

# MAPEAMENTO DE ÁREAS ALAGÁVEIS EM DUAS REGIÕES DA BACIA AMAZÔNICA E A CONSTRUÇÃO DE UMA MEDIDA DE INCERTEZA ESPACIALMENTE DISTRIBUÍDA

Trabalho final da disciplina de Introdução ao Geoprocessamento (SER-300) do curso de Mestrado em Sensoriamento Remoto ministrada pelos Dr. Antônio Miguel Vieira Monteiro e Dr. Claudio Barbosa.

Deborah Lopes Correia Lima

INPE

São José dos Campos

2020

1	IN	TRO	DUÇÃO	1
2	Me	etodo	ologia	3
	2.1	Vis	são geral	3
	2.2	Da	dos de entrada	5
	2.2	2.1	Wetmask	5
	2.2	2.2	Rede de drenagem	6
	2.2	2.3	MDE	7
	2.3	Pré	é-processamento do MDE	8
	2.3	8.1	Local Drain Directions (LDD)	8
	2.3	8.2	Área acumulada	9
	2.4	Atr	ributos hidrológicos e morfométricos	10
	2.4	.1	Área acumulada expandida	10
	2.4	.2	Declividade	10
	2.4	.3	HAND	11
	2.4	.4	Distâncias relativas à rede de drenagem: euclidiana e caminho de	e fluxo 12
	2.5	Am	nostragem	13
	2.6	Rai	ndom Forest	15
	2.7	Ent	tropia de Shannon	15
3	Re	sulta	ndos e discussão	17
	3.1	Ob	tenção de parâmetros	17
	3.2	Cla	assificação pelo algoritmo Random Forest	
	3.3	Ob	tenção dos mapas de incerteza	
4	Co	nclu	sões	
5	Re	ferêr	ncias bibliográficas	

# SUMÁRIO

# 1 INTRODUÇÃO

A bacia do Rio Amazonas é a maior bacia hidrográfica do mundo, com área de aproximadamente 6,5 milhões de km<sup>2</sup> (BARBOSA, 2007). Em função da sazonalidade da precipitação ao longo da bacia, a região possui uma estação de cheia e uma estação de seca bem definidas durante o ano. Porém, a variabilidade climática e hidrográfica ao longo da região é responsável por diferenças entre as cheias e as secas, além da amplitude média dos rios, entre regiões da bacia Amazônica (PIEDADE et al., 2013).

As características geomorfológicas da bacia Amazônica propiciam a formação de uma vasta planície de inundação, formando as chamadas áreas alagáveis (BARBOSA, 2007). Essas áreas, também chamadas de *wetlands*, são regiões onde ocorrem oscilações periódicas do nível da água, em combinação com mudanças entre a fase aquática e a fase terrestre, o que cria condições específicas para esses ambientes (JUNK ,1890).

As áreas alagáveis do Rio Amazonas formam um conjunto de habitats que vai desde ambientes permanentemente aquáticos até aqueles permanentemente terrestres (JUNK, 1989). A oscilação do nível da água entre a fase de cheia e de seca da região Amazônica tem papel fundamental na ciclagem de nutrientes e na estocagem de grandes quantidades de água nas planícies de inundação (PIEDADE et al., 2013). Esses ambientes são ricos em biodiversidade, e desempenham um significativo papel socioeconômico na região. São sistemas dinâmicos, cuja composição e extensão varia rapidamente de acordo com questões hidrológicas e climáticas (FREITAS; NOVO, 2005).

A observação dos ambientes de áreas alagáveis, bem como de outros sistemas aquáticos de águas interiores, revelam informações importantes para o entendimento de processos hidrológicos da superfície (ALSDORF, 2003). Devido à grande extensão das áreas inundáveis da Amazônia, o sensoriamento remoto surge como uma ferramenta útil para o mapeamento desses ambientes, dado que a partir de informações obtidas de sensores remotos orbitais é possível produzir mapas temáticos da superfície terrestre.

A obtenção de informações a partir de sensores remotos ópticos é limitada, principalmente em regiões tropicais por conta da alta cobertura de nuvem. Nesse sentido, a utilização de sensores de micro-ondas surge como uma boa alternativa pelo fato de

serem menos influenciados pela presença de nuvens além de tornarem possível a penetração do sinal na vegetação (HESS et al., 2003).

Hess et al. (2003), utilizando imagens SAR (*Synthetic Aperture Radar*) adquiridas pelo satélite japonês JERS-1, produziram um importante mapeamento de áreas alagáveis da parte central da bacia Amazônica. Para realizar o mapeamento foram utilizados dois conjuntos de imagens, um no período de cheia e outro no período de seca. Para validação do mapeamento foram utilizados dados de sobrevoo da área de estudo e as imagens resultantes possuem resolução espacial de aproximadamente 100 m. Importante destacar que a definição de *wetlands* dada pelos autores citados acima foi baseada em duas situações, (1) as áreas que estavam inundadas no período de aquisição das imagens (durante a cheia e/ou a seca), e (2) áreas que não apareceram como inundadas nas imagens, mas que apresentam geomorfologia relativa à de áreas alagadas e que são próximas ou cercadas por *wetlands*.

As imagens utilizadas por Hess et al. (2003) para elaboração do mapa de áreas alagáveis nem sempre refletiram o período de cheias de toda a região, considerando a variação desse período ao longo da planície amazônica, o pode ocasionar a não representação real da extensão das áreas alagáveis (ALFAYA, 2012). Além disso, para a obtenção do mapa classificado por Hess et al. (2003) não foi utilizado um Modelo Digital de Elevação (MDE), não sendo considerado assim, o relevo da região. Dessa forma, quando analisado o mapa de áreas alagáveis em conjunto com o MDE- SRTM (*Shuttle Radar Topography Mission*) da área, nota-se inconsistências em algumas regiões. Rennó et al. (2013) destacaram a existência de inconsistências, provavelmente resultantes do processo de geração do mosaico das imagens JERS-1, o que provoca deslocamentos dos polígonos do mapa.

