueados nessa posição ao longo das 1000 rodadas e a posição máxima e mínima que eles ocuparam. Entre os vinte atributos mais importantes, dezenove são atributos hidrológicos. O único atributo morfológico presente foi a declividade calculada a partir de uma janela 7x7 (r7) e ocupou a 18ª posição. Os três atributos mais importantes, HAND SWBD, HAND (Ord ≥ 7) e HAND (Ord ≥ 6) não apresentaram variação das suas posições ao longo dos 1000 modelos obtidos, aparecendo sempre na 1ª, 2ª e 3ª posição, respectivamente. Ao analisar a diferença entre a posição máxima e mínima ocupada pelos atributos, nota-se que até a 10ª posição, a maior variação foi de apenas três unidades e a partir da 19ª posição, essa variação aumentou expressivamente (dezoito unidades). Assim, nota-se uma certa estabilidade do ranqueamento para os dez primeiros atributos. Esse resultado diverge do que foi observado por Millard e Richardson (2015) que utilizaram diferentes atributos derivados do LiDAR para mapeamento de áreas úmidas. Os autores constataram que os atributos selecionados como mais importantes (5 primeiros) variaram no decorrer de 100 iterações do RF, mesmo utilizando as mesmas amostras de treinamento.

Figura 4.4 - Ranqueamento a partir da métrica MDA para os vinte atributos considerados mais importantes e a proporção de vezes que cada um ocupou essa posição.



Os atributos estão dispostos de cima para baixo, do mais importante ao menos importante (em média). Ao lado de cada barra está indicada, entre colchetes, a posição mínima e máxima ocupada pelo atributo.

Fonte: Produção do autor.

4.3 Seleção dos atributos

4.3.1 Pré-seleção dos atributos morfométricos

Ao se analisar a distribuição das posições ocupadas pelos atributos de um mesmo grupo resultantes dos 1000 modelos obtidos (Figura 4.3) é possível notar que dentre todos os atributos pertencentes ao mesmo grupo, um deles foi considerado mais importante que os demais, ou seja, entre todas as formas de representação de um mesmo grupo de atributos, uma se destacou, ocupando posições de maior importância quando comparada com as outras. Entre todos os grupos analisados, em nenhum deles a forma original de representação do atributo foi considerada a mais importante, demonstrando-se a importância de se considerar a informação contextual espacial. De toda forma, apenas o aumento do tamanho da janela na obtenção do atributo não garante uma boa

representação, pois a janela pode ser muito ampla incluindo informações diferentes, o que gera uma representação incorreta do atributo. Por isso, é importante testar diferentes tamanhos de janelas.

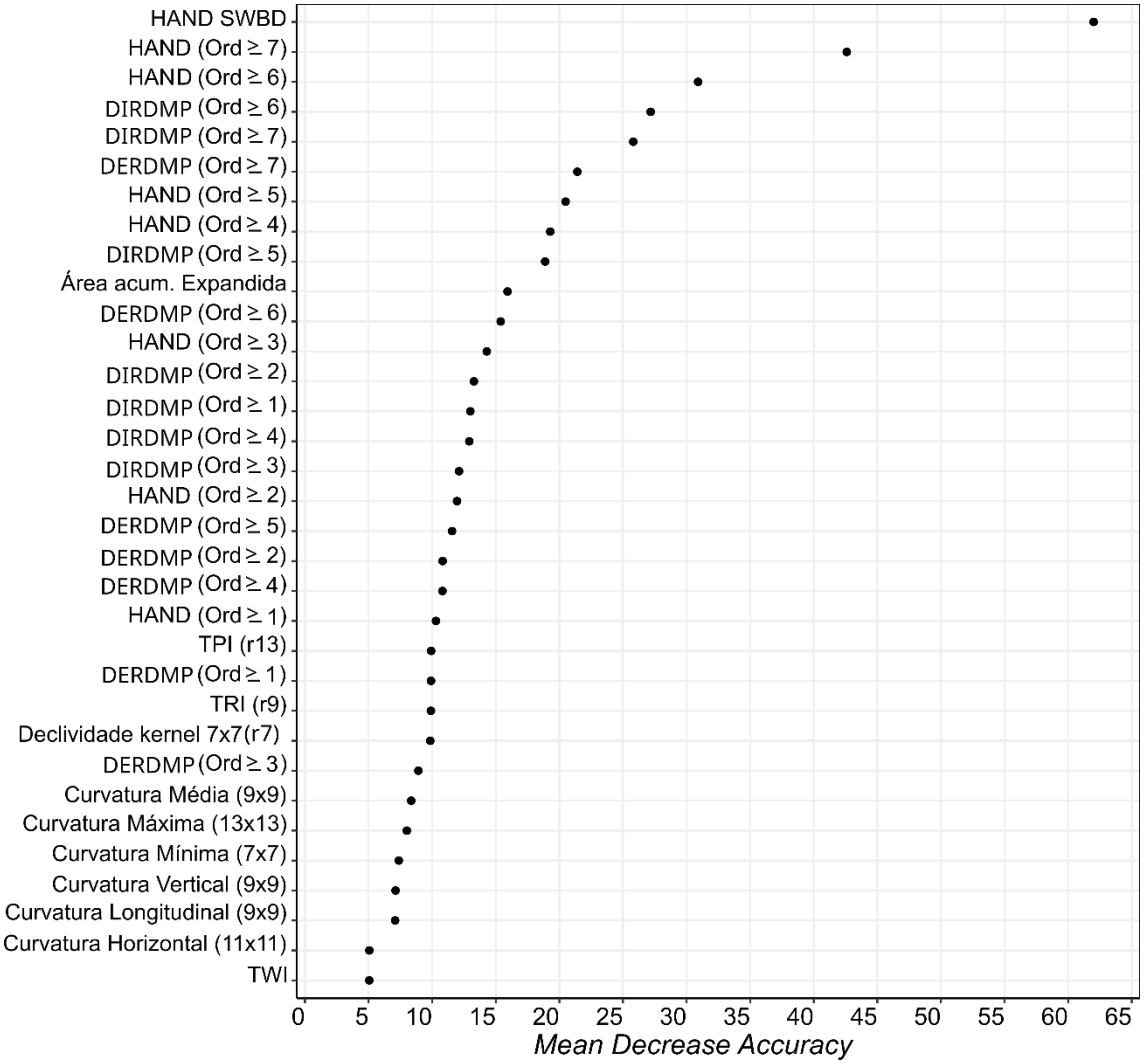
A fim de determinar um conjunto mínimo de atributos, foi implementado um método de seleção dos atributos. Como uma etapa anterior, foi realizada uma pré-seleção entre os atributos morfométricos, pois era de interesse desta pesquisa, escolher apenas uma, dentre as diferentes representações de um mesmo grupo de atributos morfométricos. Nesse sentindo, a análise da estabilidade do ranqueamento (Figura 4.3) foi utilizada para realizar essa pré-seleção sobre os atributos morfométricos. No caso, foram pré-selecionados os atributos que apresentaram maior estabilidade das posições ocupadas e que estavam associados a uma maior importância. Foram pré-selecionados os atributos: Declividade r7, TRI r9, TPI r13 (calculados utilizando uma janela de tamanho 7x7, 9x9, 13x13, respectivamente), Curvatura Horizontal 11x11, Curvatura Vertical 9x9, Curvatura Máxima 13x13, Curvatura Mínima 7x7, Curvatura Média 9x9, Curvatura Longitudinal 9x9 (calculados utilizando a partir do MDE filtrado com janelas de tamanhos 11x11, 9x9, 13x13, 7x7, 9x9, respectivamente). Após essa pré-seleção, restaram 33 atributos morfométricos e hidrológicos, que juntos passaram por um processo de seleção, como explicado a seguir.

4.3.2 Seleção dos atributos baseada na técnica *stepwise backward*

Considerando um modelo RF com 33 atributos, os parâmetros *ntree* e *mtry* foram redefinidos em 1000 e 5, respectivamente. Em seguida, foi obtido o ranqueamento dos atributos, proveniente de 1000 repetições do RF (Figura 4.5). Ao comparar esse ranqueamento com o obtido pelo modelo RF com todas as 124 variáveis (exibido anteriormente na Figura 4.4), tem-se que até a 13ª posição observam-se os mesmos atributos, na mesma ordem, e as posições seguintes foram ocupadas por atributos que variaram entre os dois ranqueamentos. O ranqueamento dos 33 atributos foi utilizado para indicar a ordem que os atributos

seriam retirados a cada passo do método de seleção baseado na técnica *stepwise backward*, como descrito na Seção 3.3.3.2.

Figura 4.5 - Ranqueamento dos 33 atributos baseado na métrica de importância *Mean Decrease Accuracy* (MDA).

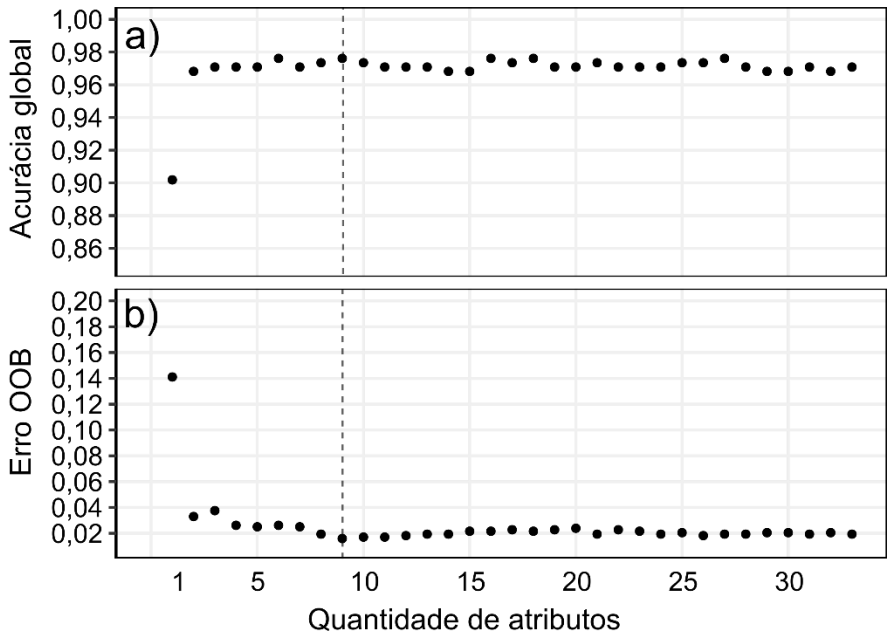


Fonte: Produção do autor.

Os primeiros atributos a serem removidos foram os que apresentaram menor importância, sendo eles: o TWI, seguido dos atributos relacionados às curvaturas. Os últimos a serem eliminados foram o HAND e o DIRDMP, considerando diferentes ordens da rede de drenagem. A cada atributo retirado, foi gerado um modelo RF e para cada modelo, foi computado o valor da acurácia

global e do erro OOB, como apresentado na Figura 4.6. É importante destacar que a acurácia foi calculada com base nas amostras de validação, ou seja, amostras independentes das utilizadas para o treinamento dos modelos.

Figura 4.6 - Acurácia global e erro OOB *versus* quantidade de atributos a cada passo do método de seleção de variáveis.



A linha tracejada indica o modelo selecionado, contendo nove atributos.

Fonte: Produção do autor.

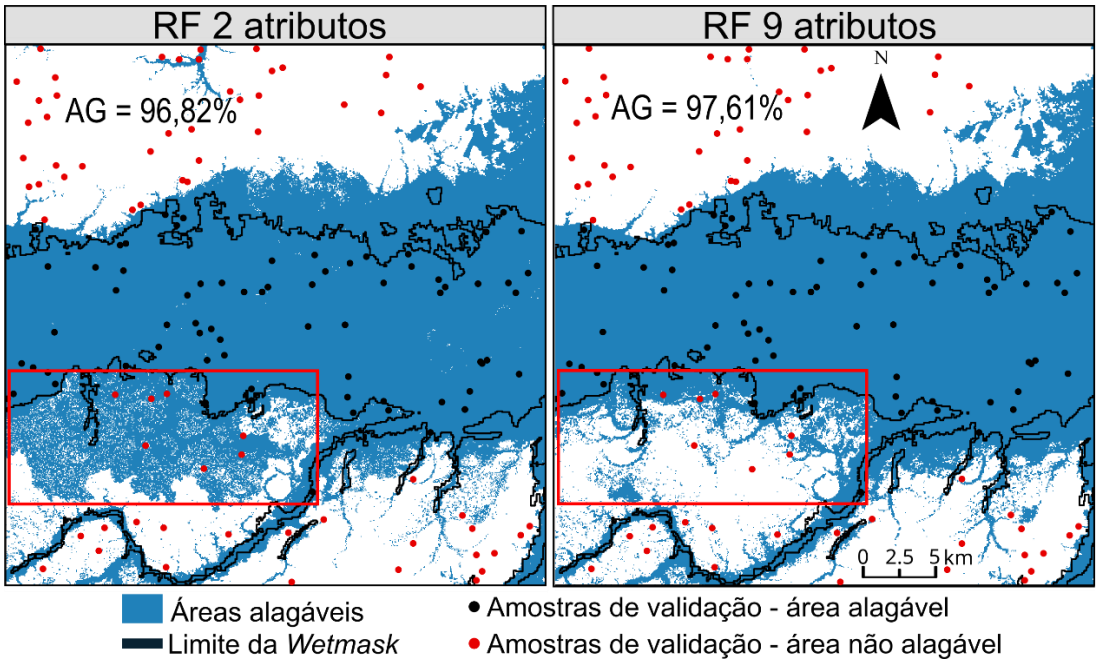
A acurácia global dos modelos (Figura 4.6a) variou entre 96 e 98% para todos os modelos com dois ou mais atributos, e caiu para 90% no modelo com apenas um atributo (HAND SWBD). Por outro lado, o erro OOB (Figura 4.6b) se manteve constante em aproximadamente 2% para o modelo considerando 33 atributos até o modelo com 9 atributos, aumentando para 14% no modelo com apenas um atributo. O modelo selecionado (linha tracejada na Figura 4.6) foi aquele que apresentou a maior acurácia. Como houve empate nesse critério, foi selecionado, entre os modelos com a maior acurácia, aquele com o menor erro OOB. Os nove atributos selecionados foram: HAND SWBD, HAND (Ord ≥ 7), DERDMP (Ord ≥ 7), HAND (Ord ≥ 6), DIRDMP (Ord ≥ 7), DIRDMP (Ord ≥ 6), DIRDMP (Ord ≥ 5), HAND (Ord ≥ 4) e HAND (Ord ≥ 5) (em ordem de

importância). Os atributos selecionados referentes à cada área podem ser visualizados no Apêndice A.

