

SENSORIAMENTO REMOTO E APRENDIZADO DE MÁQUINA APLICADOS NO MAPEAMENTO HÍDRICO DO SISTEMA CANTAREIRA

REMOTE SENSING AND MACHINE LEARNING APPLIED IN THE WATER MAPPING OF THE CANTAREIRA SYSTEM

 Lucas Oliveira ^A

 Lucas Prado Osco ^B

 José Marcato Junior ^A

 Ana Paula Marques Ramos ^B

 Mauricio de Souza ^A

^A Universidade Federal do Mato Grosso do Sul (UFMS), Campo Grande, MS, Brasil

^B Universidade do Oeste Paulista (UNOESTE), Presidente Prudente, SP, Brasil

Recebido em: 08/11/2021 | 06/09/2022 DOI: 10.12957/tamoios.2023.63354

Correspondência para: Lucas Oliveira (lucas.oliveira@ufms.br)

Resumo

O monitoramento de recursos hídricos serve como base para tomada de decisão e para amenizar os efeitos de futuras crises hídricas, como, por exemplo, a crise no Sistema Cantareira, área de estudo deste trabalho, no biênio 2013/14. Investigamos a confiabilidade da classificação de imagens orbitais, utilizando técnicas de sensoriamento remoto e aprendizado de máquina no contexto de recursos hídricos, que é um recurso indispensável para a sociedade. Os experimentos foram realizados nas seis represas que compõem o Sistema Cantareira, e utilizamos imagens multiespectrais do satélite RapidEye, que apresenta uma resolução espacial de 5 metros. Foram testados quatro métodos de classificação, sendo eles: Distância Mínima, Máxima Verossimilhança, Mapeamento do Ângulo Espectral e Random Forest. Os métodos Distância Mínima, Máxima Verossimilhança e Random Forest atingiram uma exatidão superior a 95%, sendo Random Forest, o único método de aprendizado de máquina, com a maior exatidão (98,06%). Os resultados mostram que a combinação de imagens RapidEye com as técnicas clássicas e de aprendizado de máquina permitiram o mapeamento detalhado e exato de recursos hídricos no Sistema Cantareira, podendo ser replicados em outras áreas de estudo.

Palavras-chave: Sistema Cantareira; Classificação Semi-Automática; RapidEye; Random Forest;

Abstract

The monitoring of water resources serves as a basis for decision-making and to mitigate the effects of future water crises, such as the crisis in the Cantareira System, the study area of this work, in the 2013/14 biennium. We investigate the reliability of image classification using remote sensing and machine learning techniques in the context of water resources, which is an indispensable resource for society. The experiments were carried out in the six dams that make up the Cantareira System, and we used RapidEye orbital multispectral images, which have a spatial resolution of 5 meters. Four classification methods were tested, namely: Minimum Distance, Maximum Likelihood, Spectral Angle Mapping and Random Forest. The Minimum Distance, Maximum Likelihood and Random Forest methods reached an accuracy greater than 95%, with Random Forest, the only machine learning method, with the highest accuracy (98.06%). The results show that the combination of RapidEye images with classical techniques and machine learning allowed the detailed and accurate mapping of water resources in the Cantareira System, which could be replicated in other study areas.





Keywords: Cantareira System; Semi-Automatic Classification; RapidEye; Random Forest;.

INTRODUÇÃO

Em 2014 houve o início do que foi considerado a pior crise hídrica da história do Brasil. Tal crise durou até o ano de 2016 e afetou especialmente o estado de São Paulo, que é o mais populoso do Brasil com aproximadamente 45 milhões de habitantes, segundo a Fundação SEADE (Portal de Estatísticas do Estado de São Paulo, 2022). O município de São Paulo está localizado na Bacia Hidrográfica do Paraná, que sofreu com alterações no regime de chuvas durante o período. Oliveira et. al. (2019) utilizaram técnicas de sensoriamento remoto com o intuito de analisar o regime de chuvas na bacia, associado ao armazenamento hídrico, entre janeiro de 2002 e dezembro de 2014, concluindo que o nível de chuva foi menor durante o período interferindo de forma significativa no armazenamento hídrico.