Dessa forma, o mapa produzido por Hess et al. (2003) é um relevante dado de áreas alagáveis da região Amazônica, entretanto são percebidas inconsistências nesse mapa. Nesse sentindo, surge a necessidade de realizar mapeamentos que objetivem melhorar a representação das áreas alagáveis da bacia do Rio Amazonas. Porém, é importante destacar que todo processo de mapeamento temático utilizando dados de sensores remotos orbitais possuem erros inerentes ao processo de mapeamento.

Os erros atrelados a elaboração de mapas a partir de dados orbitais estão associados com questões, como por exemplo resolução espacial e temporal dos sensores. Além disso, todo processo de classificação está sujeito a erros. Esses aspectos que estão relacionados com a produção de mapas, refletem no resultado do mapeamento, interferindo assim no que o mapa estará representando.

Sendo assim, uma boa forma de avaliar o mapa obtido é relacioná-lo com medidas de incerteza resultantes do processo de classificação. Uma maneira de associar essas informações é por meio de mapas de incerteza, onde torna-se possível visualizar espacialmente o grau de certeza com que o classificador determinou as classes para cada ponto do mapa.

Com base no que foi exposto, este estudo objetivou produzir um mapeamento de áreas alagáveis mais consistente para duas áreas inseridas na região mapeada por Hess et al. (2003), com base em diferentes atributos morfométricos e hidrológicos. Além disso, objetivou-se produzir juntamente com a classificação, uma medida de incerteza espacialmente distribuída para avaliar a incerteza do mapeamento. Diante disso, este trabalho visa contribuir com o mapeamento de áreas alagáveis além de destacar a importância da construção e análise de medidas de incerteza.

#### 2 METODOLOGIA

#### 2.1 Visão geral

A área de estudo compreende duas regiões inseridas no mapeamento realizado por Hess et al. (2003), como exposto no mapa de localização (Figura 1), e cujas características constam na Tabela 1. Estas áreas foram escolhidas por serem regiões onde foram identificadas inconsistências quando analisado o mapa de áreas alagáveis produzido por Hess et al. (2003) em conjunto com o MDE-SRTM.



Figura 1 - Mapa de localização das áreas de estudo

Fonte: Produção do autor.

Área	Retângulo Envolvente	Tamanho (pixels)	Variação altimétrica (m)
1	2° 23' 48"S, 71° 51' 25"W a 2° 03' 48" S, 71° 31' 25"W	1200 x 1200	15 a 175
2	6° 53' 48"S, 70° 13' 55"W a 6° 33' 48" S, 69° 53' 55"W	1200 x 1200	91 a 237

Tabela 1 – Características das áreas de estudo.

Na Figura 2 está ilustrado o esquema com as etapas desenvolvidas no trabalho. Como dado de entrada foram utilizados o MDE SRTM, com resolução espacial de 1 arcosegundo (aproximadamente 30 metros), um recorte da rede de drenagem elaborada por Banon e Novo (2018), obtida automaticamente para toda a bacia Amazônica e o dado de área alagável produzido por Hess et al. (2003) a partir da classificação de imagens de radar, sendo esse dado chamado de *wetmask* neste trabalho.

Para a geração da classificação e dos mapas de incerteza associados, foi necessário inicialmente obter alguns atributos considerados importantes para caracterização do processo de inundação em estudo, que serão explicados a seguir. Essas variáveis foram fornecidas ao classificador *Random Forest*, um classificador baseado em múltiplas árvores de decisão, que gerou então a imagem classificada. A partir da informação da probabilidade associada a cada pixel pertencer a uma dada classe, foi possível elaborar os mapas de incerteza. A métrica utilizada para construção da medida de incerteza foi a

Entropia de Shannon, sendo possível então representar as incertezas do processo de classificação espacialmente distribuídas. Além disso, foi feita uma avaliação do resultado do mapeamento tomando como referência a *wetmask*, onde foi produzida uma matriz de confusão e determinada a acurácia do mapeamento. Todos esses procedimentos foram feitos para cada uma das duas áreas de estudo.



Figura 2- Fluxograma com as etapas de trabalho.

#### 2.2 Dados de entrada

#### 2.2.1 Wetmask

O dado de áreas alagáveis tomado como referência neste trabalho, chamado de *wetmask*, está exposto na Figura 3 referente às Áreas 1 e 2, onde é possível visualizar as áreas alagáveis sobre o MDE-SRTM. O dado foi fornecido no formato vetorial e para utilização neste estudo, se faz necessário que os planos de informações estejam representados em formato matricial. Dessa forma, foi realizada uma operação de rasterização, onde o mapa de referência de áreas alagáveis passou a ser representado por uma grade regular

retangular. Para a realização dessa operação o MDE-SRTM foi utilizado como a grade de referência.



Figura 3- Wetmask para as duas áreas de estudo.

#### 2.2.2 Rede de drenagem

A rede de drenagem utilizada neste trabalho foi obtida por Banon e Novo (2018), por meio do projeto de pesquisa intitulado "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".

Foi então utilizado um recorte da rede de drenagem de acordo com as áreas de interesse deste estudo. Para ser utilizado para obtenção de outros atributos, o arquivo até então em formato vetorial, precisou ser representado por meio de uma matriz. Nesse sentido, foi feita a operação de rasterização, como exemplificado na Figura 4. Tomando como referência o MDE-SRTM, foi criado uma grade regular retangular em que o dado que antes era representado na estrutura vetorial do tipo linha, foi convertido para o formato matricial. A rede de drenagem das duas áreas de estudo pode ser vista na Figura 5. Cada *pixel*, de acordo com o trecho do rio ao qual pertence, recebeu como atributo a ordem de Strahler.