No gráfico da acurácia global (Figura 4.6a), notam-se altos valores de acurácia, mesmo variando a quantidade de atributos. Isso levanta algumas questões a respeito dessa métrica de avaliação da classificação. Primeiramente, tem-se que as amostras utilizadas para validação do modelo exercem influência na estimativa da acurácia. É importante notar que o tamanho do conjunto amostral e a sua distribuição espacial são fatores que devem ser considerados para se avaliar a acurácia estimada de uma classificação (FOODY, 2002). Neste estudo, foi utilizada a estratégia de apenas amostrar pontos em concordância com a *wetmask* e com maior certeza da sua classe. Ao se pensar no efeito que isso tem sobre a acurácia, tem-se que: se os pontos amostrados foram aqueles em que se tinha uma maior certeza de sua classe, e essa certeza foi proveniente de uma análise visual do MDE, é esperado que o modelo acerte mais nesses locais, pois deve haver padrões mais facilmente identificáveis pelo classificador. Assim, as altas acurácias encontradas podem estar relacionadas com a escolha das amostras.

É importante notar que uma alta acurácia estimada através de pontos amostrais nem sempre representa uma boa classificação quando se analisa espacialmente essa classificação. Para exemplificar, a Figura 4.7 apresenta o resultado da classificação para uma mesma área (Área 3), considerando um modelo com dois atributos e o modelo selecionado, com 9 atributos. O modelo com apenas dois atributos, apresentou uma acurácia global de 96,82%, enquanto a do modelo selecionado foi de 97,61%. Mesmo o valor da acurácia dos modelos sendo próximos, ao comparar visualmente, percebem-se diferenças consideráveis. A classificação com nove atributos parece ser menos ruidosa do que a classificação com 2 atributos, que classificou como área alagável, com bastante ruído, uma grande porção de área na região sudoeste da imagem (retângulo vermelho na Figura 4.7). A qualidade da classificação ao se comparar com a *wetmask* (contorno preto na Figura 4.7) não será analisada neste momento, pois será abordado a seguir, na Seção 4.5. Aqui, o intuito é apenas levantar a discussão a respeito da métrica acurácia e fatores que a influenciam.

Figura 4.7 - Comparação entre as classificações e a acurácia global (AG) dos modelos *Random Forest* com dois e nove atributos.



O retângulo vermelho indica uma área que apresentou diferença considerável entre as duas classificações.

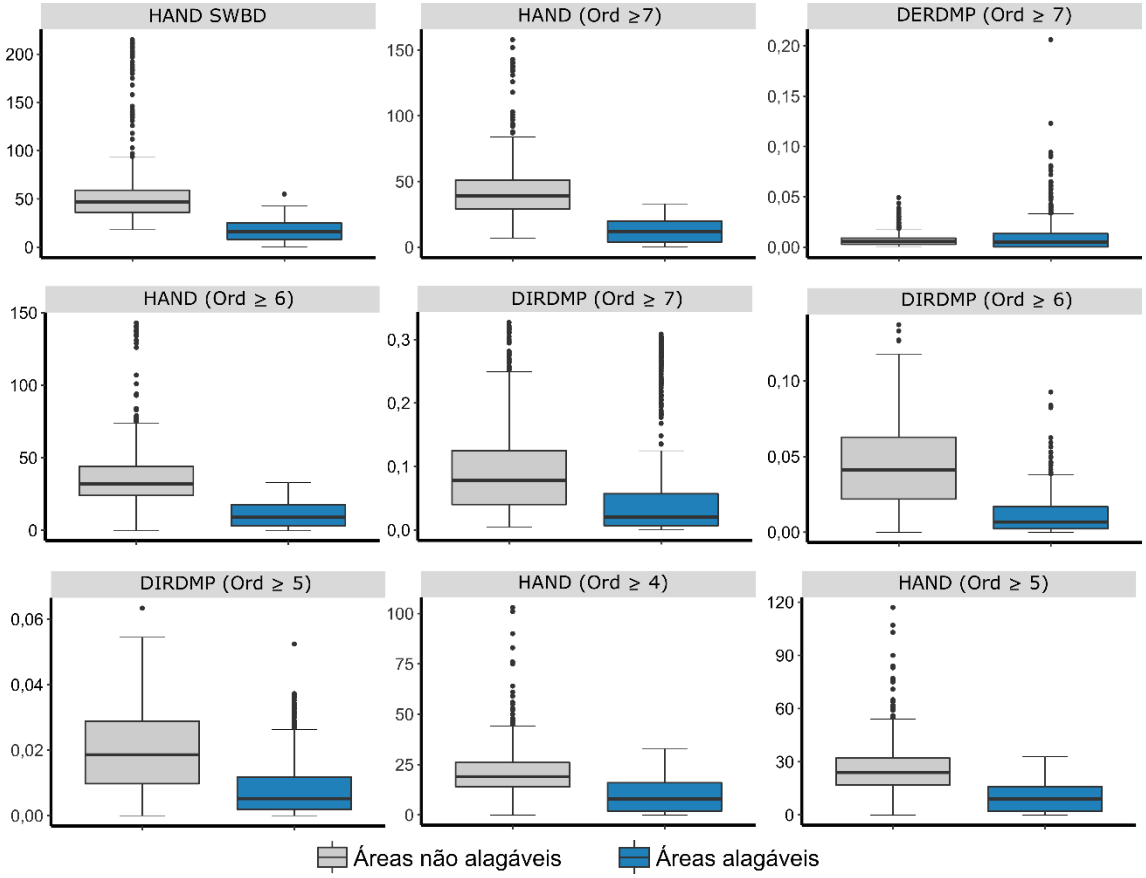
Fonte: Produção do autor.

Os resultados apresentados na Figura 4.7 evidenciam a importância de, ao se comparar diferentes modelos, também avaliar visualmente a classificação, e não se basear unicamente no resultado da métrica acurácia global. Neste caso, tem-se que, utilizando o método de seleção de variáveis, foi possível selecionar um modelo de acordo com os critérios adotados, que se mostrou visualmente coerente além de apresentar a maior acurácia e menor erro OOB.

A Figura 4.8 apresenta os *boxplots* dos nove atributos selecionados, em que pode se ter um entendimento da distribuição dos valores dos atributos para todas as amostras utilizadas (treinamento e validação). Como só foram amostrados pontos nas Áreas 1, 2, 3 e 4, esta análise não inclui a Área 5. Todos os atributos selecionados, com exceção do atributo DERDMP (Ord ≥ 7), apresentaram um alto poder discriminatório entre as classes áreas alagáveis e áreas não

alagáveis. Isso é evidenciado pelo fato de os *boxplots* estarem bem separados, sem sobreposição.

Figura 4.8 - *Boxplot* dos atributos selecionados para as classes de áreas alagáveis e áreas não alagáveis.



Fonte: Produção do autor.

Os nove atributos selecionados foram da categoria hidrológicos. Isso indica que, as informações que estão relacionadas com a rede de drenagem foram mais relevantes para mapear as áreas alagáveis, demonstrando a importância de se ter a informação do traçado da rede de drenagem para separar os ambientes de áreas alagáveis e não alagáveis. Além disso, ao analisar as ordens da drenagem consideradas no cálculo dos atributos selecionados, tem-se que a menor ordem foi de 4. Note que, quanto maior a ordem, menor é a densidade da drenagem considerada. Dessa forma, aqueles atributos que representaram a informação

considerando ordens maiores da drenagem foram mais relevantes para a classificação e foram selecionados no modelo final.

O atributo DIRDMP possui uma relação evidente com a identificação de áreas alagáveis. É esperado que áreas mais próximas da drenagem tenham um maior potencial em alagar. Entretanto, é importante notar que isso depende da ordem da rede de drenagem considerada, e isso foi percebido de acordo com os atributos selecionados. De maneira geral, considerando um alagamento que aconteça devido ao aumento do nível do rio, uma região que está a uma pequena distância do curso d'água, mas esse curso d'água se trata de um rio de pequena ordem, pode possuir menores chances de ser alagável devido seu pequeno fluxo, sem volume suficiente para alagar regiões no seu entorno. Uma exceção a essa situação ocorre quando uma drenagem de pequena ordem deságua em uma de grande ordem. Nesse caso, há um represamento do rio, ocasionando o efeito de remanso, quando o rio maior é responsável por um aumento do nível d'água no rio menor. Foram selecionados os atributos de distância à drenagem com ordem maior que cinco, seis e sete, indicando que nesse contexto, o conhecimento da proximidade com os rios de maior ordem agregou mais informação relevante para identificar as áreas alagáveis.

A partir da Figura 4.8, nota-se que, quanto mais densa a drenagem considerada (menores ordens), menores são os valores do HAND. O HAND é um atributo que normaliza a altura em função da rede drenagem e a partir disso, destaca variações locais na altura, podendo assim, evidenciar determinados ambientes (RENNÓ et al., 2008). Os HANDs mais importantes para a classificação, e por consequência, os selecionados no processo de seleção dos atributos, foram aqueles baseados em ordens mais altas da drenagem. Isso indica que, o alagamento aqui representado, ocorre no entorno dos grandes rios, provavelmente por um efeito de transbordamento.

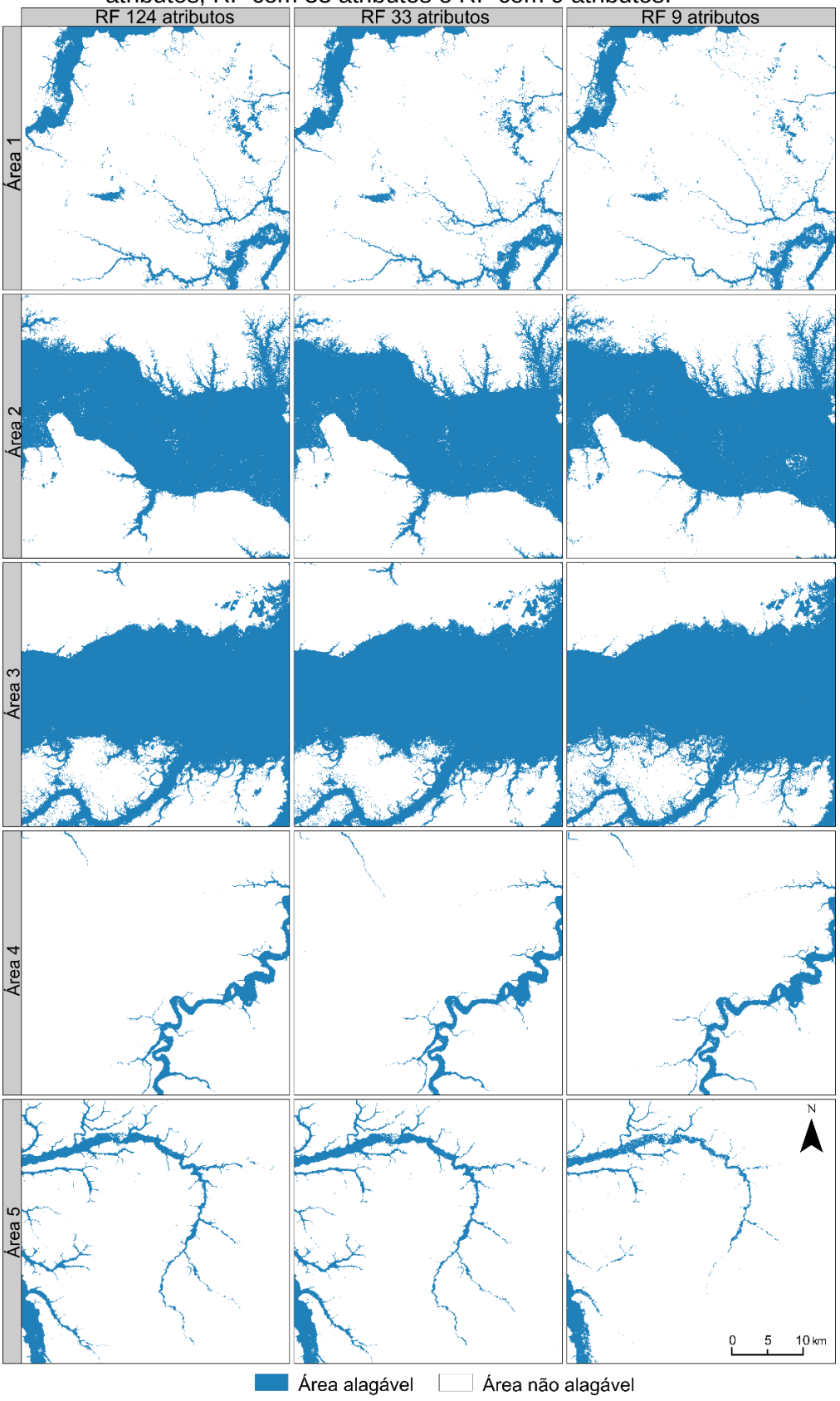
O DERDMP é um atributo obtido por meio da relação do HAND com o DIRDMP. Rennó et al. (2008) explicam que regiões distantes da drenagem e com baixos valores de HAND estão associados às áreas mais planas conectadas à drenagem, com potencial de serem alagáveis. A partir do *boxplot* desse atributo

(Figura 4.8), nota-se que sendo analisado de forma individual, o atributo não apresenta uma alta capacidade de separar os ambientes. Entretanto, é importante notar que o poder de predição do modelo selecionado é resultado das informações oriundas dos atributos individualmente e das associações entre eles. As informações complementares entre os atributos levam o classificador a aprender determinados padrões, possibilitando assim, mapear os ambientes de áreas alagáveis. Dessa forma, o atributo DERDMP deve ser importante para a classificação de algumas regiões específicas. É válido ter em mente esse tipo de situação, para evitar que atributos importantes sejam descartados. Em uma análise preliminar dos atributos, com o objetivo de se eliminar atributos com baixo poder de separabilidade, pode-se levar a descartar atributos que venham a ser importantes para a classificação.

