Revista GEOCIÊNCIAS

unesp

ISSN: 1980-900X (online)

INTEGRAÇÃO DA INCERTEZA NA AMOSTRAGEM E CLASSIFICAÇÃO *RANDOM FOREST* UTILIZANDO BANDAS E ÍNDICES ESPECTRAIS PARA O MAPEAMENTO DE INUNDAÇÃO

INTEGRATION OF UNCERTAINTY IN SAMPLING AND RANDOM FOREST CLASSIFICATION USING BANDS AND SPECTRAL INDICES FOR FLOOD MAPPING

Thiago BAZZAN[1], Camilo Daleles RENNÓ[1], Deborah Lopes Correia LIMA[1], Elisabete Weber RECKZIEGEL[2]

[1]Instituto Nacional de Pesquisas Espaciais (INPE). Avenida dos. Astronautas, 1758, São José dos Campos – SP. E-mails: tbazzan@gmail.com; camilo.renno@inpe.br; deborahlclima@gmail.com

[2]Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN). Estrada Dr. Altino Bondensan, 500 - Eugênio de Melo, São José dos Campos – SP. E-mail: elisabete.reckziegel@cemaden.gov.br

RESUMO - Classificações tradicionais apresentam limitações para o mapeamento de inundações devido à mistura da resposta espectral da água com alvos adjacentes não aquáticos ou resposta espectral similar de alvos não aquáticos com a água. Além disso, em geral, as classificações são avaliadas apenas em termos de acurácia global sem considerar as incertezas no processo de classificação. neste estudo objetivou-se integrar a incerteza na classificação Random Forest (RF) para o mapeamento de inundações auxiliando o processo de amostragem. A classificação utilizou 21 variáveis representadas por bandas e índices espectrais do sensor Operational Land Imager do satélite Landsat-8. A amostragem foi realizada inicialmente com a seleção de pontos a partir da interpretação visual da imagem de satélite e posteriormente coletando amostras com alta entropia de Shannon no mapa de incerteza. As variáveis com maior importância para a classificação foram selecionadas utilizando o algoritmo Recursive Feature Elimination (RFE). Os resultados mostram que a classificação RF final usando amostras coletadas com base no mapa de incerteza e o conjunto de variáveis selecionadas pelo RFE apresentou 98,0% de exatidão e redução das incertezas do mapeamento da água superficial em relação à classificação RF com todas as variáveis e sem considerar a amostragem baseada na incerteza.
Palavras-chave: Mapeamento de inundação. Classificador Random Forest. Bandas e índices espectrais. Seleção de variáveis. Entropia de Shannon.

ABSTRACT - Traditional classifications present limitations for mapping floods due to mixing the spectral response of water with adjacent non-aquatic targets or similar spectral response of non-aquatic targets with water. Furthermore, in general, these classifications are evaluated only in terms of overall accuracy without considering the uncertainties in the classification process. Thus, this study aimed to integrate uncertainty in the Random Forest (RF) classification process for flood mapping, which guided the sampling process. The classification used 21 variables including indices and spectral bands from the Operational Land Imager sensor of the Landsat-8 satellite. Sampling was performed initially with the selection of points from the visual interpretation of the satellite image and later by collecting samples with high Shannon entropy values in the uncertainty map. The variables with the greatest importance for classification were selected by the Recursive Feature Elimination (RFE) algorithm. The final RF classification using samples collected based on the uncertainty map and with the four selected variables by the RFE presented an accuracy of 98.0% and a reduction of uncertainty, which indicates a greater confidence in the spatial representation and quantification of water permanent and temporary surface associated with floods.
Keywords: Flood mapping. Random Forest Classifier. Spectral bands and indices. Variable selection. Shannon Entropy.

INTRODUÇÃO

As inundações são processos hidrológicos importantes para a formação de planícies fluviais devido à deposição e acúmulo de sedimentos presentes na água extravasada do canal para as áreas adjacentes (Goudie, 2004), na interação e conectividade eco-hidrológica da biota entre o

rio e a planície de inundação (Neiff, 1999), na deposição de nutrientes para a fertilização natural da planície de inundação (Ogden et al., 2007), na participação no ciclo biogeoquímico com recebimento, transporte, acúmulo, deposição nos oceanos e retorno do carbono para a atmosfera (Aufdenkampe et al., 2011), nas relações com as mudanças climáticas (Hirabayashi et al., 2013) e para o desencadeamento de desastres naturais (Rahman & Di, 2017).

Neste contexto, o sensoriamento remoto fornece dados observacionais espacialmente distribuídos e temporalmente frequentes da superfície da Terra que podem ser utilizados no mapeamento das inundações e monitoramento da dinâmica das águas superficiais (Huang et al., 2018). As imagens multiespectrais de satélites são amplamente utilizadas para a detecção da água e mapeamento de inundações (Donchyts et al., 2016; Pekel et al., 2016; Devries et al., 2017), sendo a extensão e a quantificação da inundação um dado importante para a calibração e validação de modelos hidrológicos (Khan et al., 2011) e modelos hidrodinâmicos (Giustarini et al., 2015), elaboração de cartografia de risco para prevenção, planejamento, monitoramento, alerta e resposta face a ocorrência de desastres hidrológicos associados à inundações (Campos et al., 2015).

Entre os métodos aplicados para o mapeamento das águas superficiais a partir de imagens multiespectrais estão a limiarização de índices espectrais de água (Jiang et al., 2014; Xie et al., 2016; Zhou et al., 2017; Brubacher et al., 2017; Wang et al., 2018a; Acharya et al., 2018a; Martins et al., 2019), classificações supervisionadas e não supervisionadas baseadas nos dados originais (Joyce et al., 2009; Alatorre et al., 2011; Franci et al., 2016; Moura et al., 2022). No entanto, devido à mistura da resposta espectral da água com alvos adjacentes não aquáticos ou da resposta espectral similar de alvos não aquáticos (sombras de vegetação, relevo ou de estruturas urbanas) com a água, ocorrem limitações para a diferenciação dos alvos implicando em erros de omissão e comissão que reduzem a acurácia global do mapeamento (Namikawa et al., 2016).

A limiarização do histograma da imagem dos índices espectrais é um método de processamento rápido para o mapeamento da água superficial. Os índices espectrais realçam, melhoram a distinção entre alvos e facilitam a interpretação e a comparação dos dados, tanto qualitativos quanto quantitativos (Young et al., 2017).

No entanto, os índices espectrais têm como principal limitação a determinação de um valor de limiar ideal para subdividir de forma efetiva alvos aquáticos e não aquáticos para o mapeamento da água superficial, principalmente em áreas inundadas (Jones, 2015; Silveira & Guasselli, 2019; Totaro et al., 2019; Kordelas et al., 2019). Em relação às classificações supervisionadas baseadas em estatística, estas podem ser complexas justamente devido à esta mistura ou similaridade espectral dos alvos não aquáticos com a água, necessitando a criação de muitas subclasses de mapeamento (Koko et al., 2021).

O algoritmo de aprendizado de máquina (*machine learning*) para classificação baseado em árvores de decisão *Random Forest* (Breiman, 2001) tem apresentado bons resultados e alta acurácia global ao combinar e integrar diferentes tipos de variáveis para o mapeamento da água superficial permanente (Ko et al., 2015) ou temporária associada a inundações, onde há maior mistura da resposta espectral dos alvos (Feng et al., 2015, Tulbure et al., 2016). O *Random Forest*, em relação a outros algoritmos de aprendizado de máquina, é simples de parametrizar, informa a contribuição relativa de cada variável para a classificação e fornece a exatidão do modelo (Tyralis et al., 2019).

A avaliação dos resultados das classificações é importante, não só em termos de acurácia global, mas também quanto às incertezas inerentes ao processo de classificação. Isso pode auxiliar na identificação de regiões pouco representadas (através das amostras) ou indicar a necessidade de novos atributos capazes de identificar essas feições incertas (Lima & Rennó, 2021).

A entropia de Shannon (Shannon, 1948) é uma métrica que possibilita avaliar a distribuição espacial das incertezas oriundas da classificação. No entanto, poucos estudos abordam os métodos de amostragem, a seleção de variáveis preditoras e a avaliação da distribuição espacial das incertezas associadas ao classificador *Random Forest* para melhorar o desempenho da classificação e a confiança no mapeamento da água superficial.

Neste contexto, o objetivo deste estudo foi integrar a incerteza no processo de amostragem e classificação *Random Forest* para o mapeamento da água permanente e temporária associada a inundações considerando: (i) as bandas e índices espectrais como variáveis preditoras; (ii) amostragem inicial a partir da interpretação da imagem de satélite; (iii) a seleção de variáveis

com maior importância para a classificação e; (iv) a entropia de Shannon como fonte para amostragem e métrica de avaliação de incertezas.

Área de estudo

A área de estudo (Figura 1) corresponde a um trecho densamente urbanizado, com presença de diferentes tipologias de uso do solo e cobertura vegetal na planície do Rio dos Sinos, localizado entre os municípios de São Leopoldo, Novo Hamburgo e Campo Bom na Região Metropolitana de Porto Alegre (RMPA), estado do Rio Grande do Sul. Está situada no médio curso do Rio dos Sinos, aproximadamente entre as coordenadas 29°40'15"S e 29°46'58"S de latitude e 50°59'51"W e 51°10'45"W de longitude, abrangendo 218 km^2 de extensão.

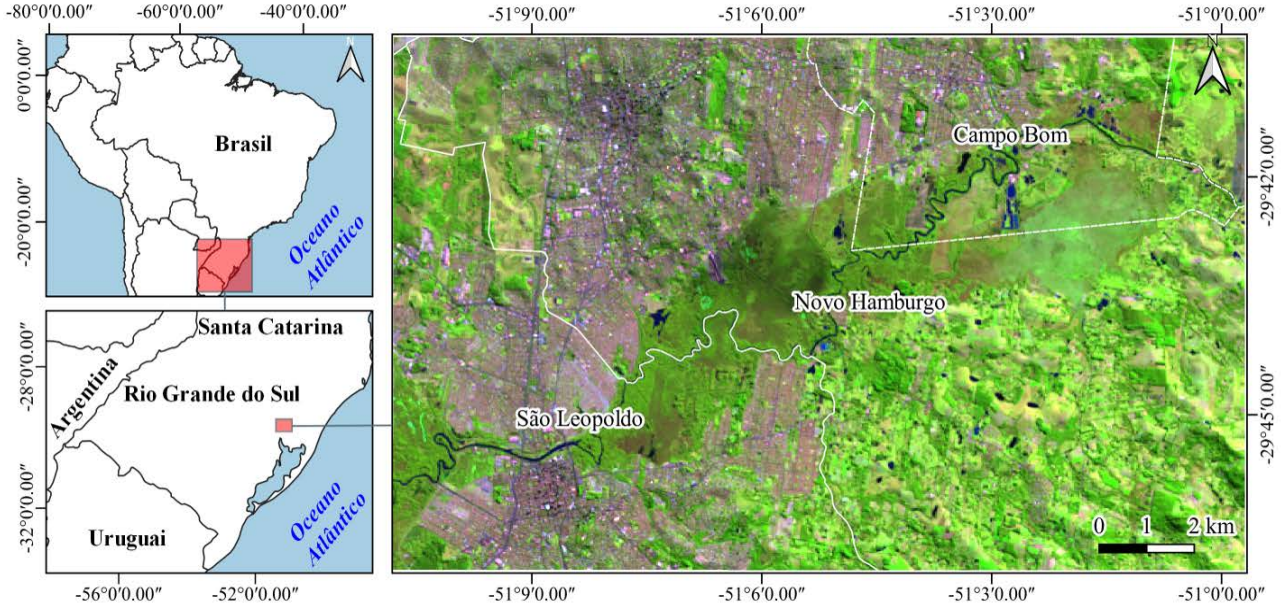

Figura 1 - Mapa de localização da área de estudo com a imagem OLI/Landsat-8 de 29/07/2013.

MATERIAIS E MÉTODOS

Os métodos foram estruturados e desenvolvidos nas seguintes etapas, conforme apresentado no fluxograma metodológico (Figura 2): (i) identificação das variáveis preditoras para o mapeamento da água superficial; (ii) processo de amostragem de classes relacionadas com alvos aquáticos e não aquáticos; (iii) classificação *Random Forest*; (iv) seleção de variáveis mais relevantes para o processo de classificação e; (v) avaliação das classificações e da incerteza (entropia de Shannon).