Figura 4 - Processo de rasterização da rede de drenagem. a) Representação por linhas. b) Representação da rede de drenagem por *pixels*.



Fonte: Produção do autor.





#### 2.2.3 MDE

O Modelo Digital de Elevação pode ser representado por uma grade regular retangular, onde cada *pixel* recebe como atributo um único valor, referente a altimetria média de toda área pertencente ao *pixel*. Dessa forma, temos que o MDE representa quantitativamente a altimetria, uma grandeza que varia de forma contínua numa região. E então essa grandeza é discretizada no espaço, em intervalos representados pelo tamanho do *pixel*.

Para o desenvolvimento deste estudo, o MDE utilizado foi o resultante do projeto SRTM (*Shuttle Radar Topography Mission*) com resolução espacial de 1 arco-segundo (aproximadamente 30 metros). Esse produto foi resultado de uma missão realizada em conjunto pela *National Aeronautics and Space Administration* (NASA) e a *National* 

*Imagery and Mapping Agency* (NIMA), no ano de 2000. O objetivo dessa missão foi, por meio de interferometria SAR, produzir modelos digitais de elevação de cerca de 80% da superfície terrestre (FARR et al., 2007).

# 2.3 Pré-processamento do MDE

# 2.3.1 Local Drain Directions (LDD)

Uma das etapas iniciais de obtenção de atributos de interesse hidrológico a partir do MDE é a determinação das direções de fluxo. Uma maneira comum de representá-las é por meio de uma grade, conhecida por *Local Drain Directions* (LDD), onde o atributo armazenado em cada célula é uma codificação que identifica a direção do fluxo da célula correspondente no MDE (JARDIM, 2017).

Há diferentes métodos para determinar as direções de escoamento da água a partir de um MDE. 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, como apresentado na Figura 6.



Figura 6 - Oito possíveis direções de fluxo.

Fonte: Produção do autor.

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

O TerraHidro (ROSIM et al., 2003), software utilizado neste trabalho para obtenção das direções de fluxo e outros atributos, trata o problema de áreas planas por meio da abordagem de cavar o MDE. Já para a resolução da problemática dos mínimos locais, o

software utiliza o método PFS (*Priority First Search*) que realiza correções no MDE de forma a simular o comportamento físico da água. Dessa forma, por meio desses dois processos, o MDE se torna hidrologicamente consistente para aplicação do algoritmo de determinação das direções de fluxo. A Figura 7 apresenta a codificação utilizada pelo TerraHidro para grade de LDD.

32	64	128
16	0	1
8	4	2

Figura 7- Codificação das direções de fluxo do TerraHidro.

Fonte: Produção do autor.

#### 2.3.2 Área acumulada

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 um dado *pixel*. Nesse sentindo, é importante destacar que a obtenção do dado de área acumulada depende de uma correção de continuidade feita no MDE, a partir da operação de obtenção das direções de fluxo, e então são feitas operações que fazem o acúmulo de área de cada *pixel*. A Figura 8 exemplifica o cálculo de área acumulada representada em número de células. Neste trabalho o resultado do cálculo da área acumulada é uma grade, com as mesmas dimensões do MDE, onde cada *pixel* recebe como atributo o valor da área acumulada, em m<sup>2</sup>, referente aos *pixels* cujos caminhos de fluxo passam pelo *pixel* em questão.



Figura 8- Exemplo de grade de área de contribuição.

Fonte: JARDIM (2017).

#### 2.4 Atributos hidrológicos e morfométricos

Com o objetivo de caracterizar o fenômeno de inundação estudado, foram obtidos alguns atributos para serem então fornecidos ao classificador. É importante destacar que as relações espaciais entre os elementos desempenham papel fundamental na obtenção dos atributos. Ou seja, as informações determinadas a partir dos atributos levaram em consideração, de forma essencial, a localização espacial dos elementos em análise.

#### 2.4.1 Área acumulada expandida

Neste trabalho, foi utilizado o atributo chamado de área acumulada expandida. Esse atributo consiste em expandir a informação do rio relacionada a área a montante para todos os pontos conectados a ele, como ilustrado na Figura 9.





Fonte: Produção do autor.

O objetivo em se utilizar essa informação é poder identificar as áreas mais relacionadas com as calhas principais dos grandes rios, pois é esperado que essas áreas tenham mais propensão a inundação. Para melhor visualizar os valores, optou-se por trabalhar com o log da área acumulada expandida em m<sup>2</sup>.

#### 2.4.2 Declividade

Com o objetivo de caracterizar a área por meio de atributos morfométricos, foi calculada a declividade. A análise desse parâmetro permite determinar a distribuição das inclinações da superfície do terreno. Com relação ao fenômeno de inundação estudado, é esperado que áreas mais planas tenham mais predisposição a inundar. Sendo assim, incialmente, a declividade é então um importante parâmetro a se considerar. Neste trabalho a declividade foi representada por uma grade, de mesmas dimensões do MDE, onde cada *pixel* recebe como atributo o valor de declividade calculada em porcentagem. Mais informações sobre como é calculado o valor da declividade no TerraHidro podem ser obtidas em (SHARY, SHARAYA, MITUSOV, 2001).

#### 2.4.3 HAND

O HAND é um algoritmo, desenvolvido por Rennó et al. (2008), 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 a rede de drenagem mais próxima, seguindo o caminho de fluxo. Essa ferramenta permite indicar a profundidade do lençol freático, sendo assim um bom atributo para identificar áreas mais úmidas.

A Figura 10 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 de LDD) identifica o *pixel* pertencente a rede de drenagem mais próxima e então obtém o valor de elevação desse *pixel* pertencente a 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 que os *pixels* pertencentes a rede de drenagem recebem como atributo o valor zero.