4.4 Comparação entre os modelos

Foram comparados os três modelos RF, os quais, estavam com seus parâmetros otimizados (todos utilizando a mesma quantidade de árvores). A Figura 4.9 apresenta as classificações resultantes utilizando os modelos RF com 124 atributos, RF com 33 atributos e modelo selecionado com 9 atributos. De maneira geral, a partir da análise visual tem-se que as classificações obtidas considerando os diferentes modelos foram similares entre si para as cinco áreas de estudo.

Figura 4.9 - Comparação das classificações baseadas nos modelos RF com 124 atributos, RF com 33 atributos e RF com 9 atributos.



Fonte: Produção do autor.

A partir da Figura 4.9, tem-se que a redução de 124 para 33 atributos não modificou visualmente o mapeamento. Nota-se que as pequenas diferenças existentes entre as classificações foram observadas ao se comparar o modelo selecionado (com 9 atributos) com os demais. Na maior parte dos casos, as diferenças foram percebidas nas regiões do contorno da área mapeada como alagável. Existiram áreas que antes foram classificadas como alagáveis e ao reduzir a quantidade de atributos passaram a ser não alagáveis, assim como ocorreu a situação inversa, áreas que eram classificadas como não alagáveis e passaram a ser alagáveis. Por exemplo, na Área 2, percebe-se um sutil aumento das áreas mapeadas como alagáveis no contorno da área na região superior, enquanto na parte inferior, houve uma redução da área mapeada em um braço do rio. Já na Área 3, na parte superior da área, uma feição deixou de ser considerada como área alagável ao se reduzir a quantidade de atributos. Ainda na Área 3, na região sudoeste, ao se reduzir os atributos, houve um aumento da área classificada como alagável, o que acentuou o efeito sal e pimenta, também constatado nas classificações obtidas a partir dos outros modelos. A Área 4 foi onde ocorreram as menores diferenças ao se comparar os modelos. Já a Área 5, onde não foram amostrados pontos para o treinamento dos modelos, apresentou algumas mudanças. Áreas que nos modelos com mais atributos foram consideradas como alagáveis, passaram a ser não alagáveis no modelo selecionado com 9 atributos. Esse comportamento foi percebido tanto nos rios mais estreitos, quanto nos rios maiores. Em alguns pontos da Área 5, localizados sobre os rios principais, notou-se descontinuidades da classificação obtida a partir do modelo reduzido, o que caracteriza o ruído do tipo sal e pimenta.

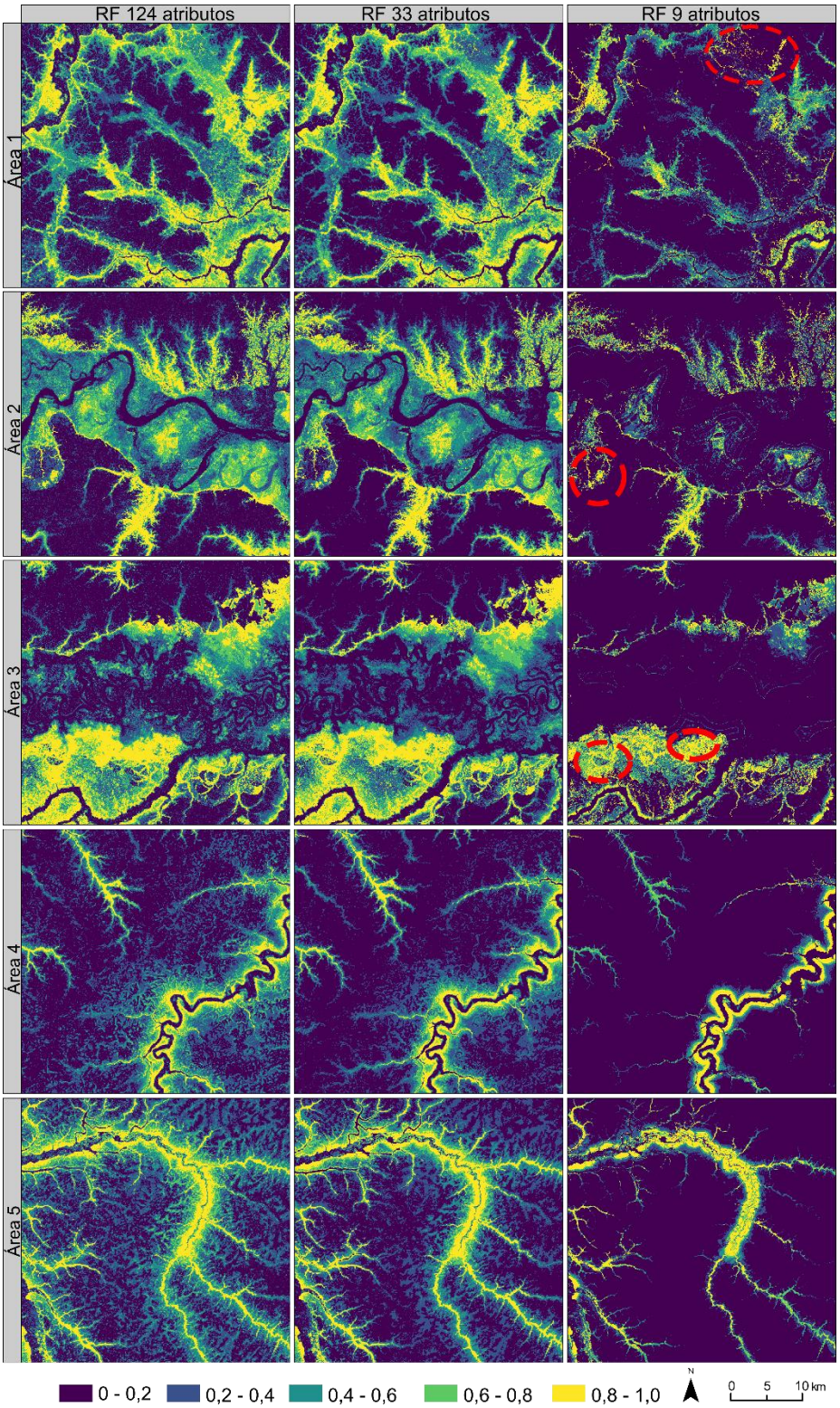
Quando comparados os mapas de incerteza, obtidos a partir da métrica Entropia de Shannon (Figura 4.10), observa-se que em todas as áreas, o modelo selecionado (com 9 atributos) apresentou uma redução expressiva da incerteza associada ao mapeamento, levando a uma maior diferenciação entre os ambientes de áreas alagáveis e não alagáveis. Isso foi indicado pela predominância da cor azul no tom mais escuro no mapa, que está associado a uma baixa entropia (de 0 a 0,2). Assim, tem-se que a redução das variáveis resultou em uma classificação com uma maior certeza associada.

A partir da Figura 4.10, nota-se que de maneira geral, as maiores incertezas estão nas áreas de transição entre as classes. Isso era esperado, visto que a transição entre os ambientes apresenta maiores confusões, o que pode dificultar para o classificador encontrar determinados padrões nessas regiões. Além disso, analisando os mapas de incerteza em conjunto com a classificação exposta na Figura 4.9, tem-se que elevadas incertezas foram encontradas em regiões que foram classificadas como áreas alagáveis com um aspecto ruidoso. Os círculos na cor vermelha, destacados na Figura 4.10 nas Áreas 1, 2 e 3 exemplificam regiões que foram classificadas como alagáveis, mas com muita confusão associada. Nesse sentido, a informação da entropia pode auxiliar em um processo pós classificação, em que essas áreas com maior entropia podem ser desconsideradas, permanecendo apenas as áreas em que o classificador apresentou uma maior certeza.

Ainda em relação à Figura 4.10, analisando o mapa de incerteza do modelo reduzido para a Área 4, é possível observar as maiores incertezas em regiões de transição entre as classes. Nota-se que a calha do rio principal foi classificada com baixa entropia, ou seja, com elevada certeza associada, e ao se afastar em direção às bordas, a incerteza assume seu valor máximo, reduz para um valor intermediário, até atingir seu valor mínimo novamente, pois foi identificado facilmente o ambiente de áreas não alagáveis. Da mesma forma, na Área 5 percebe-se que as regiões de baixa incerteza foram associadas às áreas mapeadas como não alagáveis e às áreas dentro da calha dos rios maiores. Já as altas incertezas foram visualizadas nos rios mais estreitos, provavelmente mais encaixados no relevo.

É importante destacar que considerar a informação da distribuição espacial da incerteza de classificação ao mapeamento é algo de grande relevância. Um mapa pode apresentar um alto valor de acurácia global, mas se os erros presentes estiverem associados com uma determinada área de interesse, esse mapeamento pode perder sua importância. Dessa forma, a apresentação da classificação em conjunto com o mapa de incerteza, permite que se conheça os locais em que aquele mapeamento foi gerado com maior ou menor confiança.

Figura 4.10 - Comparação entre os mapas de incerteza obtidos a partir da Entropia de Shannon para os modelos RF com 124, 33 e 9 atributos.

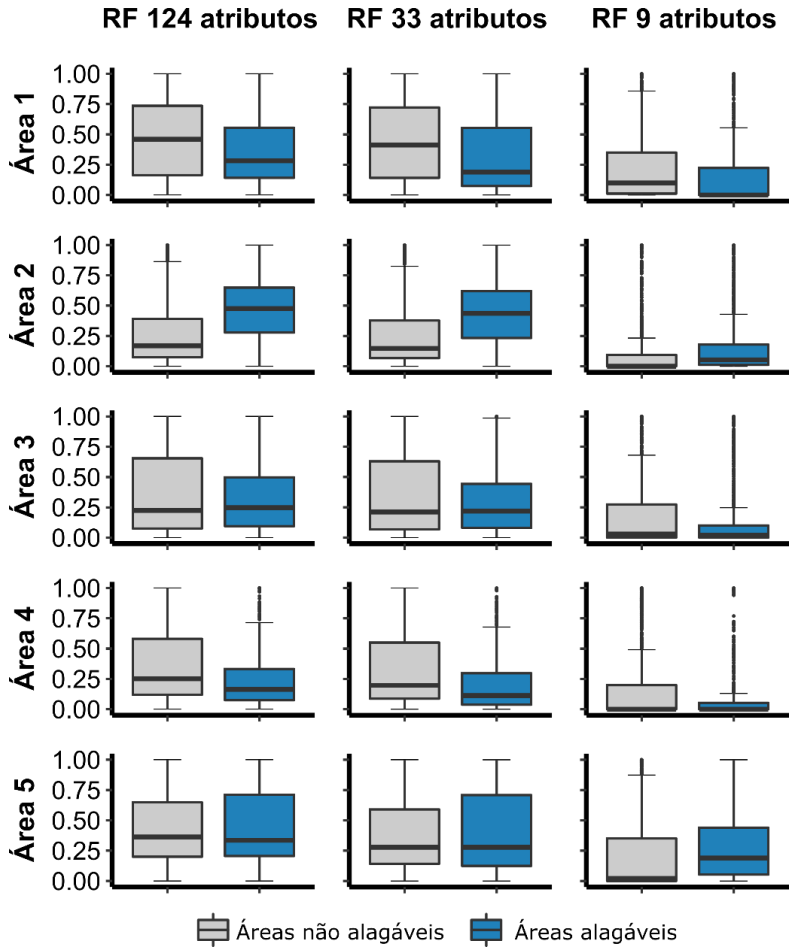


Os círculos vermelhos nas Áreas 1, 2 e 3 exemplificam regiões que foram classificadas como alagáveis, mas com elevada incerteza.

Fonte: Produção do autor.

A Figura 4.11 apresenta os *boxplots* para os valores de entropia separados por classes (áreas alagáveis e áreas não alagáveis), para cada um dos modelos analisados. Em geral, o modelo selecionado (com 9 atributos) reduziu os valores medianos da entropia. Na Área 2, a mediana da entropia para a classe de áreas alagáveis apresentou uma diferença de 0,43 quando comparado os modelos com 124 atributos e com 9 atributos (RF124 = 0,48 e RF9 = 0,05). Para as outras áreas, a diferença entre as medianas para os dois modelos foi menor que 0,30. Note que, para as Áreas 1, 2, 3 e 4, a mediana do valor da entropia no modelo selecionado tendeu a 0, enquanto que na Área 5 foi de 0,19.