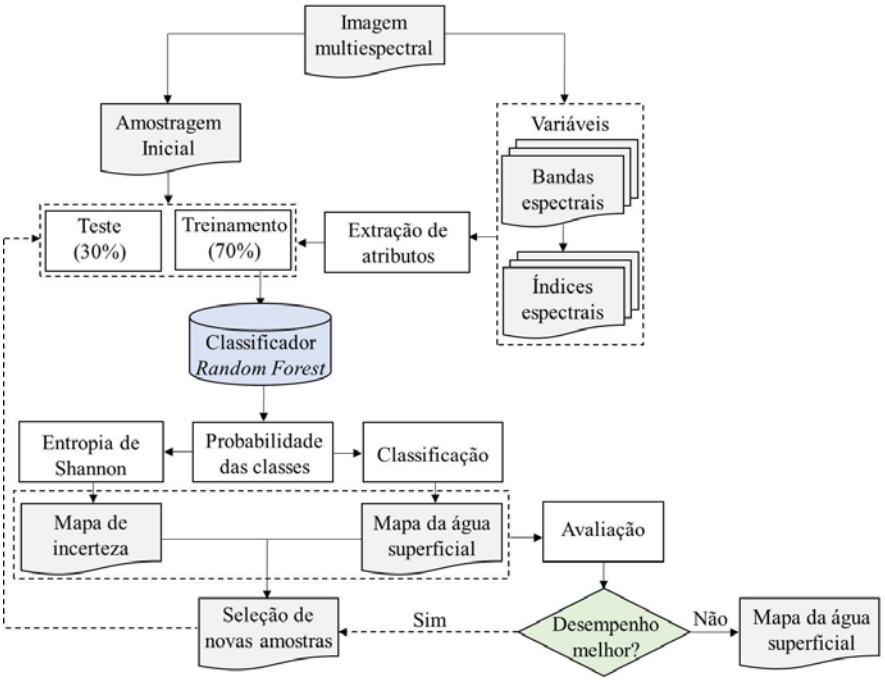

Figura 2 - Fluxograma com os procedimentos metodológicos.

Variáveis preditoras

Como variáveis foram utilizadas as bandas 2 a 7 do sensor *Operational Land Imager* (OLI) do satélite Landsat-8 e índices espectrais de água. O mapeamento das águas superficiais permanentes e temporárias na área de estudo corresponde a um evento de inundação de grande magnitude ocorrido em 30/08/2013 (Figura 3).

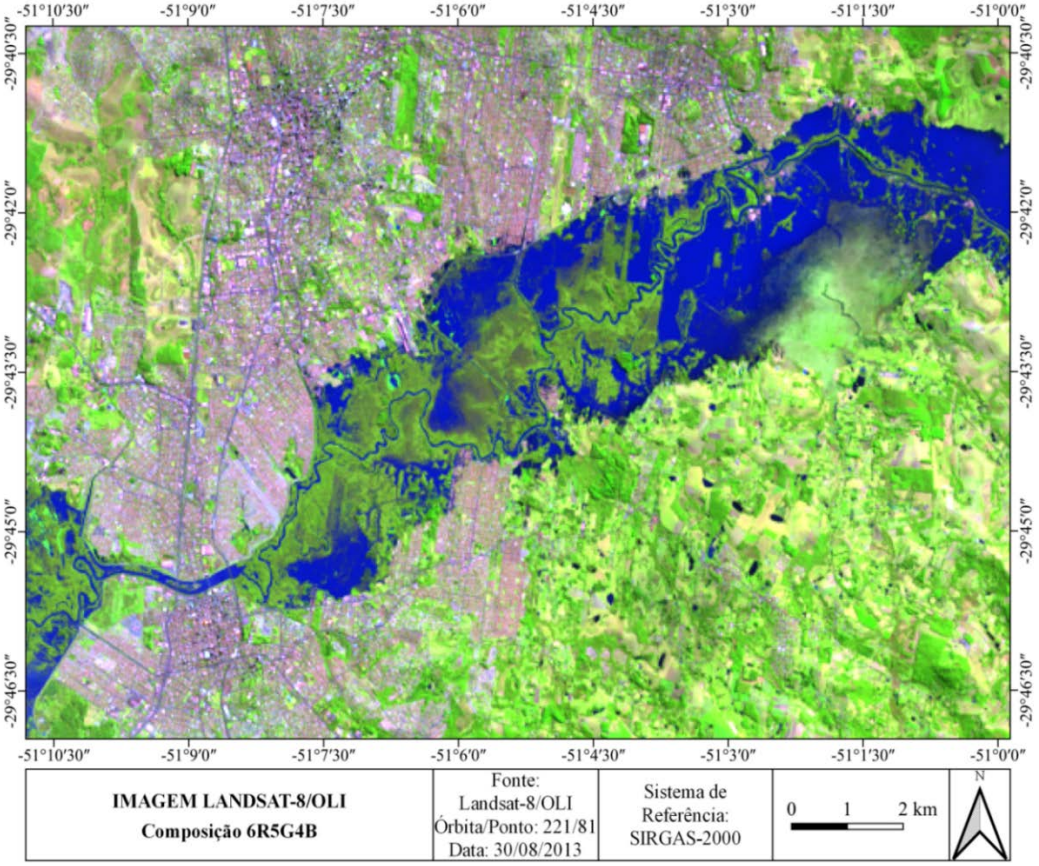

Figura 3 - Imagem OLI/Landsat-8 (composição 6R5G4B) com a inundação de 30/08/2013.

Os dados foram adquiridos por meio da plataforma *Google Earth Engine* (Gorelick et al., 2017) utilizando a linguagem de programação *JavaScript* para a exportação das bandas espectrais. Os dados da coleção 2 do Landsat-8/OLI são disponibilizados pelo *United States Geological Survey* (USGS) em refletância de superfície com correção atmosférica processados com o *Land Surface Reflectance Code* (LaSRC) (Vermote et al., 2018). O pré-processamento incluiu uma etapa complementar no *software* RStudio (RStudio Team, 2020) para a correção dos valores negativos nas bandas espectrais e a conversão para o intervalo de reflectância de superfície (ρ) entre 0 e 1.

Os índices espectrais obtidos para este estudo (Tabela 1) correspondem aos principais índices utilizados para o mapeamento da água superficial, como o *Tasseled Cap Wetness* (TCW) (Crist, 1985), *Normalized Difference Water Index* (NDWI) (McFeeters, 1996), *Modified Normalized Difference Water Index* (MNDWI) (Xu, 2006), *Water Ratio Index* (WRI) (Shen & Li, 2010), *Automated Water Extraction Index* (AWEI) (Feyisa et al., 2014), *Water Index* 2015 (WI2015) (Fisher et al., 2015) e *Multi-Band Water Index* (MBWI) (Wang et al., 2018b). A modificação do NDWI a partir da combinação de diferentes bandas espectrais do visível e do infravermelho possibilitou calcular outros sete índices espectrais da água (Ji et al., 2009; Li et al., 2013; Acharya et al., 2018b; Lefebvre et al., 2019). Os índices espectrais foram calculados a partir das bandas espectrais no *software* RStudio (RStudio Team, 2020), totalizando 21 variáveis preditoras.

Amostragem

A amostragem inicial para treinamento e teste foi produzida a partir da seleção de pontos (*pixels*) representativos da classe água e classe não água com base na interpretação visual dos alvos na imagem OLI/Landsat-8 no *software* QGIS versão 3.16.5 (QGIS Development Team, 2021). A amostragem da classe água foi realizada em águas superficiais permanentes e temporárias observáveis em diferentes situações na imagem de satélite, tais como, água no canal fluvial, reservatórios de água e em áreas inundadas.

Tabela 1 - Índices espectrais de água e equações para o cálculo

Índice	Equação
TCW	$0{,}0315 \times \rho_{blue} + 0{,}2021 \times \rho_{green} + 0{,}3102 \times \rho_{red} + 0{,}1594 \times \rho_{NIR} - 0{,}6806 \times \rho_{SWIR1} - 0{,}6109 \times \rho_{SWIR2}$
NDWI	$(\rho_{green} - \rho_{NIR}) / (\rho_{green} + \rho_{NIR})$
MNDWI	$(\rho_{green} - \rho_{SWIR1}) / (\rho_{green} + \rho_{SWIR1})$
$NDWI_{3,7}$	$(\rho_{green} - \rho_{SWIR2}) / (\rho_{green} + \rho_{SWIR2})$
$NDWI_{2,5}$	$(\rho_{blue} - \rho_{NIR}) / (\rho_{blue} + \rho_{NIR})$
$NDWI_{2,6}$	$(\rho_{blue} - \rho_{SWIR1}) / (\rho_{blue} + \rho_{SWIR1})$
$NDWI_{2,7}$	$(\rho_{blue} - \rho_{SWIR2}) / (\rho_{blue} + \rho_{SWIR2})$
$NDWI_{4,5}$	$(\rho_{red} - \rho_{NIR}) / (\rho_{red} + \rho_{NIR})$
$NDWI_{4,6}$	$(\rho_{red} - \rho_{SWIR1}) / (\rho_{red} + \rho_{SWIR1})$
$NDWI_{4,7}$	$(\rho_{red} - \rho_{SWIR2}) / (\rho_{red} + \rho_{SWIR2})$
WRI	$(\rho_{green} + \rho_{red}) / (\rho_{NIR} + \rho_{SWIR1})$
$AWEI_{sh}$	$\rho_{blue} + 0{,}25 \times \rho_{green} - 1{,}5 \times (\rho_{NIR} + \rho_{SWIR1}) - 0{,}25 \times \rho_{SWIR2}$
$AWEI_{nsh}$	$4 \times (\rho_{green} - \rho_{SWIR1}) - (0{,}25 \times \rho_{NIR} + 2{,}75 \times \rho_{SWIR2})$
WI2015	$1{,}7204 + 171 \times \rho_{green} + 3 \times \rho_{red} - 70 \times \rho_{NIR} - 45 \times \rho_{SWIR1} - 71 \times \rho_{SWIR2}$
MBWI	$2 \times \rho_{green} - \rho_{red} - \rho_{NIR} - \rho_{SWIR1} - \rho_{SWIR2}$

Em relação a classe não água, a amostragem foi realizada em diferentes tipos de uso do solo e cobertura vegetal para garantir alta variabilidade espectral dos alvos. Foi coletada a mesma proporção de amostragem para cada classe de mapeamento, totalizando 240 amostras iniciais para treinamento e validação do modelo *Random Forest* (RF) e para avaliação da classificação. O conjunto amostral foi subdividido aleatoriamente em 70% de amostras para treinamento e 30% de amostras para avaliação da classificação.

Posteriormente, para aumentar e melhorar a representatividade amostral foram realizadas novas etapas de amostragem. Para aumentar a acurácia global e reduzir as incertezas derivadas do processo de classificação, foram efetuadas novas etapas de amostragem com a seleção de novos pontos (*pixels*) considerando o mapa de incerteza (entropia de Shannon) resultante da classificação RF. O critério para a seleção de novas amostras foi determinado pela amostragem em *pixels* com alta incerteza classificados corretamente em relação à classificação inicial do *Random Forest*.

A fim de obter um bom balanceamento, a cada etapa foram adicionadas 10 amostras para cada classe de mapeamento, mantendo igual proporção de amostragem em cada classe. A cada nova etapa de amostragem, 30% das amostras foram adicionadas ao conjunto amostral inicial de teste e 70% das amostras ao conjunto amostral inicial de treinamento para gerar uma nova classificação *Random Forest* com todas as variáveis e uma classificação *Random Forest* com as variáveis selecionadas com o algoritmo *Recursive Feature Elimination* (RFE).

Classificador *Random Forest*

O *Random Forest* (RF) (Breiman, 2001) faz parte do grupo de algoritmos supervisionados de aprendizado de máquina que utiliza árvores de decisão (*decision trees*) para a classificação e regressão dos dados. É um método de aprendizado em conjunto (*ensemble-learning*) que gera múltiplas classificações e agrega seus resultados ao final do processo.

O *Random Forest* utiliza o método de ensacamento (*bagging*), no qual cada árvore de decisão é construída independentemente a partir de um subconjunto de amostras D_1, D_2, D_k (*bootstrap*) selecionado de forma aleatória de uma parte do conjunto de dados de entrada (D). No final, uma votação por maioria simples (Figura 4) é encaminhada para a predição final (Liaw & Wiener, 2002).

Para a construção de cada árvore do *Random Forest*, parte das amostras de treinamento não é selecionada no conjunto de amostras (*bootstrap*), representando cerca de um terço das amostras de treinamento (Breiman, 2001). Esse conjunto de amostras não utilizadas é definido como amostras *Out-Of-Bag* (OOB). O conjunto de amostras OOB de cada árvore de decisão é agregado e fornece a taxa de erro OOB que é utilizada para a validação do modelo (Liaw & Wiener, 2002).

Em relação aos parâmetros do RF, pode-se definir o número de variáveis que serão sorteadas aleatoriamente para cada nó da árvore de decisão comumente denominado m_{try} e o número de árvores de decisão comumente denominado n_{tree} que serão geradas pelo modelo.