Figura 10- Processos para obtenção do HAND

Fonte: Rennó et al. (2008).

Neste trabalho, foram obtidas diferentes grades do descritor HAND, de acordo com a menor ordem da rede de drenagem considerada. Dessa forma, foi calculado para cada área de estudo o HAND considerando os diferentes planos de informação da rede de drenagem (de acordo com a menor ordem considerada). O HAND calculado utilizando a rede de drenagem com ordem maior igual a um, foi denominado HAND1 e assim sucessivamente, até o HAND 5, que considera apenas os trechos de rios com ordem igual ou superior a cinco.

## 2.4.4 Distâncias relativas à rede de drenagem: euclidiana e caminho de fluxo

Outro atributo considerado para este estudo foi a distância de cada *pixel* à rede de drenagem. Os métodos utilizados para o cálculo da distância relativa foram, o da distância euclidiana e o da distância ao longo dos caminhos de fluxo.

A distância euclidiana entre um ponto na superfície e a rede de drenagem é tida como a menor distância em linha reta entre o *pixel* analisado e o *pixel* pertencente a rede de drenagem. Já no cálculo da distância ao longo dos caminhos de fluxo, o sentido do fluxo é levando em consideração. Dessa forma, a distância é obtida medindo-se a menor distância entre o *pixel* da superfície e o *pixel* pertencente a rede de drenagem mais próxima, seguindo a grade de LDD.

Na Figura 11 estão representados os métodos de cálculo da distância utilizados. O resultado das duas formas de cálculo da distância é uma grade regular onde cada *pixel* recebe como atributo a distância relativa à rede de drenagem mais próxima.

Figura 11- (a) Distância euclidiana à rede de drenagem; (b) Distância considerando o



caminho de fluxo

Fonte: Sousa et al. (2015).

Neste trabalho a distância euclidiana foi calculada utilizando o *software* QGIS, enquanto que para a determinação da distância considerando os caminhos de fluxo foi utilizado uma rotina de cálculo. Assim como para determinação do HAND, foram consideradas as distâncias relativas à rede de drenagem de acordo com a menor ordem dos rios considerada. Sendo assim, foram gerados diferentes planos de informação de distância relativa considerando inicialmente os trechos de rio de ordem maior igual a um, depois os trechos de rio de ordem maior igual a dois, até a drenagem menos ramificada considerada, que foi aquela com os rios de ordem maior igual a cinco.

#### 2.5 Amostragem

Após obtenção dos atributos, o passo seguinte foi realizar o processo de amostragem para fornecer ao classificador. Buscando gerar um bom conjunto amostral, representativo considerando todos os atributos analisados, foi obtida uma imagem onde cada banda representa um atributo e então foi feito um processo de classificação não supervisionada dessa imagem utilizando o método k*-means*, onde os *pixels* pertencentes à mesma classe apresentam padrões que o algoritmo considera semelhantes de acordo com todos os atributos escolhidos.

Como os dados possuem diferentes unidades, o passo inicial para aplicação do algoritmo k-means foi a normalização das variáveis. Esse procedimento foi feito com o objetivo de que todos os parâmetros fossem igualmente considerados no processo de classificação. A normalização foi feita utilizando um *script* em *Python*, onde os valores de todas as variáveis ficaram entre 0 e 1. Importante salientar que os atributos normalizados foram utilizados apenas para a aplicação do k-*means*. Todos os processos subsequentes se deram considerando as variáveis sem modificação em seus valores.

O k-means é um método que agrupa os dados em K grupos ou *clusters*. Inicialmente os *pixels* são agrupados aleatoriamente de acordo com o número de *clusters* prédeterminado. O centroide de cada *cluster* receberá o valor da média de todos os valores dentro do *cluster*, então cada *pixel* é avaliado com relação a distância até o centroide dos K grupos. As classes são redefinidas de acordo com o critério da menor distância até o centroide, ou seja, um *pixel* pertencerá a classe cujo centroide está mais próximo. Posteriormente são feitas novas iterações e a cada iteração as médias dos centroides são recalculadas e os *pixels* vão sendo agrupados nas classes consideradas mais próximas, de acordo com o critério de menor distância.

Para a aplicação do algoritmo k-*means* é necessário informar alguns parâmetros, sendo eles: o número de classes, limiar de mudança e o número máximo de iterações. O número de classes foi escolhido de forma arbitrária, sendo 10 um valor entendido como suficiente para formar *clusters* que permitissem gerar amostras bem distribuídas ao longo da imagem. Como essa classificação foi realizada apenas para gerar as amostras e não para a utilização do resultado da classificação em si, não houve um refinamento na escolha dos demais parâmetros, sendo escolhido o valor de 5% para o limiar de mudança e 1 para o número máximo de iterações, sendo esses os valores *default* do *software* ENVI utilizado.

Como o objetivo era tornar a amostra mais representativa possível, foram feitos dois processos de classificação. Foi feita uma classificação só para a área alagável, e outra apenas para a área de terra firme, ambas as áreas foram consideradas de acordo com a *wetmask*, como ilustrado na Figura 12. Dessa forma, foram geradas 10 classes dentro da região de área alagada e outras 10 classes na região de terra firme. As amostras aleatórias foram então distribuídas igualmente dentro de cada uma das classes. Optou-se por trabalhar com 10000 amostras em cada região (área inundável e terra firme). Dessa forma, foram amostrados por volta de 20000 pontos no total para cada uma das áreas de estudo.



Figura 12- Esquematização do método K-means

Fonte: Produção do autor.