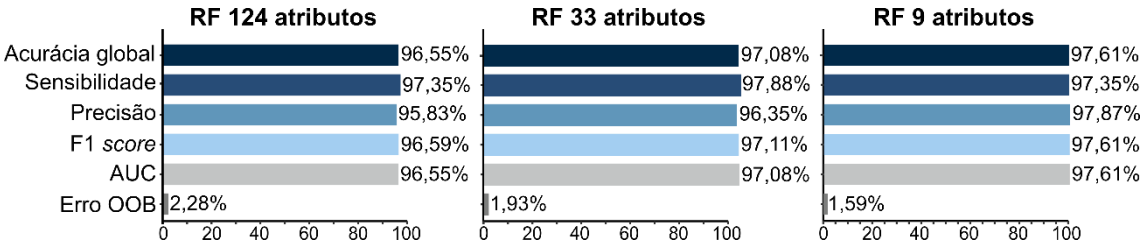
Figura 4.11 - Comparação entre os *boxplots* da entropia para os modelos RF com 124, 33 e 9 atributos. Os *boxplots* foram baseados em 2000 amostras distribuídas aleatoriamente na imagem, por área.



Fonte: Produção do autor.

As métricas de avaliação da classificação foram calculadas para cada um dos três modelos analisados, como mostra a Figura 4.12. Nota-se que a diferença entre os valores de todas as métricas é pequena comparando os três modelos. A acurácia global foi de 96,55% para o modelo com 124 atributos, passando para 97,08% no modelo com 33 atributos, chegando a 97,61% no modelo selecionado contendo 9 atributos. O teste de McNemar foi aplicado para determinar se os valores de acurácia foram estatisticamente diferentes. Constatou-se que para o RF com 124 atributos *versus* RF com 33 atributos o valor-p foi de 0,91. Já para a comparação entre o RF com 124 atributos *versus* RF com 9 atributos o valor-p foi 0,80. Por fim, o RF com 33 atributos *versus* RF com 9 atributos levou a um valor-p de 0,91. Assim, nas três comparações, os resultados indicaram que não houve diferença entre as acurácias, considerando um nível de significância de 5%. Dessa maneira, não se pode concluir que houve melhorias com relação a acurácia ao se reduzir a quantidade de atributos, comprovando que o modelo com 9 atributos consegue resultados semelhantes àqueles obtidos pelo modelo completo.

Figura 4.12 - Comparação das métricas de avaliação da classificação entre os modelos RF com 124, 33 e 9 atributos.



Fonte: Produção do autor.

Diante dos resultados aqui expostos, tem-se que a redução dos atributos levou, em geral, a uma classificação visualmente semelhante com a obtida considerando todos os atributos. Além disso, a acurácia da classificação permaneceu praticamente a mesma. Entretanto, a entropia associada à classificação reduziu consideravelmente, melhorando a separação entre as classes, o que que indica que ao reduzir a quantidade de atributos, o

classificador convergiu para o resultado com uma maior certeza associada. Dessa forma, este estudo destaca a possibilidade de se determinar um conjunto de atributos suficientemente reduzido capaz de gerar a classificação de forma satisfatória, semelhante àquela obtida ao se utilizar o conjunto completo de atributos.

Os resultados encontrados corroboram com outros estudos que realizaram a seleção dos atributos para a classificação. Por exemplo, Millard e Richardson (2013) propuseram mais de 110 atributos derivados de dados LiDAR e SAR para o mapeamento de áreas alagáveis. Após realizar a seleção de atributos, os autores verificaram que diversos atributos poderiam ser removidos sem que a acurácia da classificação fosse afetada de forma significativa, obtendo assim a classificação final proveniente de um modelo contendo apenas 8 atributos. Já Berhane et al. (2018a) testaram 37 atributos ópticos e morfométricos para a classificação de áreas alagáveis, e encontraram a maior acurácia de mapeamento utilizando um modelo com apenas três atributos.

O fato de a acurácia global não ter variado de forma estatisticamente significativa ao variar a quantidade de atributos presentes no modelo confirma o argumento de que o classificador RF consegue lidar de forma satisfatória com uma grande quantidade de atributos, incluindo dados que não sejam relevantes para a classificação (BELGIU; DRAGUT, 2016). Entretanto, deve-se notar que a utilização de uma grande quantidade de atributos está associada a um grande volume de dados, o que aumenta o custo computacional envolvido na extração e processamento desses dados. Neste estudo, inicialmente foram definidos 124 atributos para o mapeamento das áreas alagáveis em áreas testes da bacia Amazônica. Ao se pensar em expandir o mapeamento para toda a bacia, os 124 atributos deveriam ser calculados para toda a região, o que está associado a um alto custo computacional. Além disso, reduzir os atributos a um pequeno grupo pode auxiliar no entendimento dos resultados determinando-se quais atributos são realmente relevantes para a classificação, o que simplifica a interpretação de como esses atributos contribuem para a modelagem do fenômeno estudado.

4.5 Comparação entre a classificação *Random Forest* e a *wetmask*

Com o objetivo de melhorar o mapeamento final de cada uma das áreas escolhidas para este estudo, a classificação obtida a partir do modelo selecionado (RF com 9 atributos) passou por um processo de pós-classificação. Inicialmente, foi aplicado um limiar na informação da entropia, de forma que apenas os *pixels* que foram classificados com uma entropia menor que 0,80 foram considerados como áreas alagáveis na classificação final. Esse valor de entropia reduz a área alagável mapeada, mantendo aquelas em que pelo menos 75% das árvores do RF indicaram ser dessa classe. Tal procedimento teve por objetivo apresentar uma classificação com um maior nível de certeza, visto que foram consideradas como áreas alagáveis apenas aquelas áreas em que o classificador teve uma maior certeza em classificar como áreas alagáveis. Em seguida foi aplicado um filtro de moda (janela 5x5) para reduzir os ruídos do tipo sal e pimenta na classificação. A Figura 4.13 apresenta a comparação entre a classificação final e a *wetmask* (Hess et al., 2015), destacando áreas de concordância e discordância entre os mapas. De maneira geral, a classificação RF apresentou coerência quando comparada com a *wetmask*, demonstrada pela grande porção de área em concordância entre os mapas em todas as áreas de estudo.

Na Área 1, as maiores divergências entre os mapas ocorreram em feições menores e mais estreitas, que deixaram de ser consideradas áreas alagáveis na classificação RF. Essas feições, muitas vezes relacionadas a calhas de rios encaixados, podem ser mais difíceis de serem detectadas quando se usa atributos com grande abrangência espacial que tendem a mascarar essas feições. Além disso, o pós-processamento, através de técnicas de filtragem, pode eliminar feições lineares descontínuas previamente detectadas pelo RF. Outra feição importante que não foi detectada pelo RF localiza-se na parte nordeste desta área. Por se tratar de uma região muito plana, o MDE-NASADEM apresenta um nível de ruído considerável, o que pode ter dificultado a detecção de feições relacionadas às áreas alagáveis. Essas áreas precisam ser melhor investigadas para certificar de que se trata de áreas alagáveis de fato. Por outro

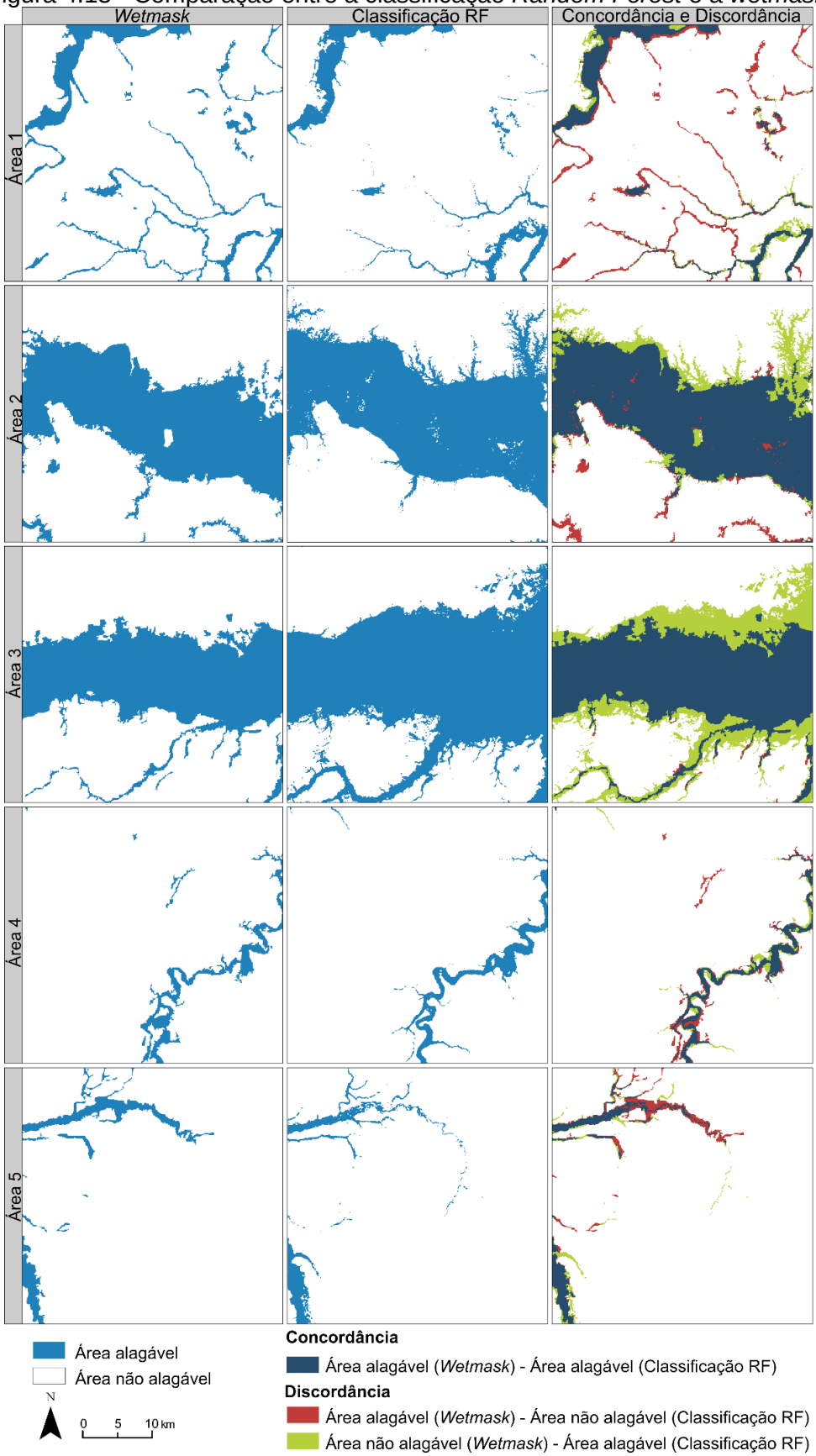
lado, melhorias na delimitação da área alagável podem ser verificadas e serão posteriormente apresentadas.

Na Área 2, houve um aumento de área considerada como alagável na parte superior da imagem, associada principalmente à inclusão de afluentes do rio principal na classificação RF. Por outro lado, na parte inferior da imagem, feições foram desconsideradas. Na região central da imagem, a classificação RF gerou algumas descontinuidades (“buracos”) ao longo da região classificada como áreas alagáveis. Ao analisar o MDE, a presença dessas descontinuidades não é justificada, o que indica que elas representam erros de classificação, ocasionados possivelmente pela falta de amostras que representem determinadas características ali presentes.

As maiores discordâncias entre os mapeamentos ocorreram na Área 3, resultantes de áreas que passaram a ser consideradas como alagáveis na classificação RF. A análise da classificação nessa área em conjunto com o MDE indicou que na parte superior da imagem, as feições mapeadas como áreas alagáveis apresentaram uma conformidade com o MDE. Já na parte inferior da imagem, em alguns pontos, a classificação RF aumentou a área mapeada como alagável de forma acentuada e não justificada quando analisado o MDE, indicando provavelmente uma superestimação dessa classe.

Na Área 4, as divergências foram pequenas, sendo a maioria delas relacionada a uma melhor representação do contorno das áreas alagáveis pela classificação RF, quando analisada em conjunto com o MDE. Isso pode ser atribuído a uma melhor representação das áreas alagáveis nessas regiões pelos atributos extraídos do MDE, levando a uma melhor separação entre os ambientes alagáveis e não alagáveis.

Figura 4.13 - Comparação entre a classificação *Random Forest* e a *wetmask*.



Fonte: Produção do autor.

Na Área 5, as divergências entre a classificação RF e a *wetmask* ocorreram principalmente na região norte da imagem, em que a classificação RF apresentou uma descontinuidade no canal principal. Inicialmente, parte desses canais foi considerada como áreas alagáveis, entretanto com uma alta entropia associada. Assim, como foi realizado o procedimento de pós-classificação que removeu áreas mapeadas com elevada incerteza, essas áreas foram desconsideradas. Entretanto, analisando o MDE, essa descontinuidade não é justificada, o que indica que há características nessa região que não foram bem representadas pelos atributos selecionados e pelas amostras utilizadas. É importante ressaltar que não foram coletados pontos na Área 5, o que indica que não foram coletados pontos suficientemente representativos dessas feições nas outras áreas.

A Tabela 4.1 apresenta, para cada área de estudo, a área classificada em cada classe na *wetmask* e na classificação RF. O total de área mapeada como alagável na classificação RF para a Área 2 e a Área 3 apresentou um aumento em relação a *wetmask* de aproximadamente 14% e 50%, respectivamente. Já nas Áreas 1 e 5, a classificação RF em comparação com a *wetmask*, reduziu a área total classificada como alagável em aproximadamente 21% e 15%, respectivamente. Na Área 4, a redução da área mapeada como alagável pela classificação RF foi de apenas 1,2%, sendo essa a área que apresentou as menores mudanças na quantidade de área mapeada como alagável.