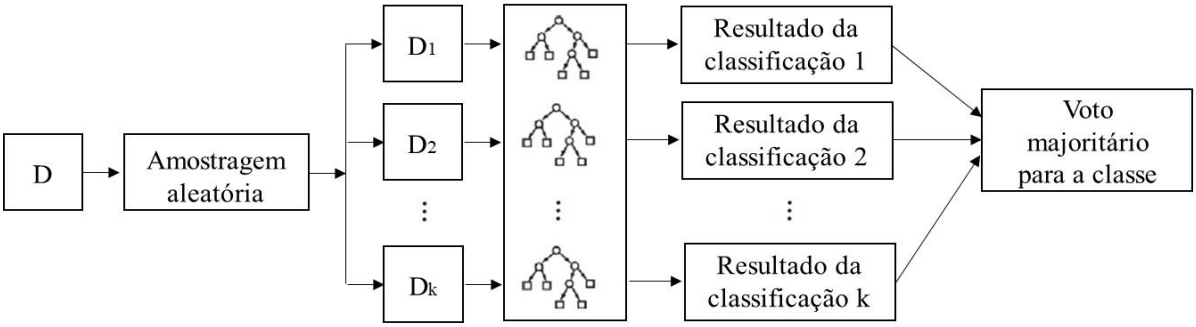

Figura 4 - Classificação *Random Forest*. Fonte: Adaptado de Wang et al. (2015).

O valor para o parâmetro m_{try} foi definido pela raiz quadrada do número de variáveis de entrada e para o n_{tree} foi definido 500 árvores de decisão (Belgiu & Dragut, 2016). A execução do algoritmo RF foi realizada com o pacote *randomForest* (Liaw & Wiener, 2022) em linguagem R (R Core Team, 2021) no RStudio (RStudio Team, 2020). Para a classificação, o pacote *randomForest* utiliza o índice Gini como critério para as divisões de nós (*splitting*) nas árvores de decisão (Liaw & Wiener, 2022

Seleção de variáveis

Uma grande quantidade de variáveis em modelos preditivos como o *Random Forest* pode conter variáveis altamente correlacionadas que aumentam o custo computacional relativo ao processamento e armazenamento sem contribuir significativamente para o aumento da acurácia da classificação final. Além disso, reduzir o número de variáveis no modelo pode ajudar na melhoria do entendimento da relação entre as variáveis e o fenômeno de interesse.

Neste estudo, foi utilizado o método de seleção de variáveis *Recursive Feature Elimination* (RFE) proposto por Guyon et al. (2002) com o objetivo de melhorar o desempenho do modelo para a classificação. Yagmur et al. (2019), por exemplo, obtiveram resultados satisfatórios empregando o RFE para selecionar variáveis representadas por bandas e índices espectrais para o mapeamento de águas superficiais.

O RFE é um algoritmo que seleciona as variáveis mais relevantes para a classificação. O RFE aplica um processo de seleção reversa (*stepwise backward*) para encontrar a combinação ideal de variáveis. Primeiro, constrói-se um modelo baseado em todas as variáveis e calcula-se a importância de cada uma para o modelo. Em seguida, ordenam-se as variáveis removendo as de menor importância de forma iterativa com base nas métricas de avaliação do modelo (exatidão e índice Kappa).

Este processo continua até que um menor subconjunto de variáveis é selecionado no modelo final quando a exatidão atinge o nível máximo (Bulut, 2021). O RFE identifica variáveis importantes e elimina variáveis com baixa importância que não estão suficientemente associadas à discriminação das classes de interesse (Kuhn & Johnson, 2019).

Dessa forma, as variáveis selecionadas pelo RFE são utilizadas como um novo conjunto de dados no modelo RF. O RFE foi calculado com a função *rfeControl* do pacote *caret* (Kuhn, 2022) no *software* RStudio (RStudio Team, 2020). Para a seleção de variáveis com o RFE foi utilizada a validação cruzada (*k-fold*) dividindo o conjunto amostral em 10 subconjuntos, sendo 80% das amostras para treinamento e 20% para teste em cada etapa de amostragem.

Medida de Incerteza da Classificação

A identificação das fontes de incertezas e sua quantificação torna-se um aspecto importante para a seleção do método ou modelo mais adequado para a redução da subestimação ou superestimação no mapeamento das inundações (Teng et al., 2017). Um dos índices usados para avaliar as incertezas espacialmente distribuídas é a medida de entropia de Shannon (E), derivada da teoria da informação (Shannon, 1948):

$$E = -\sum_{i=1}^{n} pi \; x \; log2 \; x \; pi$$

onde, pi é a probabilidade associada a classe i e n é o número de classes. A partir do cálculo da entropia de Shannon no *software* RStudio (RStudio Team, 2020), obteve-se o mapa com a distribuição espacial da incerteza derivada do processo de classificação com o *Random Forest*.

Avaliação das classificações

A partir do conjunto amostral, 30% de amostras independentes foram utilizadas como teste para a avaliação final da classificação, garantindo uma amostragem mínima de teste superior a 30 amostras, conforme recomendado por Congalton (1991).

Para a avaliação final dos mapas oriundos das classificações, foi utilizada a matriz de confusão considerando apenas duas classes (Tabela 2), sendo que a classe água corresponde ao positivo e a classe não água corresponde ao negativo.

A partir da matriz de confusão foram extraídas as seguintes métricas para medir e comparar o desempenho das classificações:

Tabela 2 - Matriz de confusão

		Referência	
		Positivo (água)	**Negativo (não água)**
Classificação	Positivo (água)	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo (não água)	Falso Negativo (FN)	Verdadeiro Negativo (VN)

$$\text{Exatidão Total (ET)} = \frac{VP + VN}{VP + FP + FN + VN}$$

$$\text{Sensibilidade (S)} = \frac{VP}{VP + FN}$$

$$\text{Precisão (P)} = \frac{VP}{VP + FP}$$

$$\text{F1-}Score = 2\,\frac{P * S}{P + S}$$

$$\text{Erro de Omissão}_{\text{água}}\ (EO_{\text{água}}) = \frac{FN}{VP+FN}$$

$$\text{Erro de Comissão}_{\text{água}}\ (EC_{\text{água}}) = \frac{FP}{VP+FP}$$

$$\text{Erro de Omissão}_{\text{não água}}\ (EO_{\text{não água}}) = \frac{FP}{FP+VN}$$

$$\text{Erro de Comissão}_{\text{não água}}\ (EC_{\text{não água}}) = \frac{FN}{FN+VN}$$

Para a avaliação do desempenho e comparação entre as classificações RF foram utilizadas as métricas ROC (*Receiver Operating Characteristics*) e AUC (*Area Under the Curve*), conforme descrito por Fawcett (2006). A ROC é um gráfico calculado em função da taxa verdadeira positiva (plotada no eixo Y) e taxa de falso positivo (plotada no eixo X). A AUC resume a linha do gráfico da curva ROC em um único valor e varia de 0,5 (classificador completamente aleatório) a 1,0 (classificador perfeitamente discriminador). As classificações RF foram realizadas até a etapa em que não ocorreu mais melhoria no desempenho do mapeamento com as métricas de avaliação.

A avaliação e quantificação da distribuição espacial das incertezas foram efetuadas a partir da amostragem aleatória com 10.000 pontos nos mapas com a diferença da entropia de Shannon entre as classificações RF. A partir da referida amostragem foram gerados gráficos de frequência relativa acumulada e de dispersão das amostras (*boxplot*) para a visualização da distribuição e quantificação das incertezas derivadas do processo de classificação *Random Forest*.

RESULTADOS E DISCUSSÕES

Classificação *Random Forest*

Avaliação dos modelos Random Forest

A classificação inicial RF com 168 amostras de treinamento (etapa 1) e 21 variáveis apresentou uma taxa de erro OOB de 2,38% (modelo RF inicial). A classificação RF com a amostragem inicial (etapa 1) e com a seleção de variáveis pelo RFE manteve a taxa de erro OOB em 2,38% (modelo RF inicial RFE). Já o aumento da amostragem considerando como referência o mapa de incerteza resultante da classificação inicial com o RFE resultou na redução da taxa de erro OOB (Tabela 3). A menor taxa de erro OOB foi para a modelo RF com 238 amostras de treinamento (etapa 6) que apresentou 1,68% (modelo RF final) e 0,00% (modelo RF final RFE). O comparativo entre as taxas de erro OOB pelo modelo RF inicial e modelo RF final indica a tendência geral de estabilização do erro após 100 árvores de decisão (Figura 5).

No modelo RF inicial, modelo RF inicial RFE e modelo RF final, a taxa de erro OOB inicia com instabilidade e após 100 árvores de decisão tende a estabilizar. Já no modelo RF final RFE a estabilidade da taxa de erro OOB ocorre antes de 50 árvores de decisão.

Tabela 3 - Métricas de avaliação de acordo com amostragem em cada etapa

Etapa	Número total de amostras	Número de amostras de treinamento	OOB com todas variáveis	OOB com variáveis selecionadas (RFE)
1	240	168	2,38%	2,38%
2	260	182	2,20%	1,65%
3	280	196	2,04%	0,51%
4	300	210	0,95%	0,95%
5	320	224	1,34%	0,45%
6	340	238	1,68%	0,00%
7	360	252	1,98%	0,00%

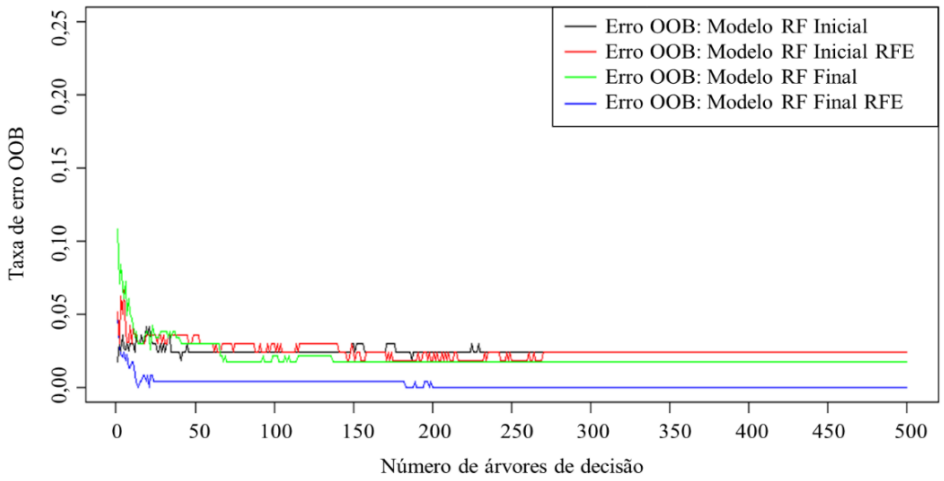

Figura 5 - Gráfico com a taxa de erro OOB em relação ao número de árvores de decisão.

No modelo RF final (etapa 6), as variáveis mais importantes para a classificação baseadas na métrica redução média da acurácia foram, por ordem de importância, o $NDWI_{4,7}$, MBWI, AWEISH, $NDWI_{3,7}$ e WI2015 (Figura 6). A seleção de variáveis pelo RFE identificou no modelo RF final que o conjunto representado pelo $NDWI_{4,7}$, MBWI, WI2015 e $NDWI_{3,7}$ apresentou a maior exatidão para a classificação

Destaca-se que índices espectrais amplamente utilizados para o mapeamento da água superficial, como o NDWI e MNDWI, apresentaram baixa importância no modelo para a predição das classes de mapeamento da água superficial do evento de inundação analisado na área de estudo.

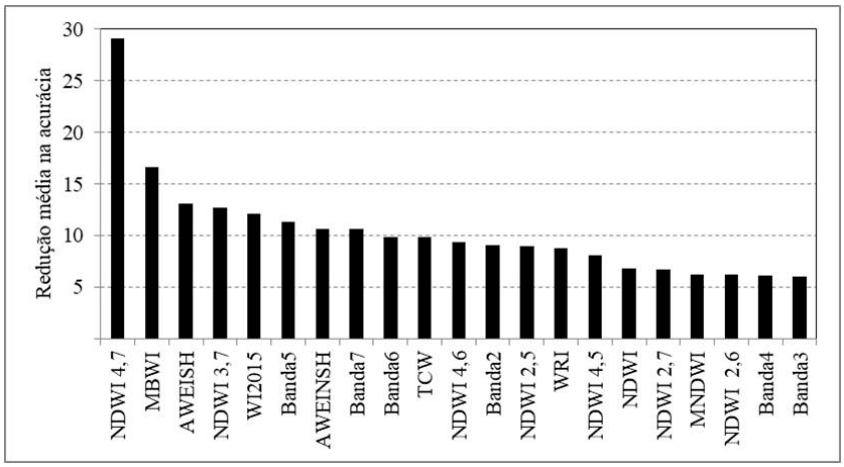

Figura 6 - Redução média da acurácia das variáveis preditoras no modelo *Random Forest* final (etapa 6).