#### 2.6 Random Forest

Neste estudo, para a obtenção do mapa de áreas alagáveis, foi utilizado o classificador *Random Forest*. O *Random Forest* (RF) é um classificador baseado em um conjunto de árvores de decisão. Por utilizar uma grande quantidade de árvores de decisão, a acurácia do método é aumentada. Cada árvore é criada no RF utilizando um algoritmo determinístico que seleciona, de forma aleatória, as variáveis presentes e um grupo de amostras, resultante de um processo de reamostragem aplicada sobre pontos amostrais originais (RAMOELO et al., 2015).

Depois que é gerado a quantidade definida de árvores de decisão, o *pixel* em análise é submetido à classificação de acordo com cada uma das árvores. A classe final escolhida para o *pixel* é aquela que apresentar maior frequência dentre todas as árvores. Ou seja, analisa-se a quantidade de votos de cada classe e a classe final é aquela que recebeu a maior quantidade de votos.

Como no *Random Forest*, a classe escolhida para um determinado *pixel* é aquela mais frequente (moda), considerando todas as árvores de decisão, cada *pixel* terá associado a ele uma probabilidade de pertencer a uma determinada classe. Sendo assim, foi obtido para cada uma das áreas de estudo, um mapa que representa a proporção de vezes que cada *pixel* foi classificado em uma determinada classe considerando o número total de árvores. Neste estudo optou-se por trabalhar com 50 árvores de decisão.

Com base na informação da probabilidade associada a cada *pixel* pertencer a classe escolhida, é possível avaliar as incertezas espacialmente distribuídas resultantes do mapeamento, avaliando assim a classificação. Importante notar que a avaliação das incertezas diz respeito a avaliação do processo de classificação e do classificador utilizado, não tendo necessariamente relação com a realidade ou o mapa de referência.

### 2.7 Entropia de Shannon

Neste trabalho a métrica utilizada para construir uma medida de incerteza espacialmente distribuída foi a Entropia de Shannon. O conceito de entropia provém da segunda lei da termodinâmica da física e pode ser entendida como uma medida da dispersão da energia. Nesse sentindo, a entropia é então relacionada com a medida da desordem do sistema (CALVALCANTI et al., 2018). Em 1948 o engenheiro e matemático Claude Elwood

Shannon (1916-2001) publicou o trabalho intitulado "*A Mathematical Theory of Communication*" que representou os passos iniciais da Teoria da Informação (PAVIOTTI, 2019). Em seu trabalho, Shannon estende o conceito de entropia e a define como uma medida de informação, escolha e incerteza (SHANNON, 1948).

A Entropia de Shannon permeia diversas áreas do conhecimento, dentre elas o processamento digital de imagens e a classificação de imagens. O processo de seleção de uma classe entre as várias existentes tem atrelado a ele a noção de incerteza. Nesse sentido, a entropia de Shannon pode ser utilizada para quantificar a incerteza associada ao processo de classificação.

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

$$H = -\sum_{i=1}^{N} b_i \times \log 2(b_i) \tag{2.1}$$

Onde:  $b_i$  representa a probabilidade associada a cada classe i; e N o número total de classes.

Analisando a expressão, tem-se que a entropia irá assumir valor máximo quando as probabilidades de ocorrência das classes forem iguais. Dessa forma, quanto maior o valor da entropia de Shannon, maior a incerteza atrelada a classificação, indicando a existência de uma maior confusão entre as classes consideradas, ou seja, indicando que houve muita confusão, por parte do classificador, para determinar a qual classe um dado pixel pertence.

#### **3 RESULTADOS E DISCUSSÃO**

#### 3.1 Obtenção de parâmetros

A primeira etapa deste trabalho consistiu na obtenção de atributos considerados importantes para caracterizar o fenômeno de inundação estudado. Primeiramente foi obtida a grade de declividade para as duas áreas de estudo, como mostradas na Figura 13.





Analisando os mapas de declividade nota-se que a variabilidade espacial da informação é muito alta, não sendo possível dessa forma ter uma visão mais geral da variável. Ou seja, não é possível identificar regiões mais homogêneas em relação a declividade, pois a informação do parâmetro é muito particionada e específica para cada *pixel*.

O passo seguinte foi a obtenção de atributos hidrológicos tomando como base o MDE. Inicialmente foi obtida a grade LDD, que representa as direções de fluxo, ou seja, qual o caminho que a água percorre em cada pixel. Utilizando a matriz de LDD, o passo posterior foi obter as áreas acumuladas e em seguida determinar o atributo utilizado que foi chamado de "expansão da área acumulada". Esse atributo representa a expansão da informação de área acumulada do pixel pertencente a rede de drenagem para todos os pixels conectados a ele. Para melhor visualização dos valores, optou-se por trabalhar com o log da área acumulada em m<sup>2</sup>, como pode ser visto na Figura 14.



Figura 14 - Log da área acumulada expandida

Por meio desse atributo pode-se agregar ao classificador uma variável que representa para cada *pixel*, qual a dimensão do rio, em termos de capacidade, no qual ele está conectado. Assim, o classificador pode identificar quais as áreas estão relacionadas com trechos de grandes rios, sendo essas talvez mais propensas a serem classificadas como áreas alagadas, dependendo da análise dos outros atributos.

Com base no MDE, na rede de drenagem e na matriz LDD foi determinado o parâmetro HAND, que representa a diferença de elevação entre um *pixel* analisado e o ponto pertencente a rede de drenagem conectado a ele. Esse atributo foi obtido utilizando diferentes grades de rede de drenagem, de acordo com o valor de ordem mínima considerada. Dessa forma, foi calculado o HAND considerando os rios de ordem maior ou igual a um, até a drenagem com rios de ordem maior ou igual a cinco. A ideia é que os diferentes HAND tenham pesos diferentes na caracterização do fenômeno em estudo. Dessa forma, foram fornecidos para o classificador as diferentes matrizes de HAND e é esperado que ele determine, a partir do processo de construção das árvores de decisão, qual ou quais são mais representativas para a classificação. Na Figura 15 estão ilustrados os diferentes HAND, nomeados de acordo com a menor ordem do rio considerada, para a Área 1. Da mesma forma na Figura 16, para a Área 2.