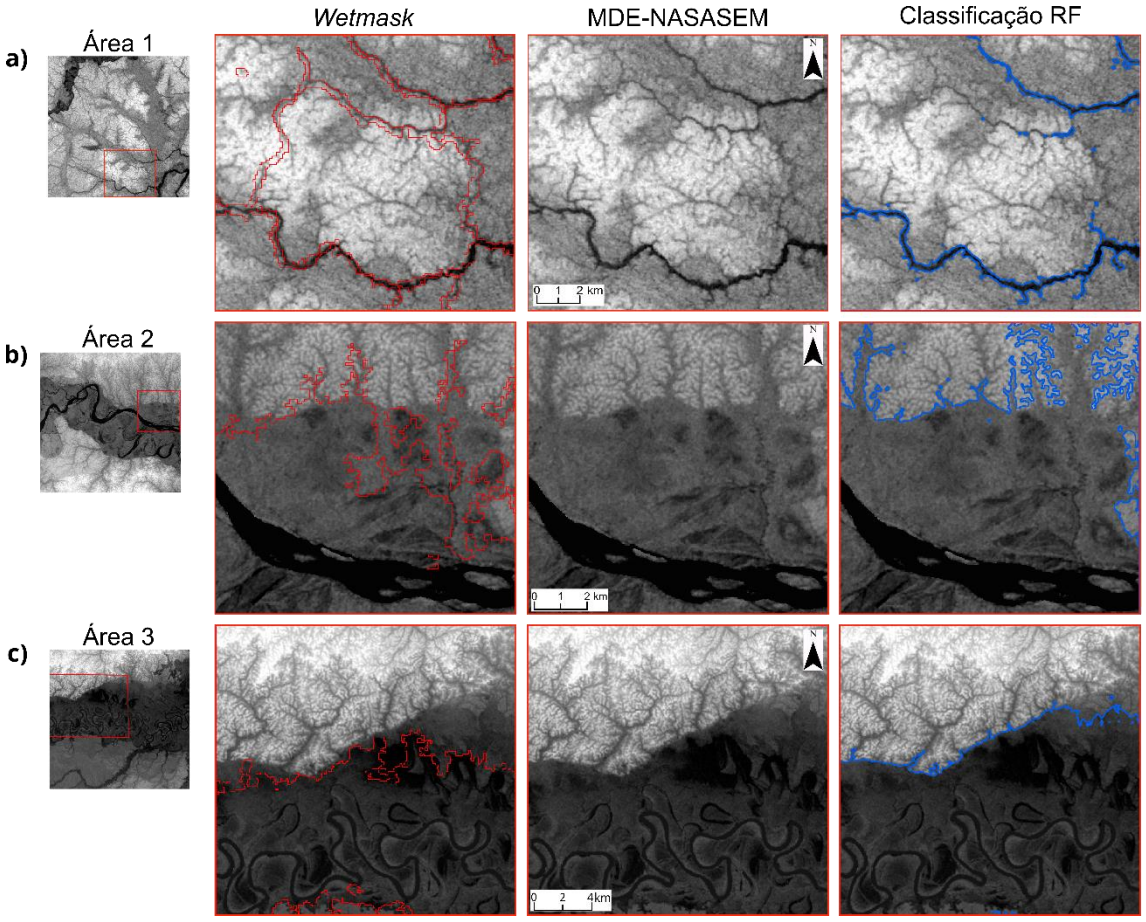
Tabela 4.1 - Quantificação da área mapeada nas classes de áreas alagáveis e áreas não alagáveis na *wetmask* e na classificação RF.

	Wetmask		**Classificação RF**	
Área	**Área alagável (km²)**	**Área não alagável (km²)**	**Área alagável (km²)**	**Área não alagável (km²)**
1	137,83	1228,32	109,30	1256,85
2	501,44	867,15	569.88	798,73
3	449,43	908,39	677,87	679,96
4	57,35	1305.25	56,69	1305,94
5	62,25	1296,56	52,80	1305,99

Fonte: Produção do autor.

Apesar das inconsistências entre os mapeamentos, em todas as áreas de estudo foram notadas melhorias em relação a *wetmask* no mapeamento das áreas alagáveis, quando analisada a classificação RF em conjunto com o MDE. A Figura 4.14 apresenta exemplos de regiões em que foi verificado que a classificação RF conseguiu refinar o mapeamento da *wetmask*. Na Figura 4.14a, é possível notar na Área 1, uma feição que deixou de ser considerada como área alagável na classificação RF, apresentando assim uma maior conformidade com o MDE, pois nele não está representada uma feição que induza a interpretação daquela área como área alagável. Na Figura 4.14b Figura 4.14c, estão exemplificadas regiões em que a classificação RF aumentou a área mapeada como alagável nas Áreas 2 e 3, respectivamente. Nesses dois exemplos, é possível notar uma maior conformidade entre a classificação RF e o MDE, que passou a representar uma maior quantidade de área como alagável. Regiões em que houve um aumento mais significativo da área mapeada como área alagável e aparentemente está de acordo com o MDE, podem estar associadas às áreas que alagam apenas devido a ocorrência de cheias mais acentuadas.

Figura 4.14 - Exemplos de áreas em que a classificação RF apresentou uma melhoria do mapeamento das áreas alagáveis em comparação com a wetmask quando analisado em conjunto com o MDE-NASADEM.



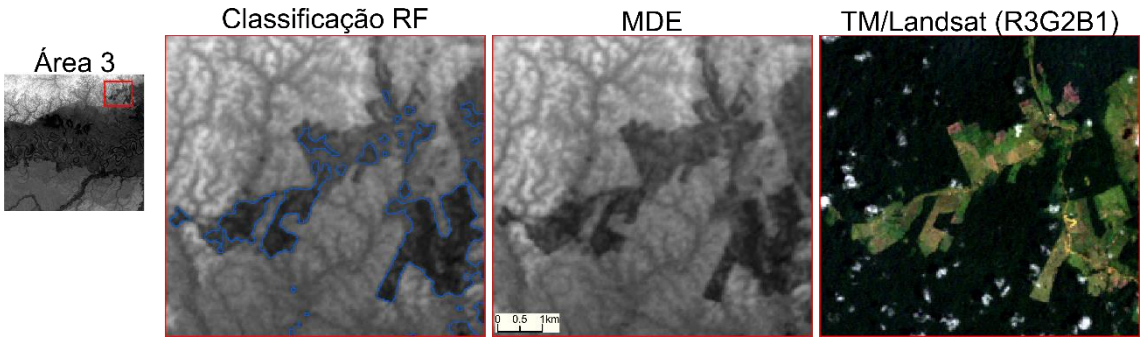
O limite da *wetmask* (cor vermelha) e o limite da classificação RF (cor azul) estão representados sobre o MDE. Tons mais escuros representam elevações mais baixas.

Fonte: Produção do autor.

É importante destacar que o MDE-NASADEM (reprocessamento do MDE-SRTM), utilizado como dado base neste estudo, apresenta limitações que influenciam a classificação obtida. Esse MDE é sensível à vegetação, apresentando o chamado efeito dossel, que é explicado pela interação da banda C do radar de abertura sintética com a copa das árvores. Como esse efeito afeta o MDE, consequentemente os atributos extraídos a partir dele também são influenciados (VALERIANO et al., 2006). Devido a esse efeito do dossel, podem surgir feições no MDE que não retratam fielmente a realidade, especialmente em

áreas de desmatamento. Nessas áreas de transição entre a floresta e a área desmatada é registrada uma falsa mudança brusca de elevação, que tem influência em praticamente todos os atributos testados. Uma situação como essa foi observada na Área 3. A Figura 4.15 apresenta a classificação RF para a uma região da Área 3 que foi classificada como área alagável pela classificação RF, entretanto nota-se que a região se trata de uma área de desmatamento, como mostrada na imagem TM/Landsat obtida em janeiro de 2000, data próxima da obtenção dos dados SRTM (fevereiro de 2000). Nessa situação, tem-se que a conformidade com o MDE não pode ser entendida como o um indício de acerto, mas sim como um erro de classificação, ocasionado devido a um erro inerente ao dado utilizado. Situação semelhante foi observada por Bwangoy et al. (2010) no mapeamento de áreas alagáveis. Os autores relataram anomalias no MDE que afetaram os atributos extraídos a partir da elevação. Essas anomalias estavam associadas às regiões em que a planície de inundação estava conectada diretamente a florestas degradadas ou áreas de savana. Os autores indicaram que em geral, esse problema foi contornado a partir da utilização de dados ópticos em conjunto com os dados topográficos, indicando a complementaridade desses dados na detecção de áreas alagáveis.

Figura 4.15 - Região identificada como área de desmatamento na Área 3 e classificada erroneamente como área alagável (representada pelo contorno em azul).



Fonte: Produção do autor.

5 CONCLUSÃO

Esta pesquisa propôs uma metodologia para o mapeamento das áreas alagáveis na bacia Amazônica baseada na extração de atributos a partir do MDE-NASADEM. Ao explorar a utilização desses atributos, confirma-se a hipótese de que eles podem representar o fenômeno do alagamento, auxiliando assim no mapeamento das áreas alagáveis. Nesta pesquisa, o RF foi utilizado para a classificação combinado com um processo iterativo de amostragem e com a seleção de atributos realizada a partir da métrica de importância *Mean Decrease Accuracy* (MDA). Nesse sentido, esta pesquisa contribuiu com: (i) a investigação do uso da entropia de Shannon para orientar o processo de amostragem; (ii) a avaliação da estabilidade da métrica MDA utilizada para ranquear os atributos no processo de seleção; (iii) a possibilidade de se reduzir o conjunto total de atributos a um subconjunto para compor um modelo reduzido; e (iv) a compreensão do impacto de se reduzir a quantidade de atributos na classificação.

A utilização da entropia de Shannon para a seleção do conjunto amostral foi uma contribuição metodológica desta pesquisa. Esse processo teve por objetivo a obtenção de amostras mais representativas, uma vez que a cada iteração, a modificação do conjunto amostral foi associada a uma avaliação da incerteza do mapeamento, com o intuito de representar melhor determinadas características da área para o processo de classificação. É importante destacar que comparações do desempenho dessa abordagem com outros métodos de amostragem comumente utilizados (e.g. amostragem aleatória) não foram realizadas nesta pesquisa, o que gera oportunidades para estudos futuros.

Com relação à análise da estabilidade da métrica MDA e do ranqueamento dos atributos baseada em múltiplas iterações do RF, tem-se que a utilização do MDA para a quantificação da importância levou a uma maior discriminação entre as importâncias para os atributos que ocuparam posições mais altas no ranqueamento, ou seja, é possível afirmar com maior certeza que um atributo é mais importante que outro entre os atributos que ocupam posições mais altas. Por outro lado, para os atributos que em média ocuparam as menores posições,

notou-se uma maior instabilidade da posição ocupada, o que significa que, a partir de uma determinada posição no ranqueamento, os atributos competem entre si, não sendo possível afirmar o nível de importância relativa entre eles. Assim, se há o interesse em ranquear os atributos utilizando a métrica MDA, recomenda-se gerar um ranqueamento médio a partir de múltiplas rodadas do RF, para lidar com a instabilidade.

A utilização da metodologia proposta para selecionar os atributos demonstrou a possibilidade de se obter um modelo reduzido, formado pelos atributos considerados mais importantes. Durante a etapa de pré-seleção dos atributos morfométricos, observou-se a importância da informação contextual. Dessa forma, este estudo destaca a relevância de se representar um atributo considerando diferentes tamanhos de janelas em sua obtenção, pois pode acontecer de um atributo não ser considerado importante para a classificação, não porque ele não é relevante para o processo, mas sim porque ele não está sendo representado da maneira adequada, ou seja, não está capturando a informação realmente relevante para descrever o fenômeno estudado.

O modelo reduzido obtido contém 9 atributos e resultou em uma classificação com acurácia de 97,61% e visualmente coerente quando analisada em conjunto com o MDE-NASADEM. De maneira geral, a seleção dos atributos não levou a mudanças expressivas no mapeamento quando comparado visualmente e por meio das métricas tradicionais de avaliação, com a classificação obtida utilizando todos os 124 atributos inicialmente propostos. Além disso, notou-se uma redução considerável da incerteza de mapeamento ao se reduzir os atributos, indicando que o modelo convergiu para o resultado com uma maior certeza associada. Além disso, reduzir os atributos facilitou o entendimento de quais variáveis são mais relevantes e de que forma elas estão relacionadas com o fenômeno em estudo. Assim, recomenda-se a implementação do processo de seleção dos atributos quando além do interesse de obter um mapeamento com uma acurácia satisfatória, há o interesse de entender quais variáveis são realmente importantes para o processo de classificação e investigar a relação delas com o fenômeno estudado. É importante destacar que, no caso deste estudo, reduzir a quantidade de atributos facilita a ampliação da metodologia aqui proposta para

a região da bacia Amazônica, pois a redução dos atributos resulta em uma redução do custo computacional envolvido na extração, processamento e armazenamento dos dados.