Avaliação das classificações Random Forest

A classificação com o modelo RF inicial (Figura 7a) apresentou bom desempenho no mapeamento.

No entanto, há ocorrência de falsos positivos em áreas urbanas devido à presença de sombras oriundas de edificações verticalizadas, implicando em superestimação da classe água nestas áreas. Com a seleção de variáveis pelo RFE no modelo RF inicial (Figura 7b) estes falsos positivos não foram removidos de forma significativa nas áreas destacadas nos mapas.

A classificação RF final (Figura 8a) e a classificação RF final com RFE (Figura 8b) resultou em maior remoção de falsos positivos nas áreas urbanas com edificações verticalizadas, porém, não eliminando totalmente os falsos positivos da classe água nas áreas destacadas nos mapas.

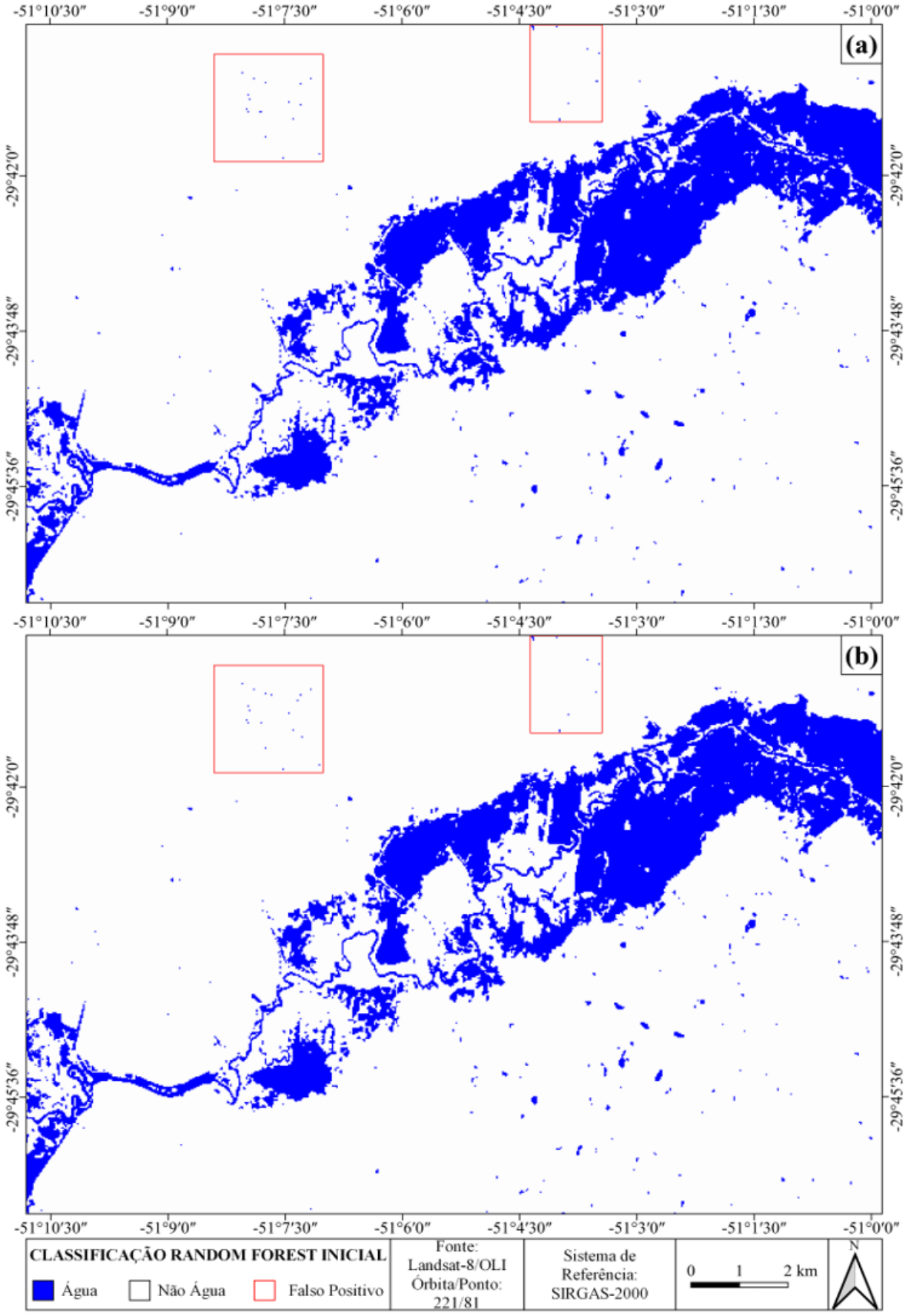

Figura 7 - Classificação *Random Forest* inicial (a) e classificação *Random Forest* inicial com RFE (b).

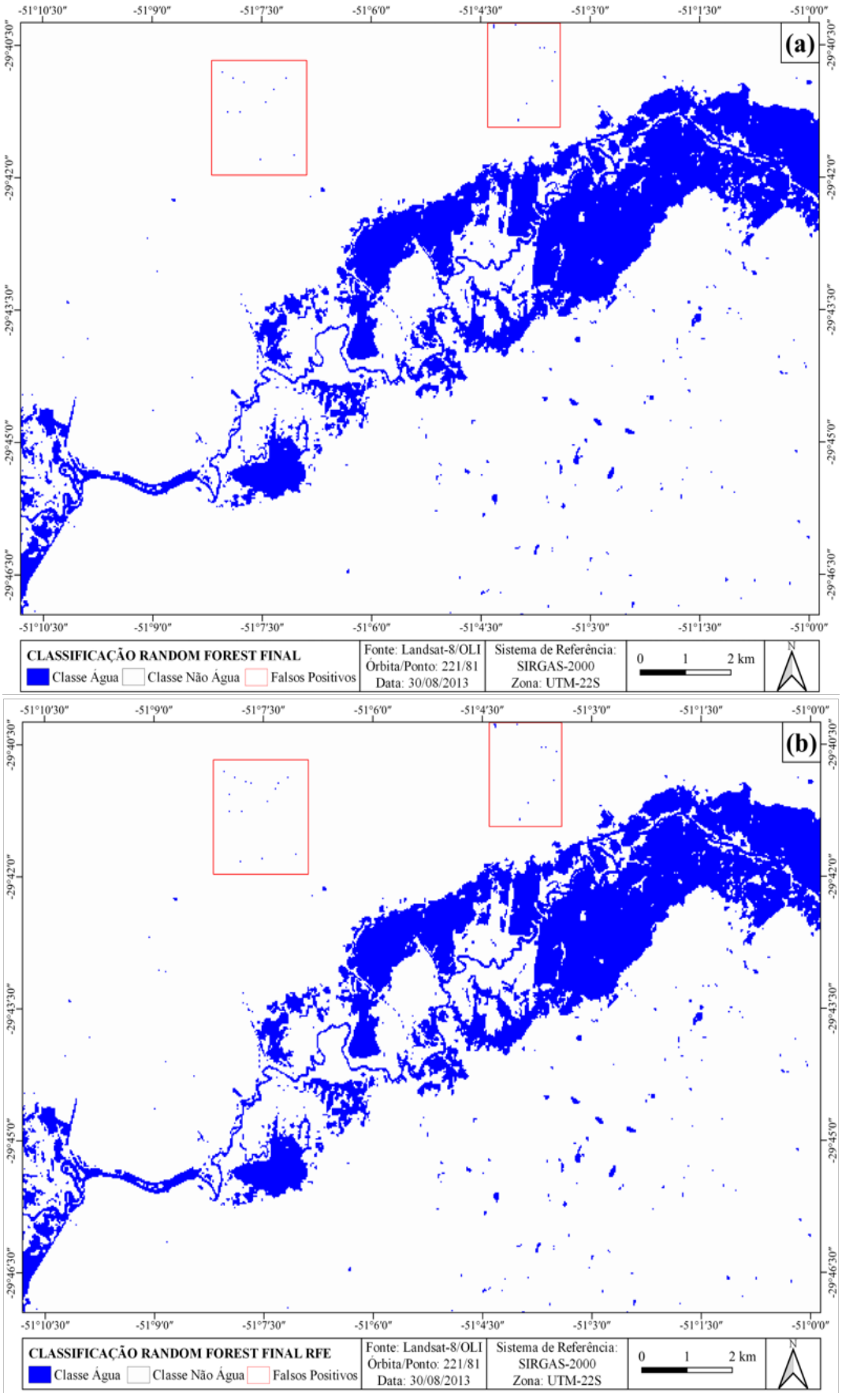

Figura 8 - Classificação *Random Forest* final (a) e classificação *Random Forest* final com *RFE* (b).

A comparação da avaliação das classificações *Random Forest* (Figura 9) mostrou que a classificação RF final com as variáveis selecionadas pelo RFE apresentou melhor desempenho em relação às métricas de exatidão (98,0%), sensibilidade (96,1%), precisão (100%) e F1-Score (98,0%).

Em relação à avaliação dos erros (Figura 10), a classificação *Random Forest* final com as variáveis

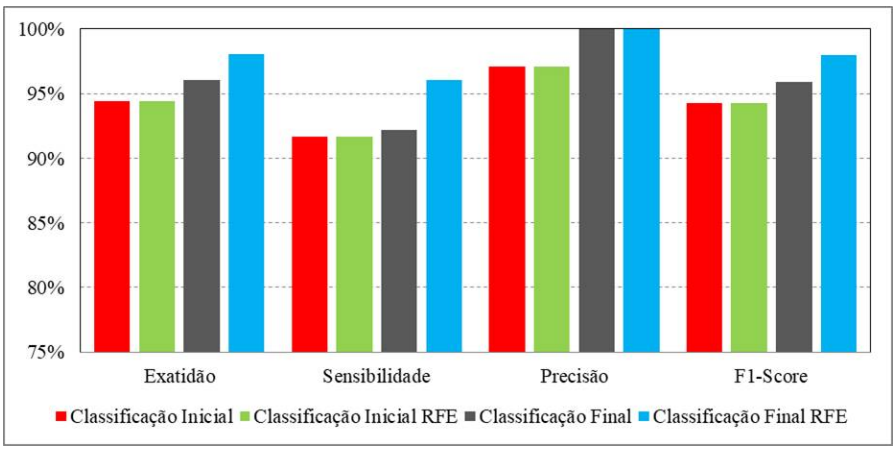

Figura 9 - Métricas de avaliação da classificação.

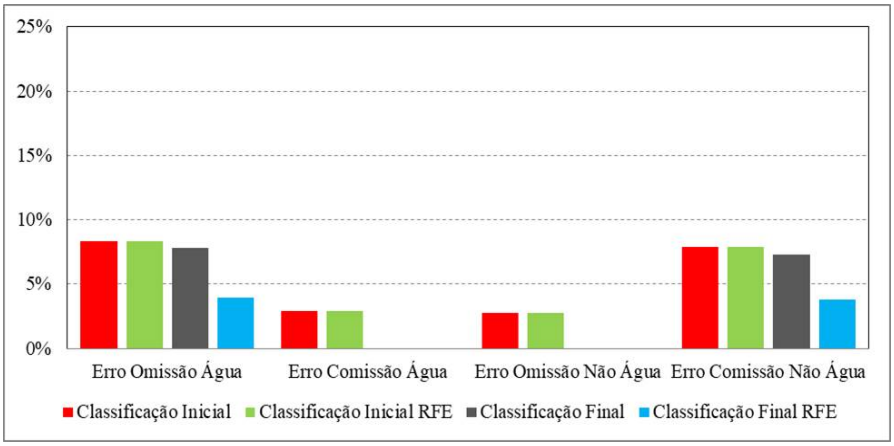

Figura 10 - Avaliação dos erros da classificação.

selecionadas pelo RFE apresentou os menores erros de omissão na classe água (3,9%) e na classe não água (0,0%) e menores erros de comissão na classe água (0,0%) e classe não água (3,8%), resultando na redução da subestimação e superestimação das classes de mapeamento.

Análise da ROC e AUC das classificações

No gráfico ROC (Figura 11) que compara o desempenho das quatro classificações *Random Forest*, é possível observar que a curva da classificação RF final RFE representou o classificador com melhor desempenho. Este desempenho também é demonstrado pela AUC que na classificação RF inicial correspondeu a 0,833 e na classificação RF final RFE foi de 0,980.

Dessa forma, os resultados das métricas de avaliação com amostras independentes indicam um bom desempenho da classificação RF final RFE para o mapeamento das águas superficiais permanentes e temporárias associadas às inundações na área de estudo.