Figura 15 - HAND de acordo com a menor ordem de rio considerada para a Área 1

Figura 16 - HAND de acordo com a menor ordem de rio considerada para a Área 2



O outro atributo fornecido ao classificador foi a distância de cada pixel à rede de drenagem. Esse parâmetro foi determinado utilizando dois métodos de cálculo de distância. Um foi a distância euclidiana até a rede de drenagem e o outro foi a distância de cada pixel à rede de drenagem considerando o caminho de fluxo percorrido. Da mesma forma que no cálculo do HAND, foram determinadas as distâncias considerando rios com diferentes ordens. Então calculou-se a distância relativa considerando a rede de drenagem formada inicialmente pelos rios de primeira ordem até a ordem máxima representada, até a rede de drenagem considerando apenas os rios de ordem maior igual a cinco. Na Figura 17 está representada a distância euclidiana para a Área 1, de acordo com a menor ordem de rio considerada. Da mesma forma, na Figura 18 está representado esse atributo para a Área 2.



Figura 17 - Distância euclidiana até a rede de drenagem para a Área 1.



Figura 18 - Distância euclidiana até a rede de drenagem para a Área 2.

Utilizando uma rotina de cálculo foi determinada a distância dos pixels a rede de drenagem considerando o caminho de fluxo para as áreas de estudo. Foram obtidos diferentes resultados de acordo com a menor ordem do rio considerada. Na Figura 19 estão expostos os mapas de distância obtidos para a Área 1, já na Figura 20 estão ilustrados os mapas para a Área 2.



Figura 19- Distância a rede de drenagem seguindo o caminho de fluxo para a Área 1.

Figura 20 - Distância a rede drenagem seguindo o caminho de fluxo para a Área 2.



Quando comparado os mapas de distância, pode-se perceber que o método de distâncias ao longo dos caminhos de fluxo gerou intervalos maiores de valores. Além de que a forma como os valores estão distribuídos na região também diferiu nos métodos. De fato, tem-se que a distância ao longo do caminho de fluxo pode ser maior ou igual, mas não menor que a distância calculada pelo método da distância euclidiana, dada a própria definição de distância euclidiana.

A ideia em se considerar esse atributo é que áreas que estejam mais próximas da rede de drenagem sejam mais influenciadas pelas mudanças no nível do rio e estejam mais propensas a serem áreas alagadas. Foram então fornecidos ao classificador todos os mapas de distância e a partir da elaboração das árvores de decisão, cada mapa terá um peso diferente, conferindo assim maior ou menor importância da informação para a caracterização das áreas alagadas.

# 3.2 Classificação pelo algoritmo Random Forest

Após obtenção de todos os atributos de interesse, foram obtidas as 10000 amostras para a área de terra firme e outras 10000 amostras para as regiões de área inundável, de acordo com o dado de referência, a *wetmask*.

Com base nos atributos e nas amostras obtidas foi utilizado o classificador *Random Forest* por meio de um programa desenvolvido na linguagem IDL e processado no *software* ENVI. Foi obtido então o mapa de áreas alagáveis para as duas regiões de estudo, como exposto na Figura 21.



Figura 21 - Mapeamento das áreas alagáveis pelo classificador Random Forest para as

duas áreas de estudo.

Analisando os mapas podemos perceber que ao comparar com a *wetmask*, houve uma superestimação das regiões de áreas alagáveis, para ambas as áreas de estudo. Isso pode ter sido caudado por diversos fatores, dentre eles a falta de atributos que descrevam melhor o fenômeno estudado. Além disso, não foram testados outros parâmetros para a geração das árvores de decisão, o que pode levar a melhores resultados.

Outra questão importante que explica o resultado obtido são as amostras utilizadas. Ao analisar os pontos amostrados percebe-se em alguns locais pontos que foram amostrados como exemplo de uma determinada classe e evidentemente não pertencem a essa classe. A Figura 22 exemplifica amostras que foram classificadas como terra firme, entretanto analisando o MDE (imagem na qual as amostras estão sobrepostas) percebe-se que essas regiões são áreas muito prováveis de serem alagáveis. Isso ocorre porque o mapa utilizado como referência para obtenção das amostras (*wetmask*) é um bom dado, porém apresenta inconsistências em algumas regiões. No caso, para a obtenção da *wetmask*, não foi considerado o relevo, além de que esse dado foi obtido por meio de um sensor diferente e possui uma resolução diferente da resolução dos atributos considerados neste estudo. Dessa forma o dado tomado como referência está sujeito a diferentes erros.



Figura 22 - Exemplo de inconsistência nos pontos amostrados.

Buscando avaliar o resultado da classificação em relação ao mapa de referência utilizado, foi feita uma matriz de confusão para cada um dos mapas de classificação obtidos. A matriz de confusão foi elaborada a partir de um conjunto de aproximadamente 1000 amostras distribuídas aleatoriamente em cada uma das duas classes. E então para cada ponto amostrado, comparou-se a o resultado do mapa classificado com a *wetmask*, sendo obtidos os resultados expostos na Tabela 2, referente à Área 1 e na Tabela 3, referente à Área 2.

	Wetmask				
RF	Classe	Área alagável	Terra firme	Total	Exatidão Usuário (%)
ção	Não classificado	2	3	5	-
fica	Área alagável	833	124	957	87.04
issi	Terra firme	129	840	969	86.69
Cl	Total	964	967	1931	
	Exatidão do produtor (%)	86.41	86.87		

Tabela 2 - Matriz de confusão referente a classificação da Área 1.