Uma das principais evidências do déficit na questão do armazenamento hídrico é encontrada no Sistema Cantareira, principal sistema de abastecimento da região metropolitana de São Paulo, que foi um dos mais afetados. Administrado pela SABESP (Companhia de Saneamento Básico do Estado de São Paulo), estima-se, de acordo com a ANA (Agência Nacional de Águas e Saneamento Básico, 2019), que ele é utilizado para abastecer aproximadamente 46% da população da Região Metropolitana de São Paulo. Durante o período de crise, o sistema chegou a operar em “nível morto”, ou seja, captação de água abaixo do nível das comportas, e houve até indícios de que ele poderia secar. Portanto, é importante a investigação de métodos para monitoramento das represas que compõem esse sistema de abastecimento. O sensoriamento remoto apresenta-se como uma alternativa, pois permite o monitoramento de áreas minimizando a exigência por trabalho de campo.

Uma opção que se mostra viável para uma análise espacial é a classificação de imagens, que pode ser realizada por meio do sensoriamento remoto, permitindo a utilização de técnicas clássicas, e por *machine learning*, utilizando técnicas mais complexas. Monard & Baranauskas (2003) define *machine learning* (aprendizado de máquina) como um sistema capaz de adquirir conhecimento de forma automática, tomando decisões baseadas no conhecimento obtido em resoluções bem sucedidas de problemas anteriores. Pereira & Guimarães (2018) utilizaram os métodos *Spectral Angle Mapper* (SAM), *Minimum Distance* (MD) e *Maximum Likelihood* (ML), técnicas clássicas na classificação de imagens, para mapear o uso da terra e concluíram que o processo apresenta resultados adequados, mesmo trabalhando em uma escala detalhada e com alta especificidade de classes. Leão Junior et. al (2018) compararam a classificação de imagem utilizando um método mais complexo (*Support Vector Machine* -SVM), que utiliza uma abordagem por pixel, com um método de abordagem por meio de árvore de decisão (OBIA-AD), onde as características espectrais, espaciais e contextuais são consideradas no agrupamento dos pixels. Através da segmentação da imagem, pixels homogêneos e semelhantes eram considerados como sendo o mesmo objeto. Verificou-se que o SVM apresenta uma acurácia geral maior que o OBIA-AD.

Noi & Kappas (2018) compararam *Random Forest* (RF), *K-Nearest Neighbors* (KNN) e SVM em uma classificação de cobertura do solo. Todos os três algoritmos apresentaram resultados confiáveis, com o SVM sendo um pouco superior com acurácia geral de 95,32%, enquanto o KNN apresentou acurácia geral de 94,59% e o RF 94,70%. Os autores também



destacam que o tamanho das amostras pode afetar na confiabilidade da classificação, pois uma amostra maior pode apresentar dados desbalanceados. Em uma comparação do RF com o SVM, Pal (2005) destacou que a confiabilidade de classificação dos dois algoritmos é semelhante, mas que o RF necessita apenas de dois parâmetros para atingir melhor exatidão, enquanto o SVM necessita de vários outros parâmetros. Outro diferencial destacado pelo autor é que o RF é capaz de lidar melhor com dados desbalanceados e com ausência de valores, enquanto o SVM não.