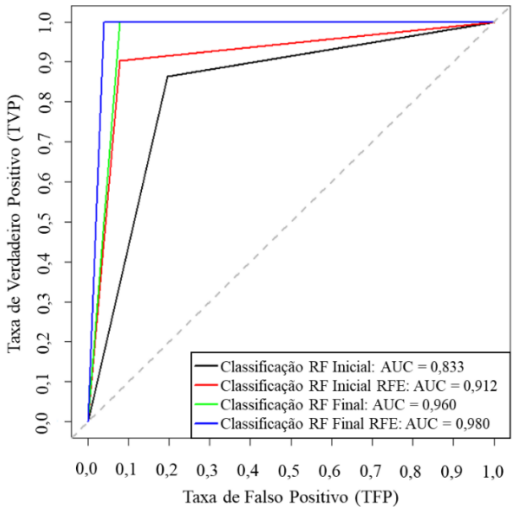

Figura 11 - Gráfico da ROC e valor da AUC comparando as classificações *Random Forest*.

Análise da Incerteza

Análise da distribuição espacial da incerteza

Nos mapas de incerteza derivados das classificações RF (Figura 12), em geral, predominam baixos valores de entropia indicando boa confiança para o mapeamento da água superficial.

Entretanto, a presença mais expressiva de altos valores de entropia de Shannon ocorre na classificação RF inicial (Figura 12a), indicando maior incerteza em áreas inundadas (I), transição entre área inundada e área não inundada (II) e área não inundada (III).

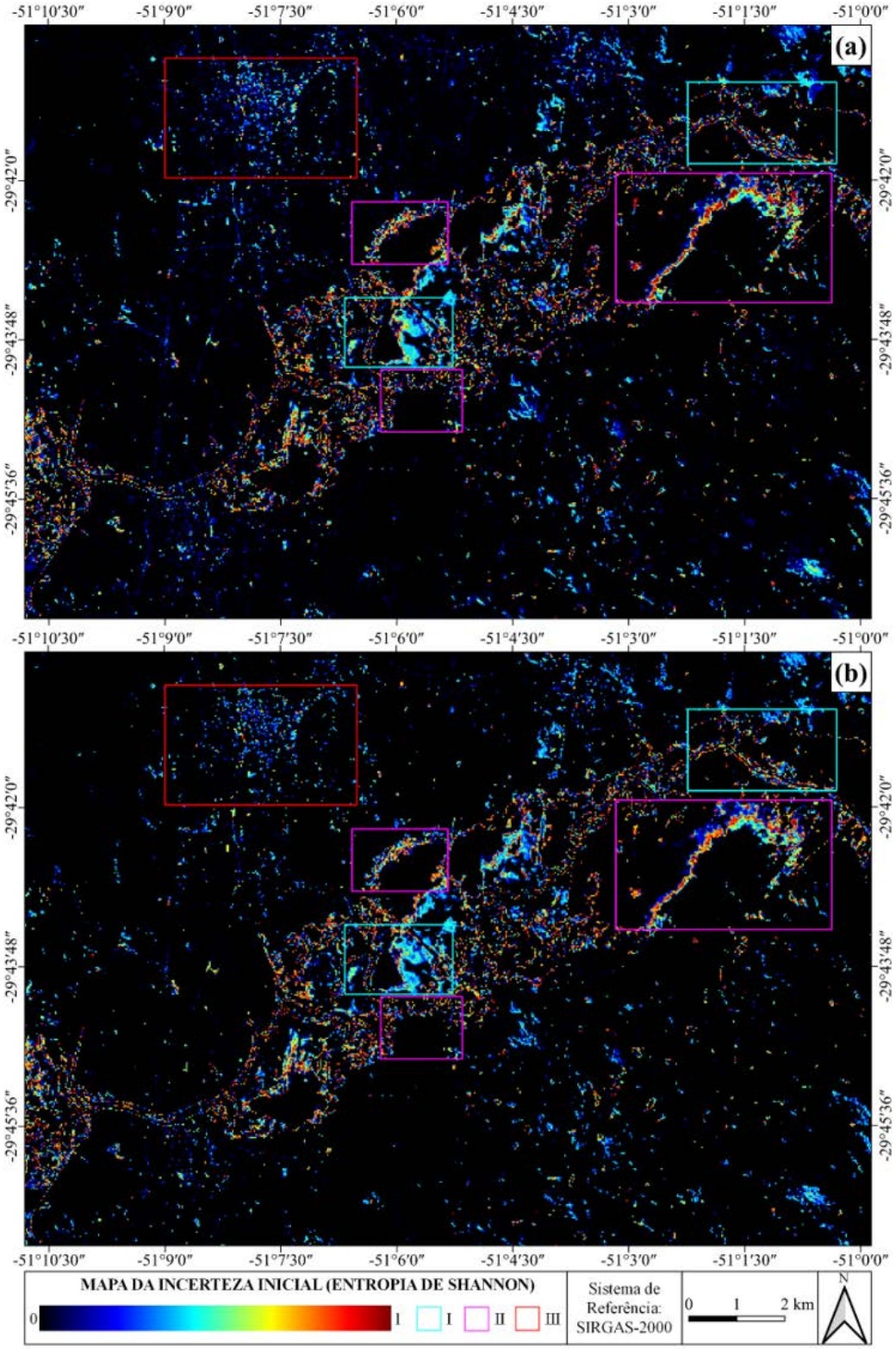

Figura 12 - Mapa de incerteza (entropia de Shannon) da classificação RF inicial (a) e inicial RFE (b).

A classificação RF inicial com a seleção de variáveis com o RFE (Figura 12b) visualmente resultou em baixa redução da entropia de Shannon nas áreas com maiores incertezas em relação à classificação RF inicial com todas as variáveis.

No mapa de incerteza derivado da classificação RF final (Figura 13a) ocorre redução da entropia de Shannon em relação ao mapa de incerteza da classificação RF inicial, porém ainda com presença de áreas com alta entropia. A distribuição espacial com as menores incertezas resultou da classificação RF final RFE (Figura 13b), indicando uma classificação com maior confiança para o mapeamento da água superficial associada às inundações.

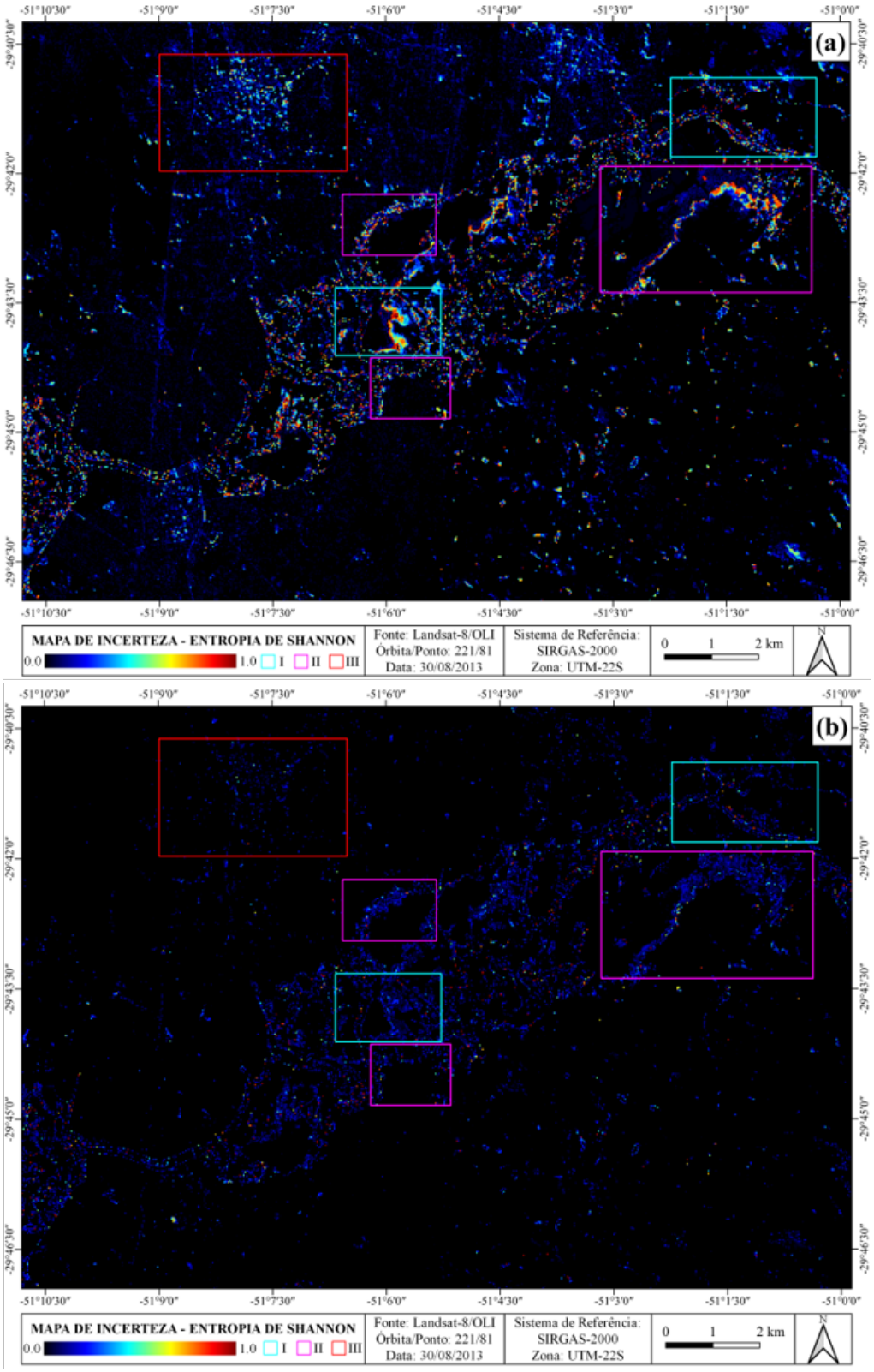

Figura 13 - Mapa de incerteza (entropia de Shannon) da classificação RF final (a) e final RFE (b).

O classificador RF final RFE conseguiu determinar a classificação com alto nível de certeza principalmente nas áreas inundadas onde a mistura da resposta espectral da água com alvos adjacentes não aquáticos é mais comum e onde ocorre maior variabilidade de uso do solo e

cobertura vegetal na área de estudo.

Conforme observa-se na figura 14, nos terrenos inundados as maiores incertezas ocorrem em trechos com vegetação arbustiva ou arbórea ciliar (I-A) e em áreas com vegetação arbustiva (I-B); na transição entre terrenos inundados e não inundados em áreas com vegetação de gramíneas (II-A), áreas com vegetação arbustiva e arbórea (II-B) e com presença de áreas urbanas (II-C) e; nos terrenos não inundados com edificações verticalizadas em áreas densamente urbanizadas (III-A).

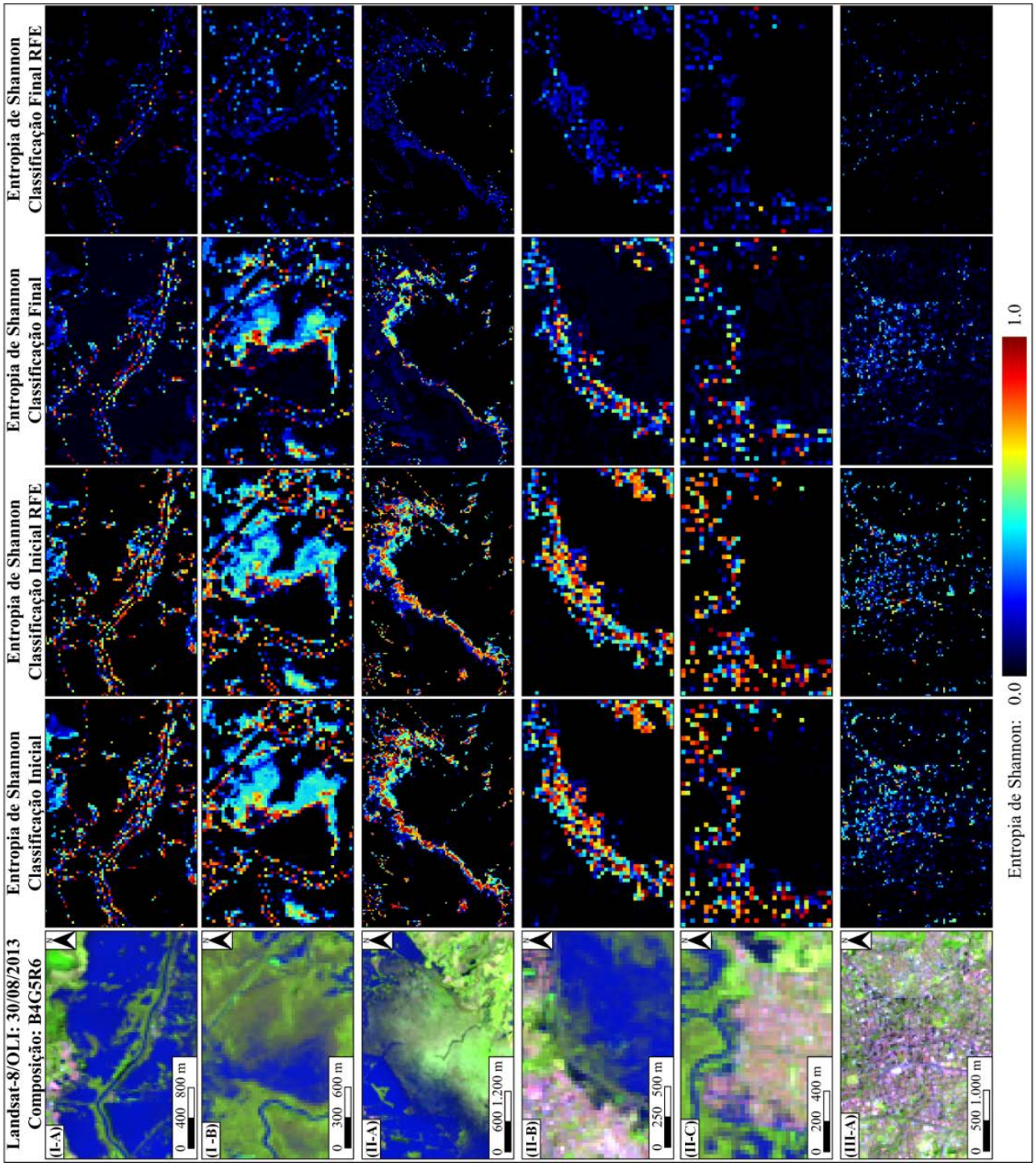

Figura 14 - Exemplos de áreas com alta incerteza (entropia de Shannon) nas classificações RF.