De maneira geral, a metodologia proposta nesta pesquisa resultou em um mapeamento coerente com o MDE-NASADEM para as áreas de estudo, indicando melhorias na representação de diferentes regiões em relação à *wetmask*. Entretanto, limitações são reconhecidas. A ausência de dados *in situ* que validem os resultados obtidos não permite indicar com certeza a correta atribuição das classes. Além disso, como só foram utilizados atributos extraídos do MDE, o resultado é mais influenciado por erros de representação desse dado. Dessa forma, estudos futuros devem incluir dados ópticos e/ou dados de radar para investigar a complementariedade dos dados e possivelmente reduzir erros de classificação.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAM, E.; MUTANGA, O.; RUGEGE, D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. **Wetlands Ecology and Management**, v. 18, n. 3, p. 281–296, 2010.

BANON, G. P. R.; BANON, G. J. F.; VILLAMARÍN, F.; ARRAUT, E. M.; MOULATLET, G. M.; RENNÓ, C. D.; BANON, L. C.; MARIONI, B.; NOVO, E. M. L. D. M. Predicting suitable nesting sites for the Black caiman (Melanosuchus niger Spix 1825) in the Central Amazon basin. **Neotropical Biodiversity**, v. 5, n. 1, p. 47–59, 2019.

BANON, L. C. **Árvores de decisão aplicadas à extração automática de redes de drenagem**. 2013. 113 p. Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013.

BANON, L. C.; NOVO, E. M. L. M. **Extração da rede de drenagem da Bacia Amazônica por um processo de mineração de dados utilizando MDE-SRTM e avaliação do resultado em aplicações nas áreas de geomorfologia, geologia, ecohidrologia e distribuição de espécies (projeto FAPESP 2016/13462-0)**. 2018. Disponível em: <http://www.dsr.inpe.br/amazondrainage>.

BANON, L. C.; SANTOS, R. D. C. DOS; VIJAYKUMAR, N. L.; RENNÓ, C. D. Definição de critérios a partir da mineração de dados para a extração automática de redes de drenagem. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2013, Foz do Iguaçu, PR. **Anais...** São José dos Campos: INPE, 2013.

BELGIU, M.; DRAGUT, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BERHANE, T. M.; LANE, C. R.; WU, Q.; ANENKHONOV, O. A.; CHEPINOGA, V. V.; AUTREY, B. C.; LIU, H. Comparing pixel- and object-based approaches in effectively classifying wetland-dominated landscapes. **Remote Sensing**, v. 10, n. 1, 2018a.

BERHANE, T. M.; LANE, C. R.; WU, Q.; AUTREY, B. C.; ANENKHONOV, O. A.; CHEPINOGA, V. V.; LIU, H. Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. **Remote Sensing**, v. 10, n. 4, 2018b.

BEVEN, K. J.; KIRKBY, M. J. A physically based, variable contributing area model of basin hydrology. **Hydrological Sciences Bulletin**, v. 24, n. 1, p. 43–69, 1979.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5–32, 2001.

BROWN, K. M.; FOODY, G. M.; ATKINSON, P. M. Estimating per-pixel thematic uncertainty in remote sensing classifications. **International Journal of Remote Sensing**, v. 30, n. 1, p. 209–229, 2008.

BUARQUE, D. C.; FAN, F. M.; PAZ, A. R.; COLLISCHONN, W. Comparação de métodos para definir direções de escoamento a partir de modelos digitais de elevação. **Revista Brasileira de Recursos Hídricos**, v. 14, n. 2, p. 91–103, 2009.

BWANGOY, J. R. B.; HANSEN, M. C.; ROY, D. P.; GRANDI, G. DE; JUSTICE, C. O. Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. **Remote Sensing of Environment**, v. 114, n. 1, p. 73–86, 2010.

CAI, J.; LUO, J.; WANG, S.; YANG, S. Feature selection in machine learning: a new perspective. **Neurocomputing**, v. 300, p. 70–79, 2018.

CAI, Y.; LI, X.; ZHANG, M.; LIN, H. Mapping wetland using the object-based stacked generalization method based on multi-temporal optical and SAR data. **International Journal of Applied Earth Observation and Geoinformation**, v. 92, p. 102164, 2020.

CASTELLO, L.; MCGRATH, D. G.; HESS, L. L.; COE, M. T.; LEFEBVRE, P. A.; PETRY, P.; MACEDO, M. N.; REN, V. F.; ARANTES, C. C. The vulnerability of Amazon freshwater ecosystems. **Conservation Letters**, v. 6, n.4, p. 217–229, 2013.

CATANI, F.; LAGOMARSINO, D.; SEGONI, S.; TOFANI, V. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. **Natural Hazards and Earth System Sciences**, v. 13, n. 11, p. 2815–2831, 2013.

CHEN, Y.; HE, X.; WANG, J.; XIAO, R. The influence of polarimetric parameters and an object-based approach on land cover classification in coastal wetlands. **Remote Sensing**, v. 6, n. 12, p. 12575–12592, 2014.

CORCORAN, J. M.; KNIGHT, J. F.; GALLANT, A. L. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota. **Remote Sensing**, v. 5, n. 7, p. 3212–3238, 2013.

CORREIA, A. H. Metodologias e resultados preliminares do projeto Radiografia da Amazônia. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO (SBSR), 15., 2011. **Anais...** São José dos Campos: INPE, 2011.

COSTANZA, R.; DE GROOT, R.; SUTTON, P.; VAN DER PLOEG, S.; ANDERSON, S. J.; KUBISZEWSKI, I.; FARBER, S.; TURNER, R. K. Changes in the global value of ecosystem services. **Global Environmental Change**, v. 26, n. 1, p. 152–158, 2014.

DEGENHARDT, F.; SEIFERT, S.; SZYMCZAK, S. Evaluation of variable selection methods for random forests and omics data sets. **Briefings in Bioinformatics**, v. 20, n. 2, p. 492–503, 2019.

DÍAZ-URIARTE, R.; ALVAREZ DE ANDRÉS, S. Gene selection and classification of microarray data using random forest. **BMC Bioinformatics**, v. 7, p. 1–13, 2006.

DUBEAU, P.; KING, D. J.; UNBUSHE, D. G.; REBELO, L. M. Mapping the Dabus Wetlands, Ethiopia, using random forest classification of Landsat, PALSAR and topographic data. **Remote Sensing**, v. 9, n. 10, p. 1–23, 2017.

FARR, T. G.; ROSEN, P. A.; EDWARD, C.; CRIPPEN, R.; DUREN, R.; HENSLEY, S.; KOBRICK, M.; PALLER, M.; RODRIGUEZ, E.; ROTH, L.; SEAL, D.; SHAFFER, S.; SHIMADA, J.; UMLAND, J.; WERNER, M.; OSKIN, M.; BURBANK, D.; ALSDORF, D. The shuttle radar topography mission. **Reviews of Geophysics**, v. 45, 2007.

FINLAYSON, M. C.; DAVIDSON, N.; PRITCHARD, D.; RANDY MILTON, G.; MACKACY, H. The ramsar convention and ecosystem-based approaches to the wise use and sustainable development of wetlands. **Journal of International Wildlife Law and Policy**, v. 14, n. 3/4, p. 176–198, 2011.

FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, v. 80, n. 1, p. 185–201, 2002.

FOODY, G. M. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. **Photogrammetric Engineering and Remote Sensing**, v. 70, n. 5, p. 627–633, 2004.

FOX, E. W.; HILL, R. A.; LEIBOWITZ, S. G.; OLSEN, A. R.; THORNBRUGH, D. J.; WEBER, M. H. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. **Environmental Monitoring and Assessment**, v. 189, n. 7, 2017.

GALIANO, V. F. R.; GHIMIRE, B.; ROGAN, J.; OLMO, M. C.; SANCHEZ, J. P. R. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 67, n. 1, p. 93–104, 2012.

GALLANT, A. L. The challenges of remote monitoring of wetlands. **Remote Sensing**, v. 7, n. 8, p. 10938–10950, 2015.

GENUER, R.; POGGI, J. M.; TULEAU-MALOT, C. Variable selection using random forests. **Pattern Recognition Letters**, v. 31, n. 14, p. 2225–2236, 2010.

GONÇALVES, L. M.; FONTE, C. C.; JÚLIO, E. N. B. S.; CAETANO, M. Evaluation of remote sensing images classifiers with uncertainty measures. In: RODOLPHE, D; HELEN, G. (Ed.). **Spatial data quality from process to decisions**. Boca Raton: CRC Press, 2009. p. 163–177.

GREGORUTTI, B.; MICHEL, B.; SAINT-PIERRE, P. Correlation and variable importance in random forests. **Statistics and Computing**, v. 27, n. 3, p. 659–678, 2017.

GUO, M.; LI, J.; SHENG, C.; XU, J.; WU, L. A review of wetland remote sensing. **Sensors**, v. 17, n. 4, p. 1–36, 2017.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**. 2. ed. New York: Springer, 2009.

HESS, L. L.; MELACK, J. M.; AFFONSO, A. G.; BARBOSA, C.; GASTIL-BUHL, M.; NOVO, E. M. L. M. Wetlands of the lowland Amazon basin: extent, vegetative cover, and dual-season inundated area as mapped with JERS-1 Synthetic Aperture Radar. **Wetlands**, v. 35, n. 4, p. 745–756, 2015.

HESS, L. L.; MELACK, J. M.; NOVO, E. M. L. M.; BARBOSA, C. C. F.; GASTIL, M. Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. **Remote Sensing of Environment**, v. 87, n. 4, p. 404–428, 2003.

JARDIM, A. C. **Direções de fluxo em modelos digitais de elevação:** um método com foco na qualidade da estimativa e processamento de grande volume de dados. 2017. 133 p. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2017.

JENSON, S. K. Applications of hydrologic information automatically extracted from digital elevation models. **Hydrological Processes**, v. 5, n. 1, p. 31–44, 1991.

JENSON, S. K.; DOMINGUE, J. O. Extracting topographic structure from digital elevation data for geographic information system analysis. **Photogrammetric Engineering and Remote Sensing**, v. 54, n. 11, p. 1593–1600, 1988.

JUNK, W. J. **Flood tolerance and tree distribution in central Amazonian floodplains**. [S.l.]: Academic Press, 1989. 47–64 p.

JUNK, W. J.; BAYLEY, P. B.; SPARKS, R. E. The flood pulse concept in river-floodplain systems. **Canadian Special Publications for Fisheries and Aquatic Sciences**, v. 106, p. 110–127, 1989.

JUNK, W. J.; PIEDADE, M. T. F.; LOURIVAL, R.; WITTMANN, F.; KANDUS, P.; LACERDA, L. D.; BOZELLI, R. L.; ESTEVES, F. A.; NUNES DA CUNHA, C.; MALTCHIK, L.; SCHÖNGART, J.; SCHAEFFER-NOVELLI, Y.; AGOSTINHO, A. A. Brazilian wetlands: their definition, delineation, and classification for research, sustainable management, and protection. **Aquatic Conservation: Marine and Freshwater Ecosystems**, v. 24, n. 1, p. 5–22, 2013.

JUNK, W. J.; PIEDADE, M. T. F.; SCHÖNGART, J.; COHN-HAFT, M.; ADENEY, J. M.; WITTMANN, F. A classification of major naturally-occurring amazonian lowland wetlands. **Wetlands**, v. 31, n. 4, p. 623–640, 2011.

KARLSON, M.; OSTWALD, M.; REESE, H.; SANOU, J.; TANKOANO, B.; MATTSSON, E. Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and random forest. **Remote Sensing**, v. 7, n. 8, p. 10017–10041, 2015.

KINGSFORD, R. T. Conservation management of rivers and wetlands under climate change: a synthesis. **Marine and Freshwater Research**, v. 62, n. 3, p. 217–222, 2011.

LIAW, A.; WIENER, M. Classification and regression by randomForest. **R News**, v. 2, n. 3, p. 18–22, 2002.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 5., 1967. **Proceedings...** Berkeley, CA: University of California Press, 1967.

MAHDAVI, S.; SALEHI, B.; GRANGER, J.; AMANI, M.; BRISCO, B.; HUANG, W. Remote sensing for wetland classification: a comprehensive review. **GIScience and Remote Sensing**, v. 55, n. 5, p. 623–658, 2018.

MAHDIANPARI, M.; SALEHI, B.; MOHAMMADIMANESH, F.; MOTAGH, M. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 13–31, 2017.

MARTINS, V. S.; KALEITA, A. L.; GELDER, B. K.; NAGEL, G. W.; MACIEL, D. A. Deep neural network for complex open-water wetland mapping using high-resolution WorldView-3 and airborne LiDAR data. **International Journal of Applied Earth Observation and Geoinformation**, v. 93, p. 102215, 2020.

MELACK, J. M.; COE, M. T. Climate change and the floodplain lakes of the Amazon basin. In: GOLDMAN, C. R.; KUMAGAI, M.; ROBARTS, R. D. **Climatic change and global warming of inland waters**: impacts and mitigation for ecosystems and societies. [S.l.]: Wiley, 2013. p. 295–310.

MELACK, J. M.; HESS, L. L.; GASTIL, M.; FORSBERG, B. R.; HAMILTON, S. K.; LIMA, I. B. T.; NOVO, E. M. L. M. Regionalization of methane emissions in the Amazon Basin with microwave remote sensing. **Global Change Biology**, v. 10, n. 5, p. 530–544, 2004.

MEYER, H.; LEHNERT, L. W.; WANG, Y.; REUDENBACH, C.; NAUSS, T.; BENDIX, J. From local spectral measurements to maps of vegetation cover and biomass on the Qinghai-Tibet-Plateau: do we need hyperspectral information? **International Journal of Applied Earth Observation and Geoinformation**, v. 55, p. 21–31, 2017.

MILLARD, K.; RICHARDSON, M. Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR-SAR fusion using a random forest classifier. **Canadian Journal of Remote Sensing**, v. 39, n. 4, p. 290–307, 2013.