Tabela 3 - Matriz de confusão referente a classificação da Are	ea 2.
--	-------

	Wetmask				
) RF	Classe	Área alagável	Terra firme	Total	Exatidão Usuário (%)
ação	Não classificado	0	2	2	-
ific	Área alagável	964	151	1115	86.46
ass	Terra firme	49	863	912	94.63
CI	Total	1013	1016	2029	
	Exatidão do produtor (%)	95.16	84.94		

Analisando os acertos da classificação em relação ao mapa de referência (*wetmask*), temse que a classificação da Área 1 apresentou uma exatidão global (ou acurácia) 86,64%, enquanto que a acurácia do mapeamento da Área 2 apresentou valor de 90,04%.

O desempenho da classificação para cada classe pode ser determinado por meio da exatidão do produtor e da exatidão do usuário. Temos que para a Área 1, a exatidão do produtor para a classe de área alagável foi de 86,41% enquanto que para a classe de terra firme foi de 86,87%. Já para a Área 2, a exatidão do produtor para a classe de área alagável e de terra firme foi de 95,16% e 84,94%, respectivamente. Essas porcentagens

representam o quanto dos pontos amostrados foram classificados corretamente de acordo com a referência.

Já a relação de número de acertos da classificação em relação ao número de pontos classificados em uma dada classe é dada pela exatidão do usuário. A classificação para a Área 1 apresentou uma exatidão do usuário de 87,04 % para a classe de área alagável e 86,69% para a classe de terra firme. Enquanto que essa exatidão para a Área 2 foi de 86,46 % para área alagável e 94,63% para terra firme.

## 3.3 Obtenção dos mapas de incerteza

O programa desenvolvido na linguagem IDL responsável por gerar as imagens classificadas utilizando o método *Random Forest*, permite a obtenção de imagens que representam a proporção de vezes que um pixel foi classificado em cada uma das classes de acordo com o número de árvores de decisão considerado.

Na Figura 23 estão apresentadas essas imagens representando a proporção de um pixel pertencer a uma dada classe de acordo com as 50 árvores de decisão utilizadas. Na imagem temos que quanto mais claro, maior a proporção com que o pixel foi classificado em uma dada classe dentre todas as árvores de decisão.

Com base nas imagens da proporção referente a classificação de cada pixel é possível então determinar a Entropia de Shannon, obtendo assim a quantificação e espacialização da incerteza associada ao processo de classificação. É importante notar que a incerteza está associada ao classificador e aos métodos utilizados na classificação, não tendo relação com o mapa de referência, a *wetmask*.





Proporção áreas alagáveis A2

Proporção terra firme A2

Para melhor visualização e comparação, na Figura 24 é exposto o mapa de incerteza para a Área 1 ao lado do resultado da classificação para a Área 1. Já na Figura 25 estão mostradas as mesmas informações para a Área 2.



## Figura 24 - Mapa de incerteza por entropia de Shannon para a Área 1.





Com base nas Figuras 24 e 25, tem-se que áreas com altos valores de entropia (regiões mais brancas nos mapas de incerteza) representam áreas onde existiu uma maior confusão por parte do classificador em determinar a classe para o pixel em questão, ou seja, houve uma maior confusão entre as classes consideradas. Dessa forma, percebemos nos mapas que houveram áreas que foram classificadas como áreas alagáveis, porém com muita incerteza associada. Esse fato está relacionado com a superestimação das áreas alagáveis, mostrando que houve muita incerteza no mapeamento de algumas dessas regiões.

Analisando comparativamente os mapas de incerteza das duas áreas temos que a Área 2 apresentou visualmente menos regiões de incerteza elevada, indicando que houve uma menor confusão na classificação dessa área quando comparada com a classificação da Área 1. Considerando o mapa de incerteza, é possível alterar o mapa de classes de acordo com um determinado valor de entropia. Isso pode ser feito para obter um resultado de classificação com maior certeza associada. Diante disso, foram gerados novos mapas de áreas alagáveis considerando um valor mínimo de entropia. Como nos mapas analisados a entropia variou de 0 a 1, escolheu-se arbitrariamente um limiar de 0,4. Dessa forma, foram gerados mapas de classificação onde os pixels com valores abaixo do limiar de entropia 0,4 receberam como atributo o valor zero, ou seja, foram associados à classe "terra firme". Nesse sentindo, foi possível obter um produto da classificação com uma maior certeza associada. Dessa forma, essa operação foi feita para exemplificar a importância de relacionar a classificação com as incertezas espacialmente distribuídas.









## 4 CONCLUSÕES

Neste trabalho foi realizada uma classificação de áreas alagáveis e obtenção de mapas de incerteza associados ao processo de classificação. Para a obtenção do mapa de classes foi utilizado o classificador *Random Forest* com base em dados oriundos do MDE-SRTM. O mapeamento resultou em uma superestimação de regiões inundáveis para ambas as regiões de estudo. Quando analisados os mapas de incerteza, a Área 1 apresentou visualmente uma maior quantidade de áreas com elevada incerteza associada no processo de classificação, mostrando que houve maior confusão na distinção das classes por parte do classificador.

Os resultados mostram um bom potencial do método. Sendo necessário para obter melhores classificações a realização de uma mineração mais aprimorada dos dados para construção das árvores de decisão, uma busca por mais atributos que expliquem melhor o fenômeno estudado e um refinamento das amostras utilizadas, tendo em vista que a referência utilizada (*wetmask*) para amostragem está sujeita a erros, pois foi gerada a partir de outro sensor, tendo uma resolução diferente do MDE. Espera-se que após refinamento do processo de classificação e melhora dos resultados, será possível descartar ou acrescentar áreas na *wetmask* que sejam classificadas com certeza elevada, gerando assim um dado melhorado de áreas alagáveis para essas regiões da bacia Amazônica.