A frequência relativa acumulada (Figura 15) obtida a partir da amostragem aleatória de 10.000 pontos nos mapas de incerteza derivados das classificações RF, indica que: a porcentagem de amostras que apresentou valor de entropia de Shannon abaixo de 0,1 foi de 90,4% no mapa de entropia inicial; 91,1% no mapa de entropia inicial RFE; 91,9% no mapa de entropia final e 96,7% no mapa de entropia final RFE, sendo este último o mapeamento que apresentou maior grau de certeza.

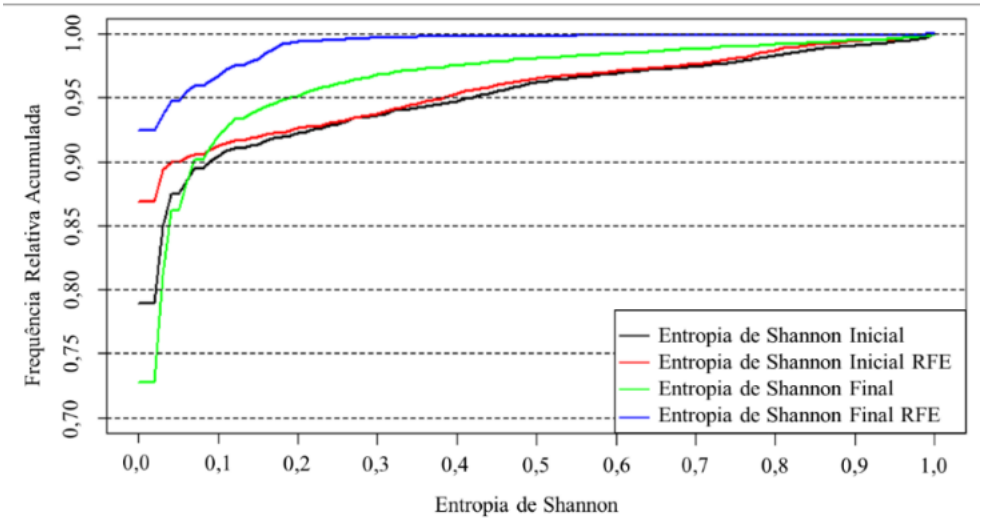

Figura 15 - Gráfico da frequência relativa acumulada com base na amostragem aleatória de 10.000 pontos nos mapas de incerteza das classificações RF.

Análise das diferenças de incertezas entre as classificações

Em geral, nos mapas das diferenças de incerteza predominam áreas sem alterações no valor de entropia entre as classificações (Figura 16).

No mapa da diferença de entropia de Shannon entre a classificação RF inicial e a classificação RF inicial RFE (Figura 16a) observa-se maior presença de áreas com redução da incerteza (intervalo entre -0,2 e 0,0) nos terrenos não inundados (III), em contraposição com a presença de áreas com aumento de incerteza (intervalo entre 0,0 e 0,2) nos terrenos inundados (I) e na transição entre terrenos (II).

O mapa da diferença de entropia entre a classificação RF inicial e a classificação RF final (Figura 16b) é caracterizado pelo aumento da incerteza (intervalo entre 0,0 e 0,2) nos terrenos não inundados (III) e pela maior redução da incerteza (intervalo entre -0,2 a 0,0 e inferior a -0,2) nos terrenos inundados (I) e na transição entre os terrenos (II).

No mapa de diferença de entropia entre a classificação RF inicial e a classificação RF final RFE (Figura 16c), observa-se a maior redução de incerteza (intervalo entre -0,2 e -0,4 e inferior a -0,4), principalmente nos terrenos inundados (I) e na transição entre os terrenos (II). Já nos terrenos não inundados (III) observa-se a redução da entropia principalmente no intervalo entre -0,2 e 0,0. Em termos gerais, a redução da incerteza foi mais significativa nos terrenos inundados (I) e na transição entre os terrenos (II) em relação aos terrenos não inundados (III) localizados em áreas urbanas. A redução da incerteza (Figura 17) foi maior nas áreas inundadas (I-A e I-B) e na transição entre áreas inundadas e não inundadas (II-A, II-B e III-C), caracterizadas pela mistura da resposta espectral da água com alvos adjacentes não aquáticos. A redução da incerteza nestas áreas contribuiu para eliminar *pixels* classificados incorretamente como não água (falsos negativos), promovendo a classificação correta desses *pixels* na classe água

Nas áreas não inundadas com presença de edificações verticalizadas em locais densamente urbanizados (III-A), caracterizadas pela maior presença de alvos com resposta espectral similar a água (sombra e pavimento asfáltico), a redução das incertezas foi menor em relação às áreas inundadas, resultando na remoção parcial de *pixels* classificados incorretamente como água (falsos positivos). Nestas áreas, pode ocorrer ainda a mistura da resposta espectral de diferentes alvos urbanos que podem resultar na resposta similar à da água, aumentando a confusão entre as classes e contribuindo para a ocorrência de falsos positivos.

A frequência relativa acumulada (Figura 18) calculada a partir da amostragem aleatória de 10.000 pontos nos mapas de diferença de incerteza derivados das classificações RF indica que: no mapa de diferença entre a entropia inicial e entropia inicial RFE 15,7% das amostras apresentaram valor abaixo de zero; no mapa de diferença entre a entropia inicial e entropia final 13,8% das amostras apresentaram valor abaixo de zero e; por fim, no mapa de diferença entre a entropia inicial e entropia final RFE a porcentagem de amostras que apresentou valor abaixo e zero foi de 20,6%, indicando uma maior redução na incerteza do mapeamento ao considerar na classificação uma amostragem modificada a partir da entropia e um subconjunto de variáveis selecionadas.

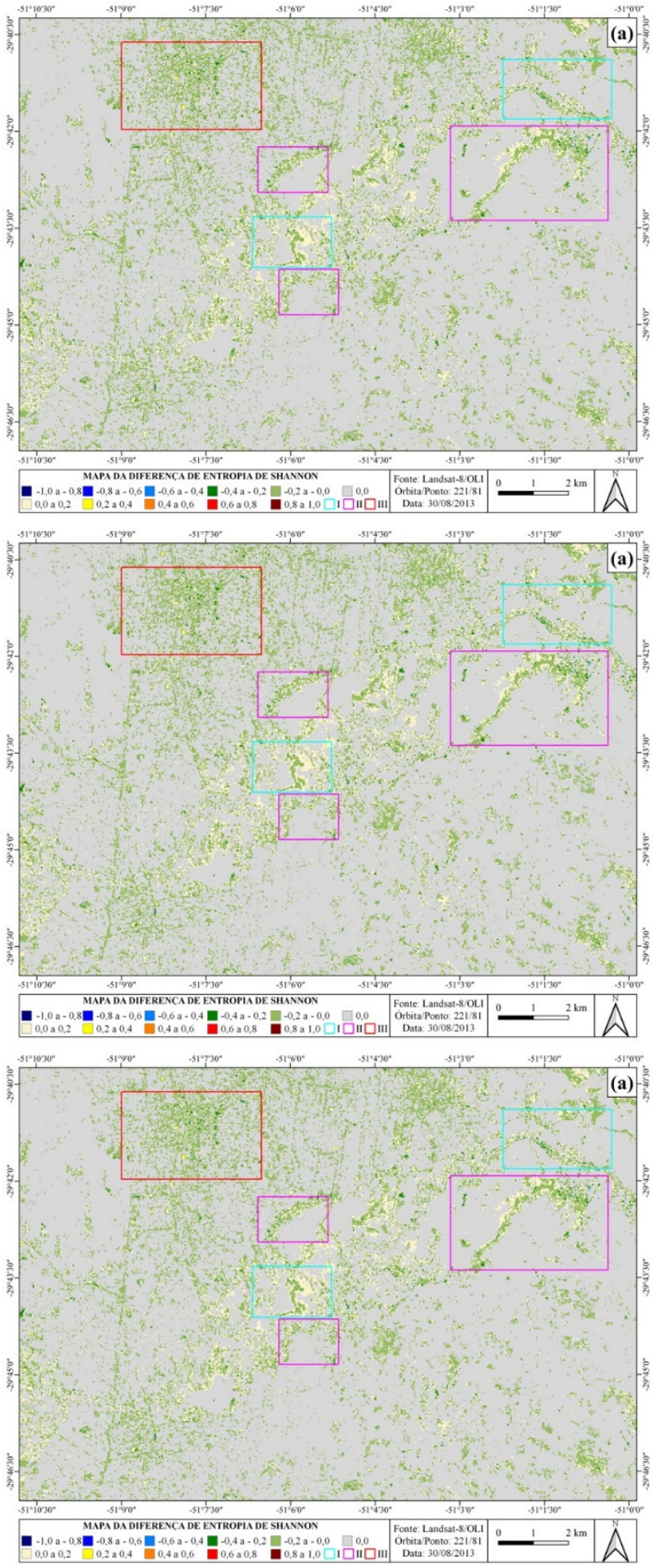

Figura 16 - Mapa de diferença de incertezas (entropia de Shannon) da classificação RF inicial e classificação RF inicial RFE (a), classificação RF inicial e final (b) e classificação RF inicial e final RFE (c).

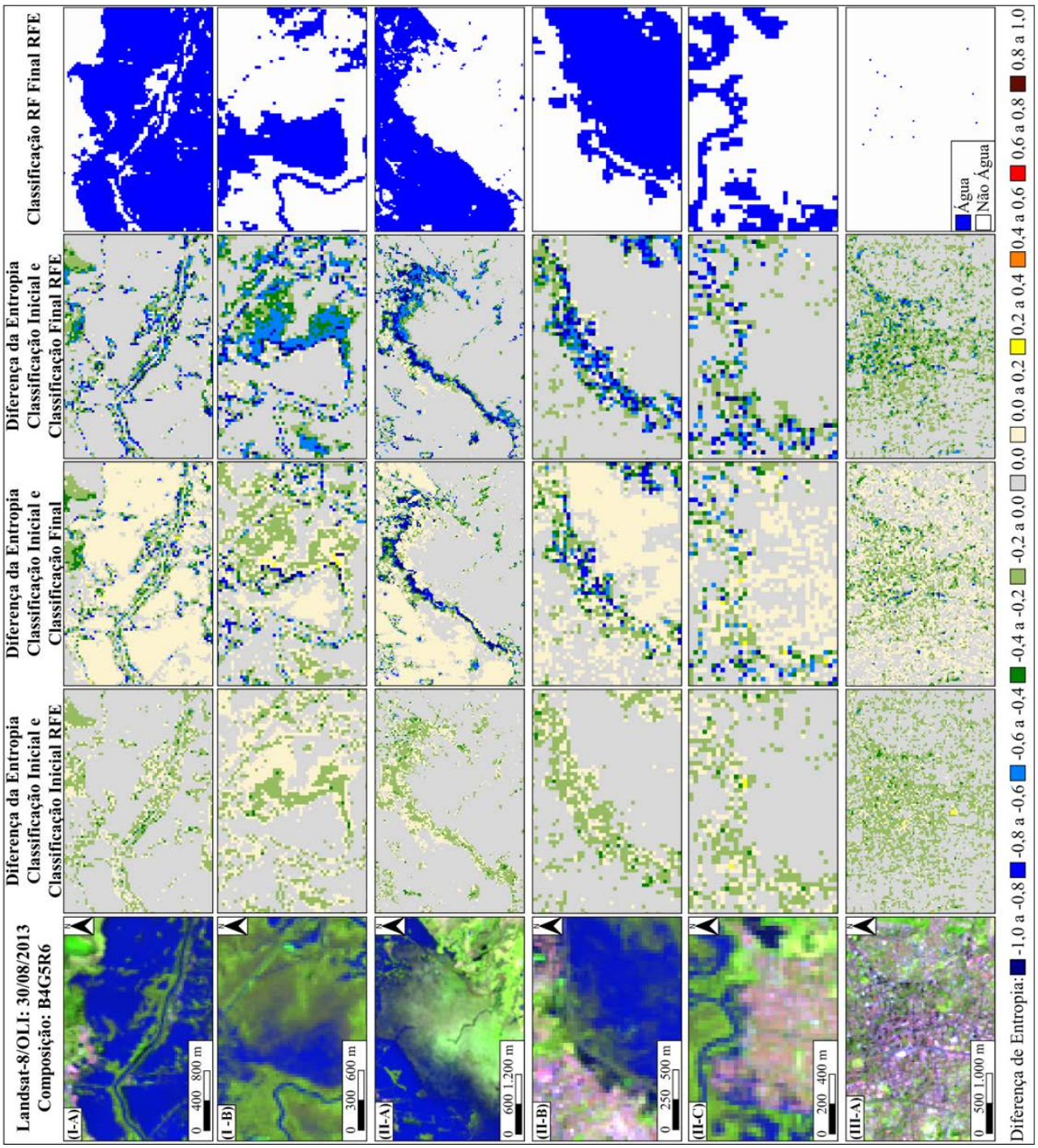

Figura 17 - Exemplo de áreas com diferenças de incerteza nas classificações RF.