MILLARD, K.; RICHARDSON, M. On the importance of training data sample selection in Random Forest image classification: a case study in peatland ecosystem mapping. **Remote Sensing**, v. 7, n. 7, p. 8489–8515, 2015.

MITSCH, W. J.; BERNAL, B.; HERNANDEZ, M. E. Ecosystem services of wetlands. **International Journal of Biodiversity Science, Ecosystem Services and Management**, v. 11, n. 1, p. 1–4, 2015.

MITSCH, W. J.; BERNAL, B.; NAHLIK, A. M.; MANDER, Ü.; ZHANG, L.; ANDERSON, C. J.; JØRGENSEN, S. E.; BRIX, H. Wetlands, carbon, and climate change. **Landscape Ecology**, v. 28, n. 4, p. 583–597, 2013.

MOORE, I. D.; GRAYSON, R. B.; LADSON, A. R. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. **Hydrological Processes**, v. 5, n. 1, p. 3–30, 1991.

OZESMI, S. L.; BAUER, M. E. Satellite remote sensing of wetlands. **Wetlands Ecology and Management**, v. 10, n. 5, p. 381–402, 2002.

PIEDADE, M. T. F.; SCHÖNGART, J.; WITTMANN, F.; PAROLIN, P.; JUNK, W. J. Impactos ecológicos da inundação e seca na vegetação das áreas alagáveis amazônicas. In: BORMA, L. S.; NOBRE, C. (Ed.). **Secas na Amazônia:** causas e consequências. São Paulo: Oficina de Textos, 2013. p. 409–461.

PIKE, R. J.; EVANS, I. S.; HENGL, T. Geomorphometry: a brief guide. **Developments in Soil Science**, v. 33, n. C, p. 3–30, 2009.

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, n. 9, p. 3469–3481, 2008.

RENÓ, V. F.; NOVO, E. M. L. M.; SUEMITSU, C.; RENNÓ, C. D.; SILVA, T. S. F. Assessment of deforestation in the Lower Amazon floodplain using historical Landsat MSS/TM imagery. **Remote Sensing of Environment**, v. 115, n. 12, p. 3446–3456, 2011.

RICHEY, J. E.; MELACK, J. M.; AUFDENKAMPE, A. K.; BALLESTER, V. M.; HESS, L. L. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO2. **Nature**, v. 416, n. 6881, p. 617–620, 2002.

RILEY, S. J.; DEGLORIA, S. D.; ELLIOT, R. A Terrain Ruggedness Index that quantifies topographic heterogeneity. **Intermountain Journal of Sciences**, v. 5, n. 1–4, p. 23–27, 1999.

ROBERTSON, L. D.; KING, D. J.; DAVIES, C. Object-based image analysis of optical and radar variables for wetland evaluation. **International Journal of Remote Sensing**, v. 36, n. 23, p. 5811–5841, 2015.

ROSA, S. A.; SILVA, A. F.; CASTRO, N.; FEITOSA, Y. O.; PIEDADE, M. T. F. Entre a água e a terra: Áreas Úmidas (AUs). In: LOPES, A.; PIEDADE, M. T. F. (Ed.). **Conhecendo as áreas úmidas amazônicas: uma viagem pelas várzeas e igapós**. Manaus: INPA, 2015. p. 23-31.

ROSIM, S.; MONTEIRO, A. M. V.; RENNÓ, C. D.; SOUZA, R. C. M.; SOARES, J. V. TerraHidro - uma plataforma computacional para o desenvolvimento de aplicativos para a análise integrada de recursos hídricos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11., 2003. **Anais...** São José dos Campos: INPE, 2003. p. 2589–2596.

SLATER, J. A.; GARVEY, G.; JOHNSTON, C.; HAASE, J.; HEADY, B.; KROENUNG, G.; LITTLE, J. The SRTM data “finishing” process and products. **Photogrammetric Engineering and Remote Sensing**, v. 72, n. 3, p. 237–247, 2006.

SPEISER, J. L.; MILLER, M. E.; TOOZE, J.; IP, E. A comparison of random forest variable selection methods for classification prediction modeling. **Expert Systems with Applications**, v. 134, p. 93–101, 2019.

TAVARES DA COSTA, R.; MAZZOLI, P.; BAGLI, S. Limitations posed by free DEMs in watershed studies: the case of river Tanaro in Italy. **Frontiers in Earth Science**, v. 7, June 2019.

VALERIANO, M. M.; KUPLICH, T. M.; STORINO, M.; AMARAL, B. D.; MENDES, J. N.; LIMA, D. J. Modeling small watersheds in Brazilian Amazonia with shuttle radar topographic mission-90 m data. **Computers and Geosciences**, v. 32, n. 8, p. 1169–1181, 2006.

VAN ZYL, J. J. The shuttle radar topography mission (SRTM): a breakthrough in remote sensing of topography. **Acta Astronautica**, v. 48, n. 5–12, p. 559–565, 2001.

WEISS, A. D. Topographic position and landforms analysis. In: ESRI USER CONFERENCE, 2001, San Diego, CA. **Proceedings…** 2001. Poster presentation.

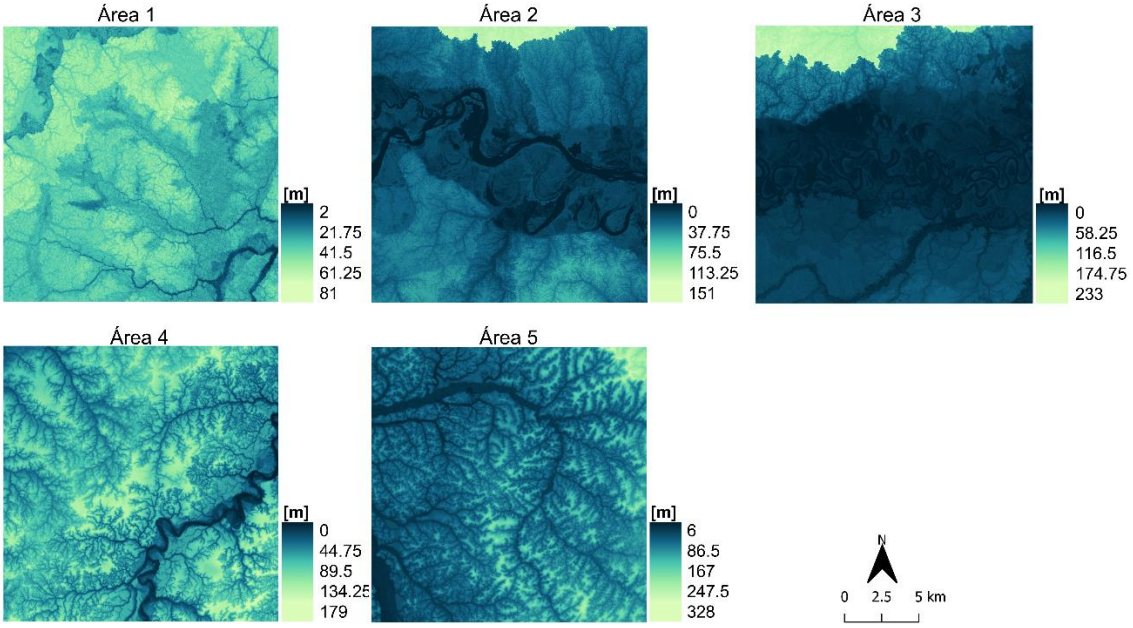
WHITE, L.; BRISCO, B.; DABBOOR, M.; SCHMITT, A.; PRATT, A. A collection of SAR methodologies for monitoring wetlands. **Remote Sensing**, v. 7, n.6, p. 7615–7645, 2015.

WOOD, J. **The geomorphological characterisation of digital elevation models.**1996. 466p. Thesis (PhD) - University of Leicester, Leicester, UK, 1996.

ZHU, Y.; LIU, K.; LIU, L.; MYINT, S. W.; WANG, S.; LIU, H.; HE, Z. Exploring the potential of world view-2 red-edge band-based vegetation indices for estimation of mangrove leaf area index with machine learning algorithms. **Remote Sensing**, v. 9, n. 10, 2017.

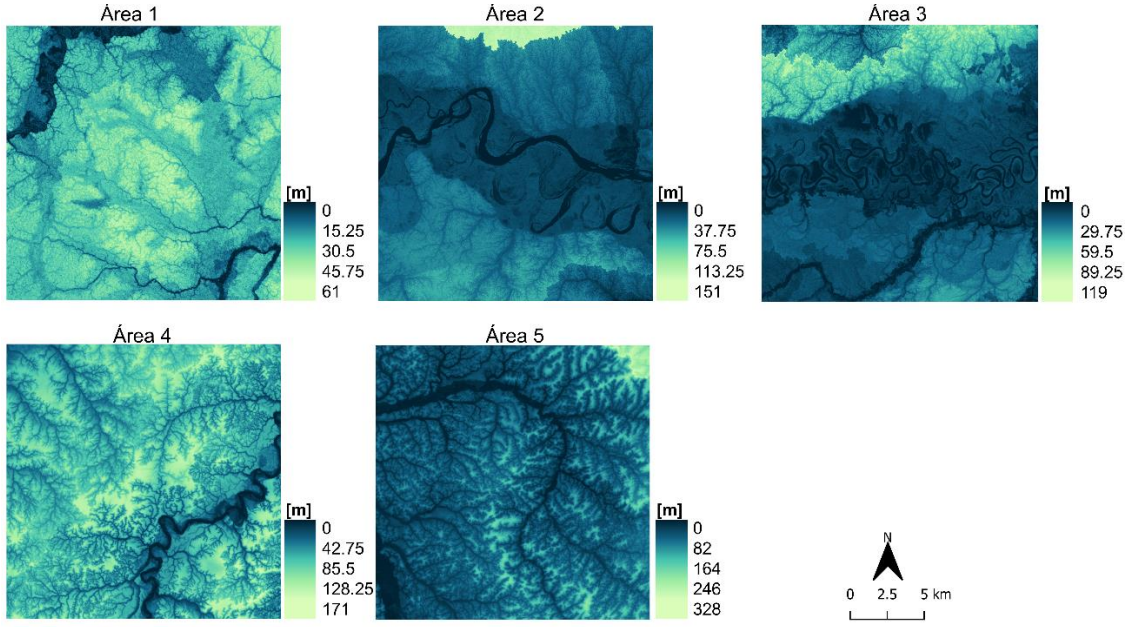
APÊNDICE A – ATRIBUTOS SELECIONADOS

Figura A.1 – Atributo HAND SWBD para as cinco áreas de estudo.



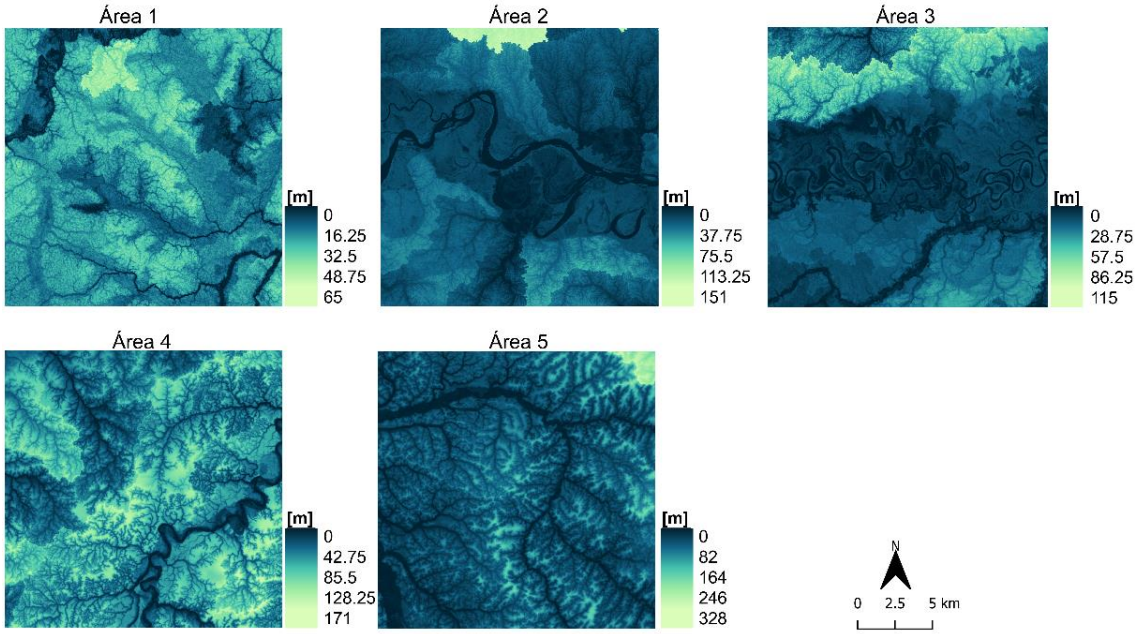
Fonte: Produção do autor.

Figura A.2 – Atributo HAND (Ord ≥ 7) para as cinco áreas de estudo.