Como opção de dados de entrada (insumos) para as técnicas citadas anteriormente, foram obtidas no site da Planet as imagens orbitais RapidEye, que se apresentam como uma fonte de dados adequada para as mais diversas finalidades, como monitoramento de áreas agrícolas, pastagens, áreas florestais e para monitoramento ambiental de forma geral (ANTUNES & SIQUEIRA, 2013). A missão RapidEye é composta por cinco satélites, equipados com o mesmo sensor REIS (RapidEye Earth Imaging System), com uma altitude orbital de 630km. Sua faixa de abrangência é de 77km com 1500km de extensão. Possui uma resolução temporal diária e resolução espacial de 6,5m (nadir) e 5m (imagem ortorretificada), com uma resolução radiométrica de 12bits. Adam et. al (2014) propuseram avaliar o desempenho das imagens RapidEye utilizando dois métodos de classificação de imagem, o RF e o SVM, em uma área de estudo costeira com paisagem heterogênea. Foram criadas onze classes e constataram que as imagens forneceram um resultado altamente preciso em ambos os casos, acima dos 93% de precisão.

Considerando a finalidade e a área desse estudo, as imagens RapidEye se apresentam como a melhor opção dentre os dados orbitais disponíveis. Se comparada às imagens Landsat 8 e outras imagens disponibilizadas gratuitamente (Sentinel, CBERS, etc), as imagens RapidEye apresentam maior resolução espacial, e também maior resolução temporal, possibilitando a visita diária, possibilitando um monitoramento mais detalhado e diário.

Sendo assim, o objetivo desse trabalho é avaliar se as técnicas clássicas de sensoriamento remoto e *machine learning* são capazes de classificar com exatidão as imagens RapidEye, permitindo a análise de recursos hídricos no contexto do Sistema Cantareira.

METODOLOGIA

Área de Estudo

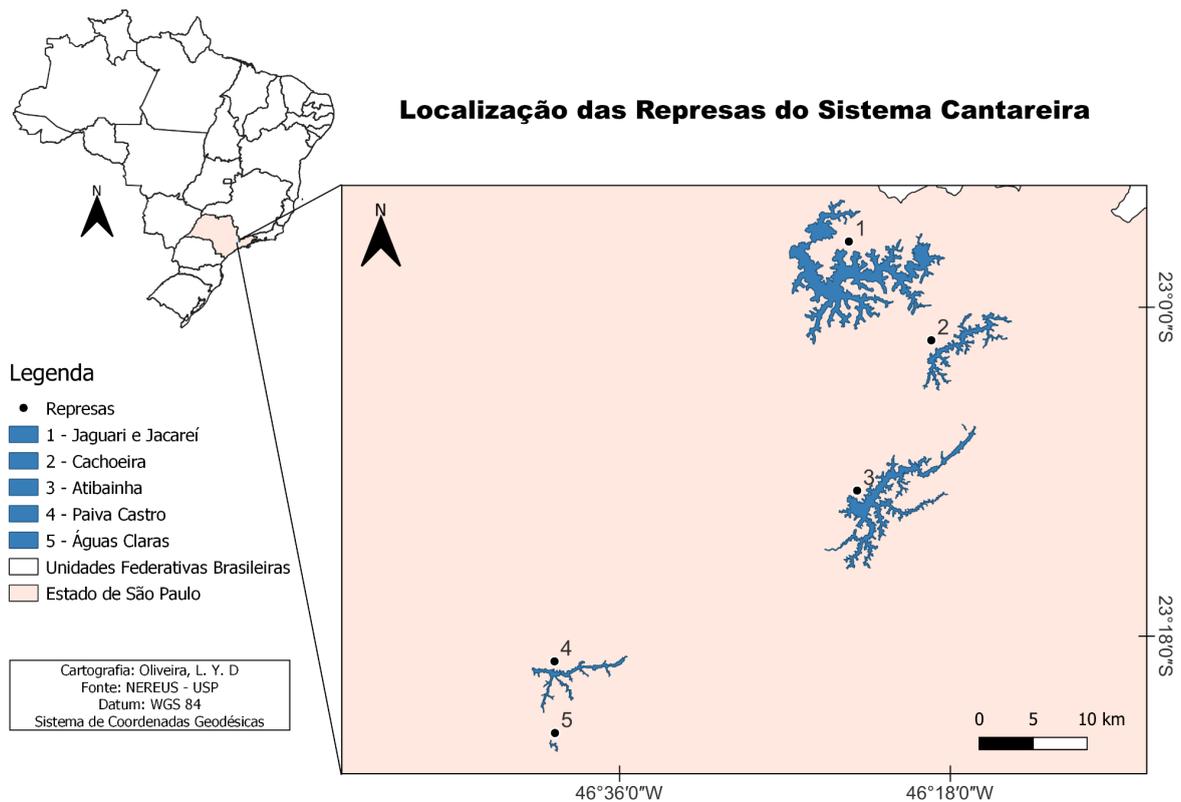
A área de estudo deste trabalho foi o Sistema Cantareira, localizado no estado de São Paulo, Brasil. É o maior sistema sob administração da SABESP (Companhia de Saneamento Básico do Estado de São Paulo), sendo o responsável pelo abastecimento de aproximadamente 9 milhões de pessoas, e um dos maiores sistemas produtores de água do mundo. O sistema está presente nos municípios de Bragança Paulista, Piracaia, Vargem, Joanópolis, Nazaré Paulista, Franco da Rocha, Mairiporã, Caieiras e São Paulo. Sua estrutura é composta por seis represas, em diferentes níveis (Figura 1), sendo elas Jaguarí, Jacareí, Paiva Castro, Águas Claras, Cachoeira e Atibainha.