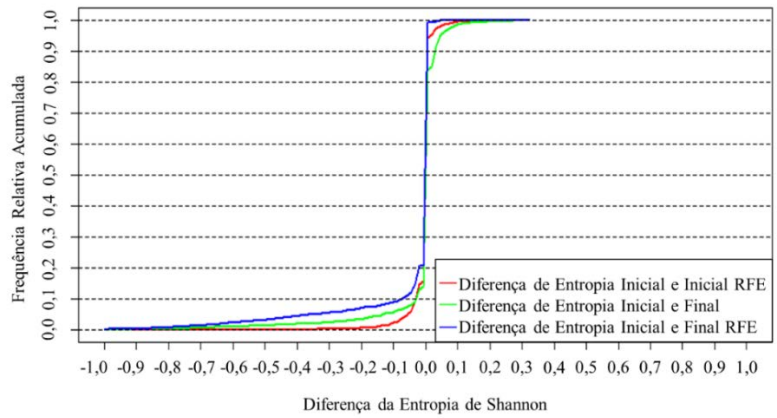

Figura 18 - Gráfico da frequência relativa acumulada com base na amostragem aleatória de 10.000 pontos nos mapas com a diferença de incerteza das classificações RF.

Destaca-se que a maior porcentagem das amostras está associada ao valor de diferença de entropia igual à zero.

A quantificação das diferenças de incerteza (Figura 19) com a amostragem aleatória de 10.000 pontos sobre o mapa da diferença da entropia de Shannon da classificação RF inicial e da classificação RF inicial RFE totalizou uma redução de incerteza em 1.570 amostras com valor de diferença de entropia inferior a zero e aumento de incerteza em 568 amostras com valor de diferença superior a zero. Um total de 7.862 amostras apresentou valor igual à zero, indicando que nestes pontos não ocorreu alteração na entropia de Shannon.

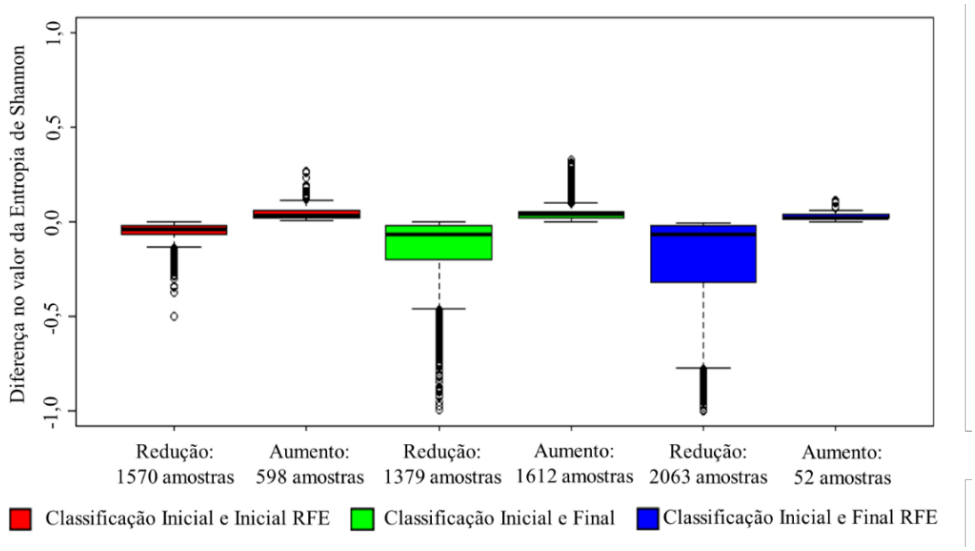

Figura 19 - Gráfico (*boxplot*) da amostragem aleatória nos mapas de diferença de entropia de Shannon.

Em relação à diferença de entropia de Shannon entre a classificação RF inicial e classificação RF final, houve uma redução de incerteza em 1.379 amostras e aumento de incerteza em 1.612 amostras. Em 7.009 amostras não ocorreu alteração na entropia de Shannon.

A redução de entropia foi mais expressiva entre a classificação RF inicial e a classificação RF final RFE com redução do valor em 2.063 amostras e o aumento do valor em 52 amostras, o que indica maior confiabilidade no processo de classificação com o *Random Forest* empregando o aumento da amostragem juntamente com a seleção de variáveis. Um total de 7.885 amostras apresentou valor igual à zero, indicando que nestes pontos não ocorreu alteração na entropia de Shannon. A amostragem aleatória estratificada de 10.000 pontos sobre a classificação RF final RFE e o mapa da diferença entre a classificação RF inicial e a classificação RF final RFE indicam que a redução da entropia foi mais significativa na classe água com 2.545 amostras com valor de diferença de entropia inferior a zero (Figura 20).

Dessa forma, o mapeamento da classe água apresenta maior confiança na classificação RF final RFE, quando comparada à classe não água.

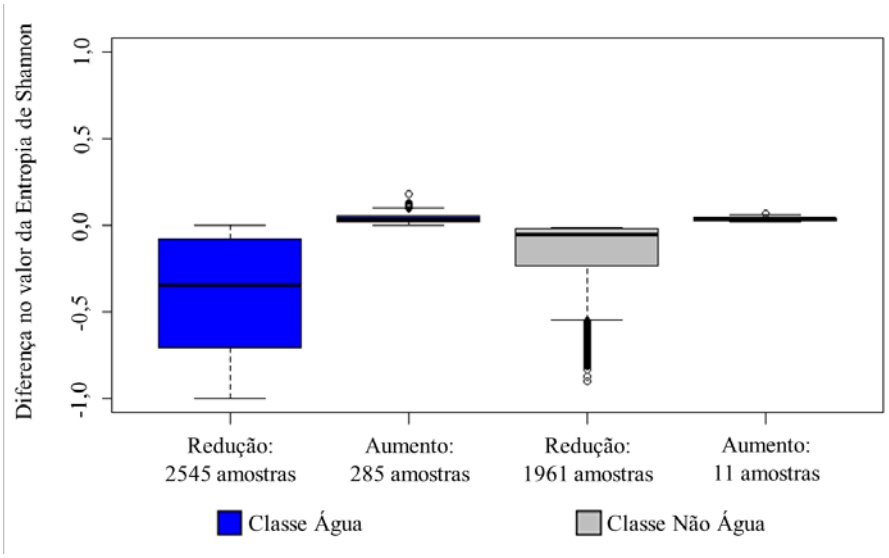

Figura 20 - Gráfico (*boxplot*) da amostragem aleatória para as classes em relação à diferença de entropia de Shannon.

CONCLUSÕES

A integração da incerteza para a classificação *Random Forest* com bandas e índices espectrais foi aplicada neste estudo para o mapeamento da água superficial permanente e temporária associada a um evento de inundação de grande magnitude em áreas com presença de diferentes tipologias de uso do solo e cobertura vegetal.

As variáveis propostas representadas pelas bandas espectrais e índices espectrais, a seleção de variáveis com o algoritmo *Recursive Feature Elimination* (RFE) e o método de amostragem considerando o mapa de incertezas derivado da entropia de Shannon, de acordo com as métricas de avaliação empregadas, apresentaram bom desempenho (exatidão total de 98,0%) para o mapeamento da água superficial.

Além disso, o RF permitiu identificar e filtrar as variáveis mais importantes para a classificação, entre as 21 variáveis preditoras, quatro delas foram suficientes para o processo de classificação apresentar alta acurácia. Em geral, a classificação RF resultou em baixos erros de comissão de água, evitando a superestimação da água.

A entropia de Shannon mostrou-se uma métrica adicional importante para avaliar o desempenho das classificações RF. A análise e avaliação das incertezas espacialmente distribuídas constituíram uma fonte de recurso para amostragem e uma métrica importante para garantir a confiabilidade do processo de classificação, principalmente por destacar elementos do terreno onde a classificação não foi satisfatória.

A coleta de amostras nestes pontos contribuiu para melhorar o desempenho da classificação e para reduzir as incertezas aumentando a confiança do mapeamento da água superficial, melhorando a representação espacial e a quantificação das classes de mapeamento. Em relação aos *pixels* classificados incorretamente como água em áreas urbanas, o aumento do número de variáveis e da amostragem poderia reduzir ou eliminar de forma mais efetiva estes falsos positivos.

Dessa forma, a classificação RF pode ser uma alternativa confiável de mapeamento em relação às classificações tradicionais baseadas apenas na limiarização de índices espectrais ou classificação baseada nas estatísticas das bandas espectrais. Enfatiza-se que o mapeamento da água superficial neste estudo utilizou dados espectrais detectados pelo sensor óptico, não classificando, portanto, a presença de água superficial abaixo do dossel florestal ou sob estruturas urbanas.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

ACHARYA, T.D.; SUBEDI, A.; LEE, D.H. Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. **Sensors**, v. 18, n. 8, p. 1–15, 2018b.

ACHARYA, T.D.; SUBEDI, A.; YANG, I.T.; LEE, D.H. Combining Water Indices for Water and Background Threshold in Landsat Image. **Proceedings**, v. 2, n. 3, p. 1–6, 2018a.

ALATORRE, L.C.; SÁNCHEZ-ANDRÉS, R.; CIRUJANO, S.; BEGUERÍA, S.; SÁNCHEZ-CARRILLO, S. Identification of Mangrove Areas by Remote Sensing: The ROC Curve Technique Applied to the Northwestern Mexico Coastal Zone Using Landsat Imagery. **Remote Sensing**, v. 3, n. 8, p. 1568–1583, 2011.

AUFDENKAMPE, A.K.; MAYORGA, E.; RAYMOND, P.A.; MELACK, J. M.; DONEY, S.C.; ALIN, S.R.; AALTO, R.E. Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. **Frontiers in Ecology and the Environment**, v. 9, p. 53–60, 2011.

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

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BRUBACHER, J.P.; GUASSELLI, L.A.; OLIVEIRA, G.G. Delimitação de áreas inundáveis a partir de Modified Normalized Difference Water Index (MNDWI) no Município de Esteio (RS, Brasil). **Pesquisas em Geociências (UFRGS)**, v. 44, n. 2, p. 367–376, 2017.

BULUT, O. Effective Feature Selection: Recursive Feature Elimination Using R. Disp. em: https://towardsdatascience.com/effective-feature-selection-recursive-feature-elimination-using-r-148ff998e4f7. Acesso em: dez. 2021.

CAMPOS, S.J.A.M.; STEFANI, F.L.; FACCINI, L.G.; BITAR, O.Y. Mapeamento de áreas sujeitas à inundação para planejamento e gestão territorial: cartas de suscetibilidade, perigo e risco. **Revista Brasileira de Geologia de Engenharia e Ambiental**, v. 5, n. 1, p. 67–81, 2015.

CONGALTON, R.G. A review of assessing the accuracy of classifications of remotely sensed data, **Remote Sensing of Environment**, v. 37, n. 1, p. 35–46, 1991.

CRIST, E.P. A TM Tasseled Cap equivalent transformation for reflectance factor data. **Remote Sensing of Environment**, v. 17, n. 3, p. 301–306, 1985.