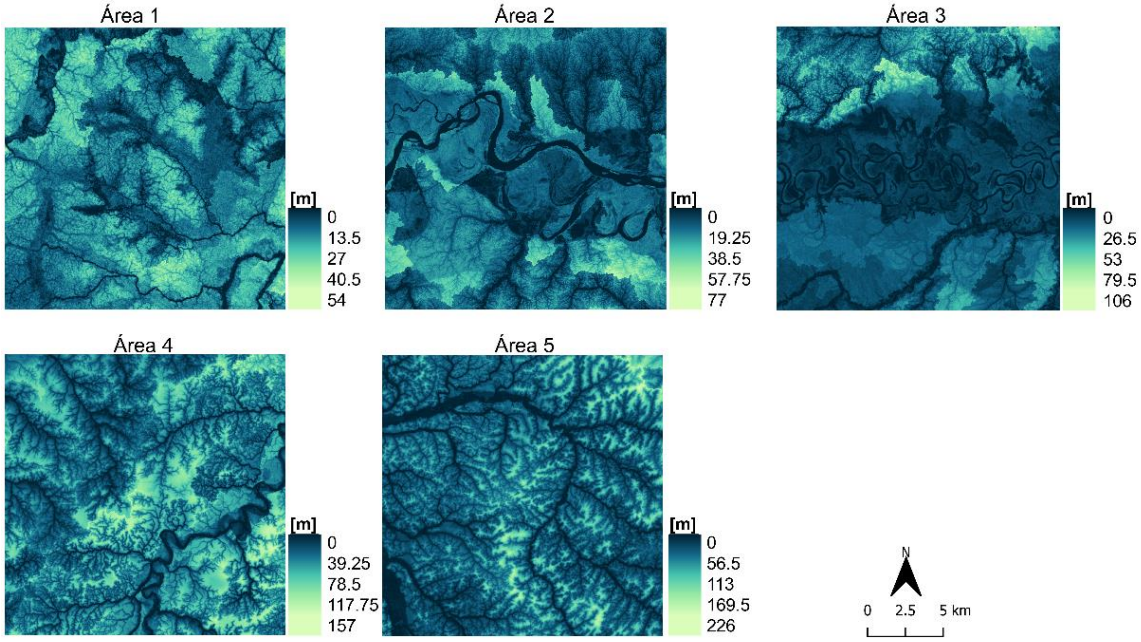
Fonte: Produção do autor.

Figura A.3 – Atributo HAND (Ord ≥ 6) para as cinco áreas de estudo.



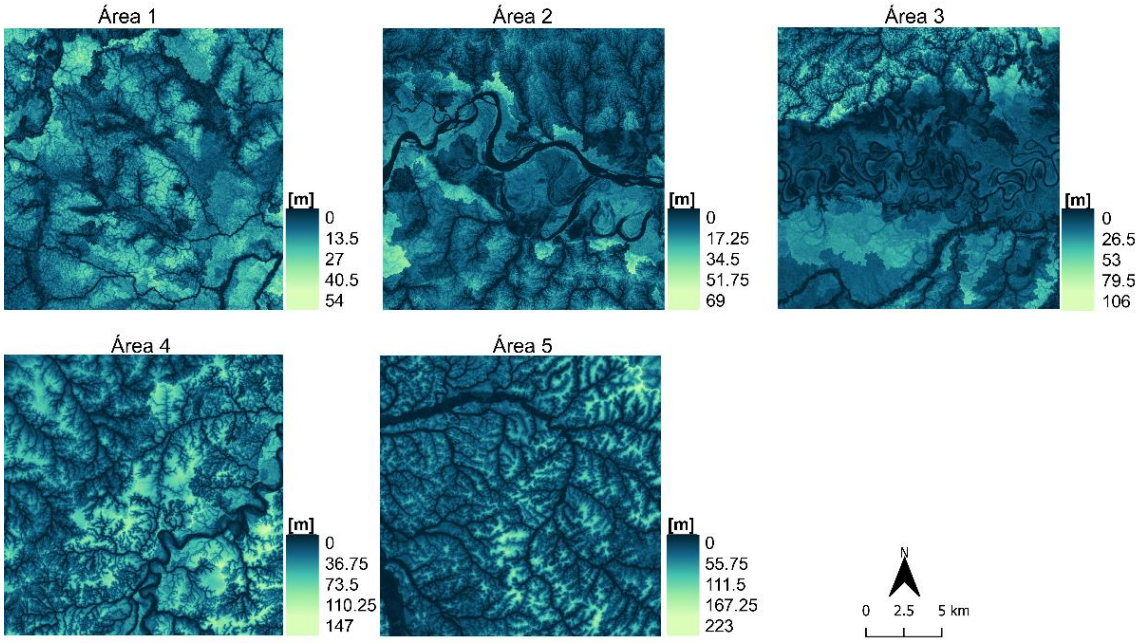
Fonte: Produção do autor.

Figura A.4 – Atributo HAND (Ord ≥ 5) para as cinco áreas de estudo.



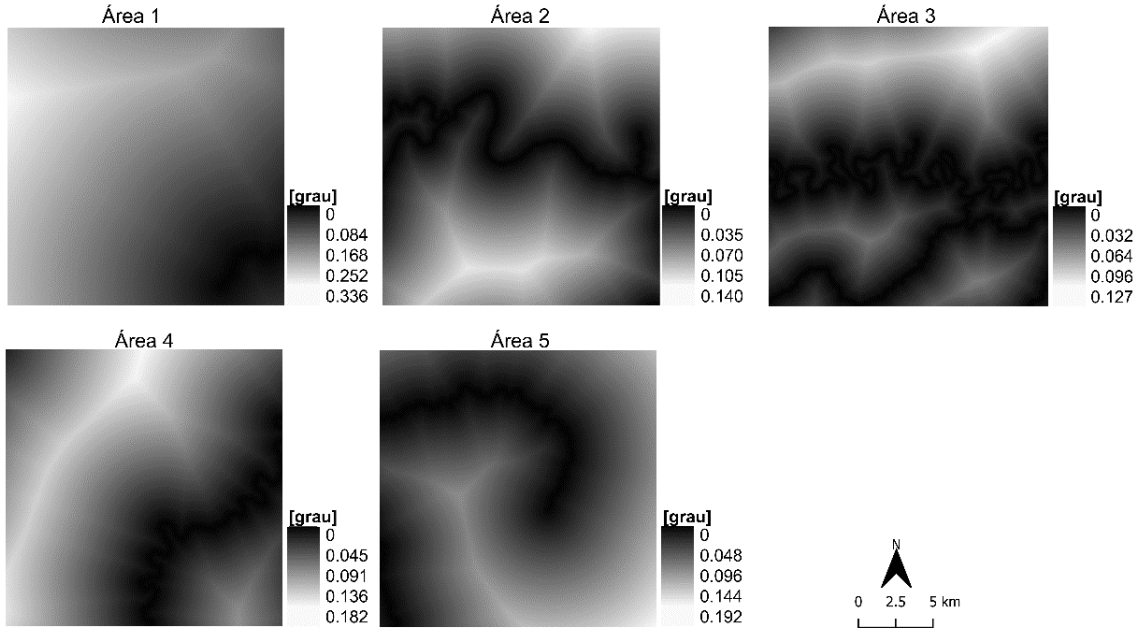
Fonte: Produção do autor.

Figura A.5 – Atributo HAND (Ord ≥ 4) para as cinco áreas de estudo.



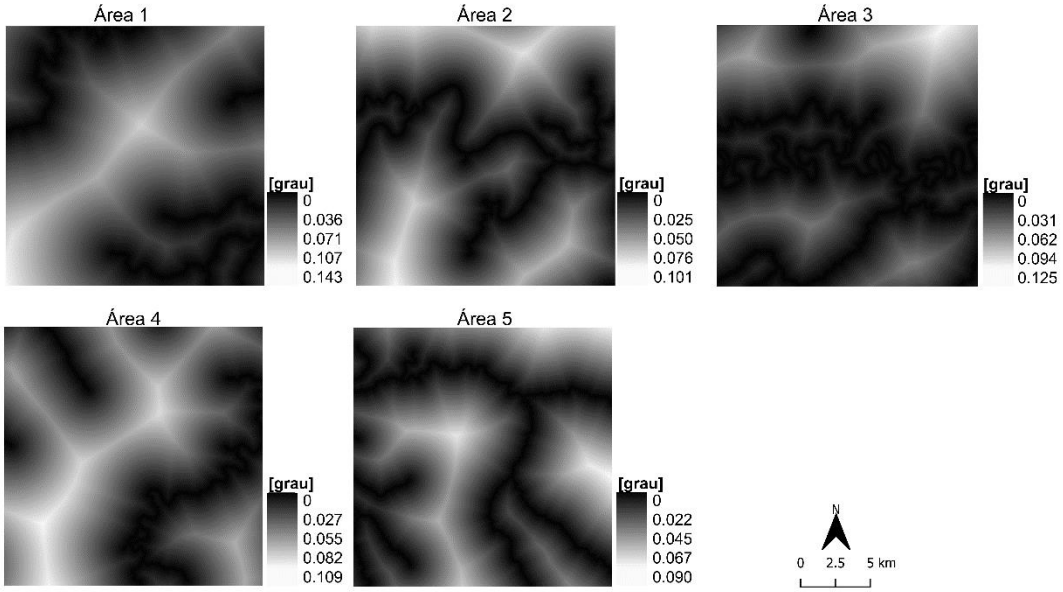
Fonte: Produção do autor.

Figura A.6 – Atributo DIRDMP (Ord ≥ 7) para as cinco áreas de estudo.



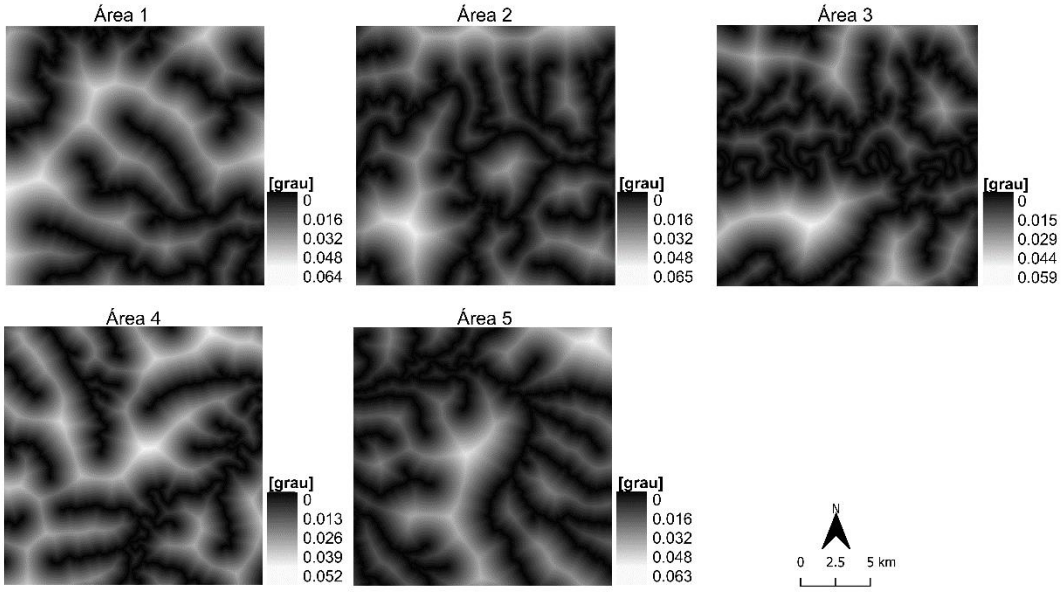
Fonte: Produção do autor.

Figura A.7 – Atributo DIRDMP (Ord ≥ 6) para as cinco áreas de estudo.



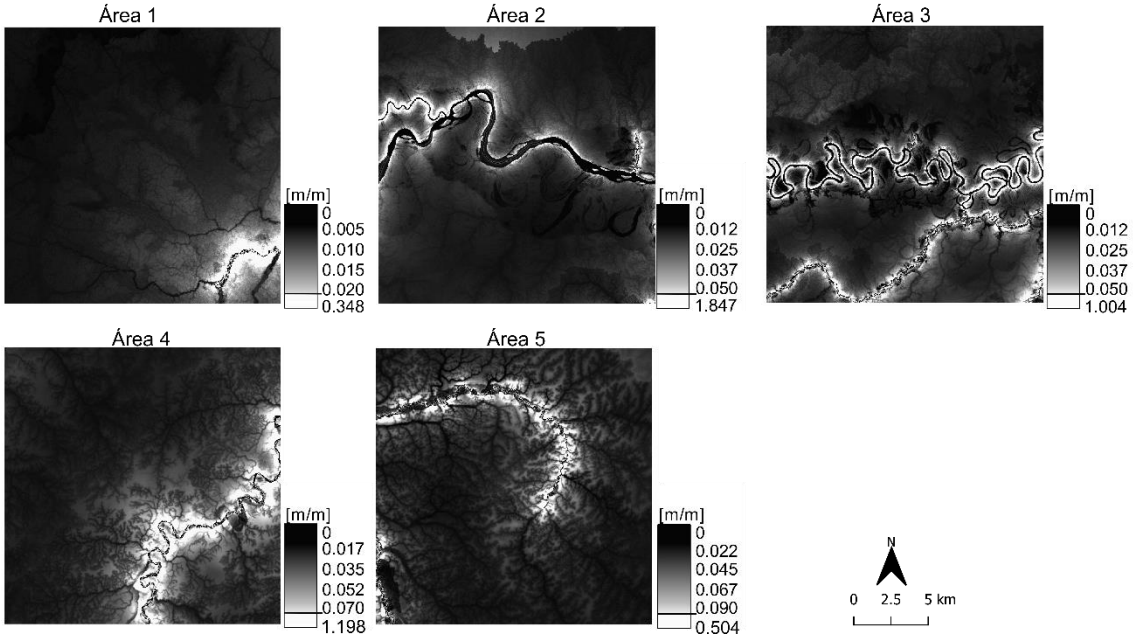
Fonte: Produção do autor.

Figura A.8 – Atributo DIRDMP (Ord ≥ 5) para as cinco áreas de estudo.



Fonte: Produção do autor.

Figura A.9 – Atributo DERDMP (Ord ≥ 7) para as cinco áreas de estudo.



Fonte: Produção do autor.