De acordo com a SABESP (2020), o sistema armazena um total de 1,2 trilhão de litros de água, sendo 982 bilhões de litros acima do nível das comportas, o chamado volume útil, e



o restante abaixo das comportas (reserva técnica). Ainda de acordo com a SABESP, simulações indicavam que haveria um esgotamento do volume útil do sistema, durante o período da crise hídrica (2013 e 2014), e que foram feitas reformas, como a construção emergencial de barragens, canais, tubulações e 17 bombas flutuantes, que tornaram viável a utilização da reserva técnica.

Figura 1 - Mapa de localização das represas do Sistema Cantareira



Fonte: Elaborado pelo Autor

Imagens RapidEye

RapidEye é um programa de observação da Terra composto por cinco satélites que tendo o primeiro satélite sido lançado em 2008, possui imagens disponíveis desde agosto do ano de lançamento, e no atual momento ainda se encontra em operação. Suas imagens são ortorretificadas, corrigidas e com exatidão na localização geográfica compatível com sua resolução espacial, conforme demonstrado por Marcato Junior et. al (2015), onde a qualidade geométrica das imagens foi testada a partir de ortofotos e pontos de controle levantados por receptores GNSS (Global Navigation Satellite System).



Sua altitude orbital é de 630km, em uma órbita sincronizada com o sol. Seu sensor é multiespectral, capaz de capturar as bandas vermelha, verde, azul, red edge e NIR (infravermelho-próximo), com resolução espacial de 5m em uma imagem ortorretificada. Sua faixa de abrangência na coleta de imagens é de 77 km de largura, com 1500 km de comprimento. Devido sua capacidade de revisitar o mesmo ponto diariamente, consegue capturar uma área de 5 milhões de km² por dia, sendo desta forma um bom produto em monitoramento de agricultura, florestas, recursos hídricos, usos cartográficos, entre outros. Conforme demonstrado por Antunes & Siqueira (2013), as imagens ortorretificadas do RapidEye possuem qualidade geométrica para mapeamento de superfície.

Utilizando os dados diários de volume e nível, disponibilizados no site da SABESP (<https://mananciais.sabesp.com.br/HistoricoSistemas?SistemaId=0>) desde 2000, selecionamos imagens tanto no período de seca quanto no período de cheia, conforme mostrado no Quadro 1.