DeVRIES, B.; HUANG, C.; LANG, M.W.; JONES, J.W.; HUANG, W.; CREED, I.F.; CARROLL, M.L. Automated quantification of surface water inundation in wetlands using optical satellite imagery. **Remote Sensing**, v. 9, n. 8, p. 1–22, 2017.

DONCHYTS, G.; BAART, F.; WINSEMIUS, H.; GORELICK, N.; KWADIJK, J.; GIESEN, N.V. Earth's surface water change over the past 30 years. **Nature Climate Change**, v. 6, p. 810–813, 2016.

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

FENG, Q.; GONG, J.; LIU, J.; LI, Y. Flood mapping based on multiple endmember spectral mixture analysis and Random Forest Classifier - The case of Yuyao, China. **Remote Sensing**, v. 7, n. 9, p. 12539–12562, 2015.

FEYISA, G.L.; MEILBY, H.; FENSHOLT, R.; PROUD, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. **Remote Sensing of Environment**, v. 140, p. 23–35, 2014.

FISHER, A.; FLOOD, N.; DANAHER, T. Comparing Landsat water index methods for automated water classification in eastern Australia. **Remote Sensing of Environment**, v. 175, p. 167–182, 2016.

FRANCI, F.; BOCCARDO, P.; MANDANICI, E.; ROVERI. E.; BITELLI, G. Flood mapping using VHR satellite imagery: a comparison between different classification approaches. Edinburgh, 2016. In: **Earth Resources and Environmental Remote Sensing/GIS Applications VII**, p. 1-9, 2016.

GIUSTARINI, L.; CHINI, M.; HOSTACHE, R.; PAPPENBERGER, F. Flood hazard mapping combining hydrodynamic modeling and multi annual remote sensing data. **Remote Sensing**, v. 7, n. 10, p. 14200–14226, 2015.

GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18–27, 2017.

GOUDIE, A.S. **Encyclopedia of geomorphology**. International Association of Geomorphologists. 1156 p., 2004.

GUYON, I.; WESTON, J; BARNHILL, S; VAPNIK, V. Gene Selection for Cancer Classification Using Support Vector Machines. **Machine Learning,** v. 46, n. 1, p. 389–422, 2002.

HIRABAYASHI, Y.; MAHENDRAN, R.; KOIRALA, S.; KONOSHIMA, L.; YAMAZAKI, D. Global flood risk under climate change. **Nature Climate Change**, v. 3. p. 816–821, 2013.

HUANG, C.; CHEN, Y.; ZHANG, S.; WU, J. Detecting, extracting, and monitoring surface water from space using optical sensors: A review. **Reviews of Geophysics**, v. 56, n. 2, p. 333–360, 2018.

JI, L.; ZHANG, L.; WYLIE, B. Analysis of Dynamic Thresholds for the Normalized Difference Water Index. **Photogrammetric Engineering & Remote Sensing**, v. 75, n. 11, p. 1307–1317, 2009.

JIANG, H.; FENG, M.; ZHU, Y.; LU, N.; HUANG, J.; XIAO, T. An Automated Method for Extracting Rivers and Lakes from Landsat Imagery. **Remote Sensing,** v. 6, n. 6, p. 5067–5089, 2014.

JONES, J.W. Efficient wetland surface water detection and monitoring via Landsat: Comparison with in situ Data from the Everglades Depth Estimation Network. **Remote Sensing**, v. 7, n. 9, p. 12503–12538, 2015.

JOYCE, K.E.; BELLISS, S.E.; SAMSONOV, S.V.; MCNEILL, S.J.; GLASSEY, P.J. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. **Progress in Physical Geography: Earth and Environment,** v. 33, n. 2, p. 183–207, 2009.

KHAN, S.I.; HONG, Y.; WANG, J.; YILMAZ, K.K. Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria Basin: Implications for hydrologic prediction in Ungauged Basins. **IEEE Transactions on Geoscience and Remote Sensing**, v. 49, n.1, p.85–95, 2011.

KO, B.C.; KIM, H.H.; NAM, J.Y. Classification of potential water bodies using Landsat 8 OLI and a combination of two boosted Random Forest Classifiers, **Sensors**, v.15, n. 6, p. 13763–13777, 2015.

KOKO, A.F.; YUE, W.; ABUBAKAR, G.A.; HAMED, R.; ALABSI, A.A.N. Analyzing urban growth and land cover change scenario in Lagos, Nigeria using multi-temporal remote sensing data and GIS to mitigate flooding. **Geomatics, Natural Hazards and Risk**, v. 12, n. 1, p. 631-652, 2021.

KORDELAS, G.A.; MANAKOS, I.; LEFEBVRE, G.; POULIN, B. Automatic Inundation Mapping Using Sentinel-2 Data Applicable to Both Camargue and Doñana Biosphere Reserves. **Remote Sensing**, v. 11, n. 19, p. 1–20, 2019.

KUHN, M. **Package 'caret':** Classification and regression training. Disp. em: https://CRAN.R-project.org/package=caret. Acesso em: fev. 2022.

KUHN, M. & JOHNSON, K. **Feature Engineering and Selection**: A Practical Approach for Predictive Models. 2019. Disp. em: http://www.feat.engineering/

LEFEBVRE, G.; DAVRANCHE, A.; WILLM, L.; CAMPAGNA, J.; REDMOND, L.; MERLE, C.; GUELMAMI, A.; POULIN, B. Introducing WIW for Detecting the Presence of Water in Wetlands with Landsat and Sentinel Satellites. **Remote Sensing**, v. 11, n. 19, p. 1–18, 2019.

LI, W.; DU, Z.; LING, F.; ZHOU, D.; WANG, H.; GUI, Y.; SUN, B.; ZHANG, X. A Comparison of Land Surface Water Mapping Using the Normalized Difference Water Index from TM, ETM+ and ALI. **Remote Sensing**, v. 5, n. 11, p. 5530–5549, 2013.

LIAW, A. & WIENER, M. Classification and Regression by randomForest. **R News,** v. 2, n. 3, p. 18–22, 2002. Disp. em: https://CRAN.R-project.org/doc/Rnews/.

LIAW, A. & WIENER, M. **Package 'randomForest'**. Disp. em: https://cran.r-project.org/web/packages/randomForest/index.html. Acesso em: fev. 2022.

LIMA, D.L.C. & RENNÓ, C.D. Mapeamento de áreas alagáveis na bacia Amazônica a partir de dados extraídos do MDE-SRTM e avaliação da incerteza por meio da entropia de Shannon. In: XXIV SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 2021, Belo Horizonte. **Anais ...** Porto Alegre: Asso-ciação Brasileira de Recursos Hídricos – ABRHidro, 2021, p. 1–10.

MARTINS, V.S.; KALEITA, A.; BARBOSA, C.C.F.; FASSONI-ANDRADE, A.C.; LOBO, F. L.; NOVO, E.M.L.M. Remote sensing of large reservoir in the drought years: Implications on surface water change and turbidity variability of Sobradinho reservoir (Northeast Brazil). **Remote Sensing Applications: Society and Environment**, v. 13, p. 275–288, 2019.

McFEETERS, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n. 7, p. 1425–1432, 1996.

MOURA, M.P.; RIBEIRO NETO, A.; COSTA, F.A. Application of satellite imagery to update depth-area-volume relationships in reservoirs in the semiarid region of Northeast Brazil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 26, n. 1, p. 44–50, 2022.

NAMIKAWA, L.M.; KÖRTING, T.S.; CASTEJON, E.F. Water Body Extraction from Rapideye Images: An Automated Methodology Based on Hue Component of Color Transformation from RGB to HSV Model. **Brazilian Journal of Cartography**, v. 6, n. 68, p. 1097–1111, 2016.

NEIFF, J.J. El régimen de pulsos en ríos y grandes humedales de Sudamérica. En: A.I. Malvárez y P. Kandus (Eds.), **Tópicos sobre grandes humedales sudamericanos,** ORCYT-MAB (UNESCO), Montevideo, p. 1–49, 1999.

OGDEN, R.; REID, M.; THOMS, M. Soil fertility in a large dryland floodplain: Patterns, processes and the implications of water resource development. **Catena**, 70, n. 2, p. 114–126, 2007.

PEKEL, J.F.; COTTAM, A.; GORELICK, N.; BELWARD, A.S. High-resolution mapping of global surface water and its long-term changes. **Nature**, v. 540, p. 418–422, 2016.

QGIS Development Team. **QGIS Geographic Information System**. Open-Source Geospatial Foundation Project, 2021. Disp. em: http://qgis.osgeo.org.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2021. Disp. em: https://www.R-project.org/.

RAHMAN, S. & DI, L. The state of the art of spaceborne remote sensing in flood management. **Natural Hazards**, v. 85, p. 1223–1248, 2017.

RStudio Team. **RStudio: Integrated Development for R**. RStudio, PBC, Boston, MA. 2021. Disp. em: http://www.rstudio.com/.

SHANNON, C.E. A Mathematical Theory of Communication. **Bell System Technical Journal**, 1948.

SHEN, L. & LI, C. Water body extraction from Landsat ETM+ imagery using adaboost algorithm. **18th International Conference on Geoinformatics**, p. 1–4, 2010.

SILVEIRA, G.V.; GUASSELLI, L.A. Mapeamento das Inundações a partir de NDWI no Município de Itaqui, Rio Uruguai – RS. **Geociências**, UNESP, v. 38, n. 2, p. 521–534, 2019.

TENG, J. & JAKEMAN, A.J.; VAZE, J.; CROKE, B.F.W.; DUTTA, D.; KIM, S. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. **Environmental Modelling & Software**, v. 90, p. 201–216, 2017.

TOTARO, V.; PESCHECHERA, G.; GIOIA, A.; IACOBELLIS, V.; FRATINO, U. Comparison of Satellite and Geomorphic Indices for Flooded Areas Detection in a Mediterranean River Basin. In: **Computational Science and Its Applications – ICCSA 2019**. Lecture Notes in Computer Science, Springer, Cham, v. 11622, p. 173–185, 2019.

TULBURE, M.G.; BROICH, M.; STEHMAN, S.V.; KOMMAREDDY. A Surface water extent dynamic from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-arid region. **Remote Sensing of Environment**, v. 178, n. 1, p. 142–157, 2016.

TYRALIS, H.; PAPACHARALAMPOUS, G.; LANGOUSIS, A. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. **Water**, **v.** 11, n. 5, p. 1–37, 2019.

USGS – United States Geological Survey. Earth Resources Observation and Science (EROS) Center. Disp. em: https://www.usgs.gov/centers/eros/. Acesso em: jan. 2022.

VERMOTE, E.; ROGER, J.C.; FRANCH, B.; SKAKUN, S. LaSRC (Land Surface Reflectance Code): Overview, application and validation using MODIS, VIIRS, LANDSAT and Sentinel 2 data. In: **IEEE International Geoscience and Remote Sensing Symposium** - IGARSS 2018, p. 8173–8176, 2018.

WANG, X.; XIE, S.; ZHANG, X.; CHEN, C.; GUO, H.; DU, J; DUAN, Z. A robust Multi-Band Water Index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 68, p. 73–91, 2018b.

WANG, Z.; LAI, C.; CHEN, X.; YANG, B.; ZHAO, S.; BAI, X. Flood hazard risk assessment model based on random forest. **Journal of Hydrology**, v. 527, p. 1130–1141, 2015.

WANG, Z.; LI, H.; CAI, X. Remotely Sensed Analysis of Channel Bar Morphodynamics in the Middle Yangtze River in Response to a Major Monsoon Flood in 2002. **Remote Sensing**, v. 10, n. 8, p. 1–14, 2018a.

XIE, H.; LUO, X.; XU, X.; PAN, H.; TONG, X. Evaluation of Landsat 8 OLI imagery for unsupervised inland water extraction. **International Journal of Remote Sensing**, v. 37, n. 8, p. 1826–1844, 2016.

XU, H. Modification of Normalised Difference Water Index (NDWI) to enhance open water features in remotely sensed imagery. **International Journal of Remote Sensing**, v. 27, n. 14, p. 3025–3033, 2006.

YAGMUR, N.; MUSAOGLU, N.; TASKIN, G. Detection of Shallow Water Area with Machine Learning Algorithms. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XLII-2/W13, p. 1269–1273, 2019.

YOUNG, N.E.; ANDERSON, R.S.; CHIGNELL, S.M.; VORSTER, A.G.; LAWRENCE, R.; EVANGELISTA, P.H. A survival guide to Landsat preprocessing. **Concepts & Synthesis**, v. 98, n. 4, p. 920–932, 2017.

ZHOU, Y.; DONG, J.; XIAO, X.; XIAO, T.; YANG, Z.; ZHAO, G.; ZOU, Z.; QIN, Y. Open Surface Water Mapping Algorithms: A Comparison of Water-Related Spectral Indices and Sensors. **Water**, v. 9, n. 4, p. 1–16, 2017.

Submetido em 25 de maio de 2022
Aceito para publicação em 9 de março de 2023