A associação de uma medida de incerteza ao resultado da classificação é uma ferramenta poderosa que torna o mapa mais significativo, pois se torna possível a visualização de quais regiões apresentaram maior ou menor confusão na separação entre as classes. Podendo garantir assim, uma maior confiabilidade do mapeamento, por permitir visualizar as áreas com maior certeza associada. Além disso, a manipulação do resultado partir de um valor de entropia surge como uma opção quando se busca entregar um resultado destacando apenas as áreas que foram classificadas com uma certeza maior. Dessa forma, a construção de uma medida de incerteza espacialmente distribuída agrega mais significado ao mapa temático gerado a partir de dados orbitais.

# **5 REFERÊNCIAS BIBLIOGRÁFICAS**

ALFAYA, F. A. V. S. Mapeamento de áreas alagáveis da calha Solimões/Amazonas utilizando análise de imagens baseada em objeto com dados MDE-SRTM. 2012. 59

p. (sid.inpe.br/mtc-m19/2012/04.10.11.22-TDI). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012.

ALSDORF, D.; LETTENMAIER, D.; VOROSMARTY, C.; NASA Surface Water Working Group. The need for global, satellite-based observations of terrestrial surface waters. **Eos, Transactions American Geophysical Union** v. 84, n. 29, p. 269–280, Jul 2003.

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: <a href="http://www.dsr.inpe.br/amazondrainage">http://www.dsr.inpe.br/amazondrainage</a>.

BARBOSA, C. C. F. Sensoriamento remoto da dinâmica da circulação da água do sistema planície de Curuai/Rio Amazonas. 2007. 282 p. (INPE-14614-TDI/1193). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2007.

BRAGA, C. B.; SANT'ANA, S. J. S.; FREITAS, C. C. Distâncias estocásticas e mapas de incerteza aplicados à classificação de dados de múltiplos sensores. **Revista Brasileira de Cartografia**, v. 67, n. 7, p. 1391-1411, Nov/Dez 2015.

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, Abr/Jun 2009.

CAVALCANTI, H. L. B.; FERREIRA, E. A.; ABRANTES, P. G.; CAVALCANTI, G. N.; As Muitas Interpretações da Entropia e a Criação de Um Material Didático Para o Ensino da Interpretação Probabilística da Entropia. **Química Nova**, v. 40, n. 3, p. 169-177, Ago 2018.

FARR, T.G.; ROSEN, P.A.; CARO, E.; CRIPPEN, R.; DUREN, R.; HENSLEY, S.; KOBRICK, M.; PALLER, M.; RODRIGUEZ, E.; Roth, L.; *et al.* The shuttle radar topography mission. **Rev. Geophys.** 2007, *45*, 361–364.

FREITAS, R. M.; NOVO, E. M. L. M. Mapeamento de áreas alagáveis na planície do rio Amazonas a partir de dados do sensor MODIS – região do Lago Curuaí – PA. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia. Anais... São José dos Campos: INPE, 2005. p. 533-540. Disponível em: <http://urlib.net/rep/ltid.inpe.br/sbsr/2005/02.12.16.31>.

HESS, L. L.; MELACK, J. M.; NOVO, E. M. L. M.; BARBOSA, C. C. F.; GASTIL,M. Dualseason mapping of wetland inundation and vegetation for the central Amazon basin. Remote Sensing of Environment, v. 87, 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.
109 p. (sid.inpe.br/mtc-m21b/2017/05.17.13.26-TDI). Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2017.

J.O.O. structure JENSON. S.K.K.: DOMINGUE, Extracting topographic 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. Áreas inundáveis - Um desafio para Limnologia. Acta Amazonica. p. 775-795, 1980.

JUNK, W. J. Flood tolerance and tree distribution in central Amazonian floodplains. In: Holm-Nielsen, L. B.; Nielsen, I. C.; Balslev, H. (eds). Tropical forest: botanical dynamics, speciation and diversity. New York: Academic **Press**, p.47-64, 1989.

PAVIOTTI, J. R. **Considerações sobre o conceito de entropia na Teoria da Informação**. 2019. 118 p. Dissertação (Mestrado em Tecnologia) – Universidade Estadual de Campinas, Limeira, 2019.

32

PIEDADE, M. T. F.; SCHONGART, J.; WITTMANN, F. PAROLIN, P.; JUNK, W. J. Impactos ecológicos da inundação e seca na vegetação das áreas alagáveis amazônicas. In: BORMA, S.; NOBRE, C. (Org.). **Secas na Amazônia: causas e consequências**. São Paulo: Oficina de Textos, 2013. cap. 16, p. 268-304.

SOARES. RENNÓ. C.D.: NOBRE, A.D.: CUARTAS, L.A.; J.V.: HODNETT, M.G.; TOMASELLA, J.; WATERLOO, M.J. HAND, a new descriptor Mapping terrain using SRTM-DEM: terra-firme rainforest environments in Amazonia. Remote Sensing of Environment, v. 112, n. 9, p.3469–3481, 2008.

RENNÓ, C. D.; NOVO, E. M. L. M.; BANON, L. C. Correção geométrica da máscara de áreas alagáveis da bacia amazônica. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16. (SBSR), 2013, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2013. p. 5507-5514. Disponível em: <a href="http://urlib.net/rep/3ERPFQRTRW34M/3E7GJHH">http://urlib.net/rep/3ERPFQRTRW34M/3E7GJHH</a>.

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. (SBSR)., 2003, Belo Horizonte. **Anais...** São José dos Campos: INPE, 2003. p. 2589 - 2596. Disponível em: <http://urlib.net/rep/ltid.inpe.br/sbsr/2002/11.17.20.15>.

SOUSA, T. M. I.; SARAIVA, A. G. S.; PAZ, A. R. Distâncias relativas à rede de drenagem: euclidiana x caminho de fluxo. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17. (SBSR), 2015, João Pessoa, Anais... São José dos Campos: INPE, 2015. p. 6227-6234.