Quadro 1 - Datas das imagens RapidEye

Águas Claras	Atibainha	Cachoeira	Jaguari/Jacareí	Paiva Castro
20/03/2010 (cheia)	16/05/2010 (cheia)	05/02/2010 (cheia)	16/04/2010 (cheia)	28/01/2011 (cheia)
30/06/2015 (seca)	10/10/2014 (seca)	16/11/2014 (seca)	10/10/2014 (seca)	08/10/2014 (seca)
-	07/04/2016 (cheia)	-	-	19/03/2016 (cheia)

Fonte: Autor/SABESP

As imagens RapidEye foram obtidas gratuitamente no site da Planet (planet.com). A escolha de imagens da década passada se deu pelo conhecimento de períodos contrastantes na questão hídrica no estado de São Paulo. Enquanto o estado apresentou um forte período chuvoso durante o biênio 2010/11, no biênio 2013/14 o estado apresentou uma das piores crises hídricas de sua história.

Técnicas de *Machine Learning*

Para os algoritmos de aprendizado de máquina utilizamos o plugin Dzetsaka, complemento do software QGIS (<https://qgis.org/>). Programado inicialmente para treinar amostras utilizando o algoritmo Gaussian Mixture Model, o plugin também permite realizar o treinamento utilizando os algoritmos Random Forest (RF), k-Nearest Neighbors (KNN) e Support Vector Machine (SVM) após a instalação de um pacote de dados adicionais (<https://github.com/nkarasiak/dzetsaka>).



Tian et. al (2016) utilizaram RF, comparando com SVM e Artificial Neural Networks (ANN), para classificação de áreas úmidas em Xinjiang, China. A base de dados orbitais foram imagens Landsat-8 e o estudo constatou que o RF apresentou uma superioridade de até 10% em relação aos outros dois métodos, com uma exatidão de 93% e índice Kappa de 0,92. Em uma comparação do RF com o SVM, Pal (2005) destacou que a confiabilidade de classificação dos dois algoritmos é semelhante, mas que o RF necessita apenas de dois parâmetros para atingir melhores acurácias, enquanto o SVM necessita de vários outros parâmetros. Outro diferencial destacado pelo autor é que o RF é capaz de lidar melhor com dados desbalanceados e com ausência de valores, enquanto o SVM não.

A escolha do uso do RF se deu devido a simplicidade que o mesmo apresenta quando comparado a outros métodos, além de possuir uma alta confiabilidade em seus resultados.

Técnicas Clássicas de Sensoriamento Remoto

Utilizamos o complemento (plugin) Semi-Automatic Classification (SCP), do QGIS, que trabalha com técnicas tradicionais de sensoriamento remoto. O plugin traz três algoritmos de classificação: Distância Mínima (MD), Máxima Verossimilhança (ML) e Mapeamento do Ângulo Espectral (SAM).

Congedo (2016) explica os três métodos clássicos de classificação, onde o algoritmo de MD calcula a distância euclidiana entre os pixels da imagem e os pixels da amostra. Ele calcula para cada pixel da imagem e atribui à classe espectral mais próxima daquilo para o pixel, de acordo com as amostras coletadas. Já o ML utiliza o Teorema de Bayes para determinar a qual classe o pixel pertence. Por fim, temos o SAM que calcula o ângulo espectral entre os pixels da imagem e os das amostras, considerando os pixels como vetores.

Processamento dos Dados

Tanto no plugin SCP quanto no DZetsaka o procedimento de treinamento e teste foi feito utilizando os mesmos conjuntos. Primeiro, as amostras foram coletadas para cada uma das imagens, sendo que elas foram divididas em duas classes: Água e Não Água. As represas em todas as imagens foram delimitadas manualmente, gerando assim o banco de dados rotulado. O algoritmo de classificação foi escolhido e, por fim, obtivemos a classificação com a respectiva matriz de confusão do treinamento.

Após esse treinamento foi feito o teste utilizando imagens de diferentes datas. Para o teste do modelo ajustado com as amostras de 2015 utilizamos amostras da imagem de 2010, inversamente utilizamos as amostras de 2015 para testar o modelo de 2010. Com isso, verificamos a capacidade de generalização do modelo treinado com imagem de apenas um ano. No final também foi gerado uma imagem com a classificação e a matriz de confusão, que permitiu calcular a acurácia geral da classificação e o valor do índice Kappa.

Coefficiente de Kappa (Teste de Concordância) e acurácia geral



Silva e Paes (2012) definem o modelo proposto por Jacob Cohen em 1960 como uma medida de concordância inter-observador (trabalhado por dois observadores diferentes, utilizando o mesmo método) que permite avaliar se a concordância está além do esperado tão somente pelo acaso, quanto o grau dessa concordância.

Quanto mais próximo de zero menor seu nível de concordância, tendo o valor máximo unitário como ápice. Landis e Koch (1977) sugerem a seguinte classificação, conforme demonstrado no Quadro 2. Além do valor do índice Kappa, a matriz de confusão, que é um relatório indicando o comportamento do modelo (erros e acertos), permite também sabermos o valor da exatidão geral, ou seja, a performance e a confiabilidade do modelo. O valor da exatidão varia entre 0 e 100%, sendo quanto mais próximo de 100% mais confiável e exato.

Quadro 2 - Classificação do valor Kappa

Valor Kappa	Nível de Concordância
< 0	Ausente
0.00 - 0.20	Leve
0.21 - 0.40	Razoável
0.41 - 0.60	Moderado
0.61 - 0.80	Substancial
0.81 - 1.00	Quase Perfeito

Fonte: Landis & Koch (1977).

RESULTADOS

Tanto os métodos clássicos de sensoriamento remoto, quanto o método de aprendizado de máquina, apresentaram desempenho satisfatório e resultados confiáveis. Em todos os quatro métodos testados obtivemos uma exatidão geral de, no mínimo, 88%, e valor Kappa de no mínimo 0,68, determinando um nível de concordância dos modelos como Substancial ou Quase Perfeito, conforme apresentado na Tabela 1.

Tabela 1 - Valor Kappa e acurácia da classificação. (MD = Distância Mínima, ML = Máxima Verossimilhança, SAM = Mapeamento do Ângulo Espectral, RF = Random Forest, AG = Acuraria Geral e K = Coeficiente Kappa)

Represa/Ano	MD		ML		SAM		RF	
	AG %	K	AG%	K	AG%	K	AG%	K
Águas Claras - 2010	98,03	0,85	99,95	0,99	99,09	0,92	100,00	1,00
Águas Claras - 2015	98,35	0,88	100,00	1,00	99,88	0,99	99,93	0,99



Atibainha - 2010	94,49	0,84	99,39	0,95	85,40	0,62	99,97	1,00
Atibainha - 2014	99,23	0,98	99,88	0,99	87,04	0,66	99,96	1,00
Atibainha - 2016	98,70	0,96	99,99	1,00	88,96	0,67	99,99	1,00
Cachoeira - 2010	86,90	0,54	86,91	0,51	86,90	0,53	86,91	0,65
Cachoeira - 2014	95,80	0,77	95,71	0,72	95,99	0,77	96,17	0,79
Jaguari e Jacareí - 2010	89,30	0,75	95,40	0,84	82,71	0,64	97,57	0,94
Jaguari e Jacareí - 2014	91,22	0,75	97,15	0,87	84,97	0,55	97,79	0,93
Paiva Castro - 2011	99,91	1,00	99,77	0,97	82,19	0,56	100,00	1,00
Paiva Castro - 2014	99,04	0,97	99,99	1,00	84,74	0,59	99,99	1,00
Paiva Castro - 2016	99,93	1,00	99,98	1,00	83,03	0,60	98,48	0,96
Média	95,91	0,86	97,84	0,90	88,41	0,68	98,06	0,94

Fonte: Elaborado por Lucas Oliveira

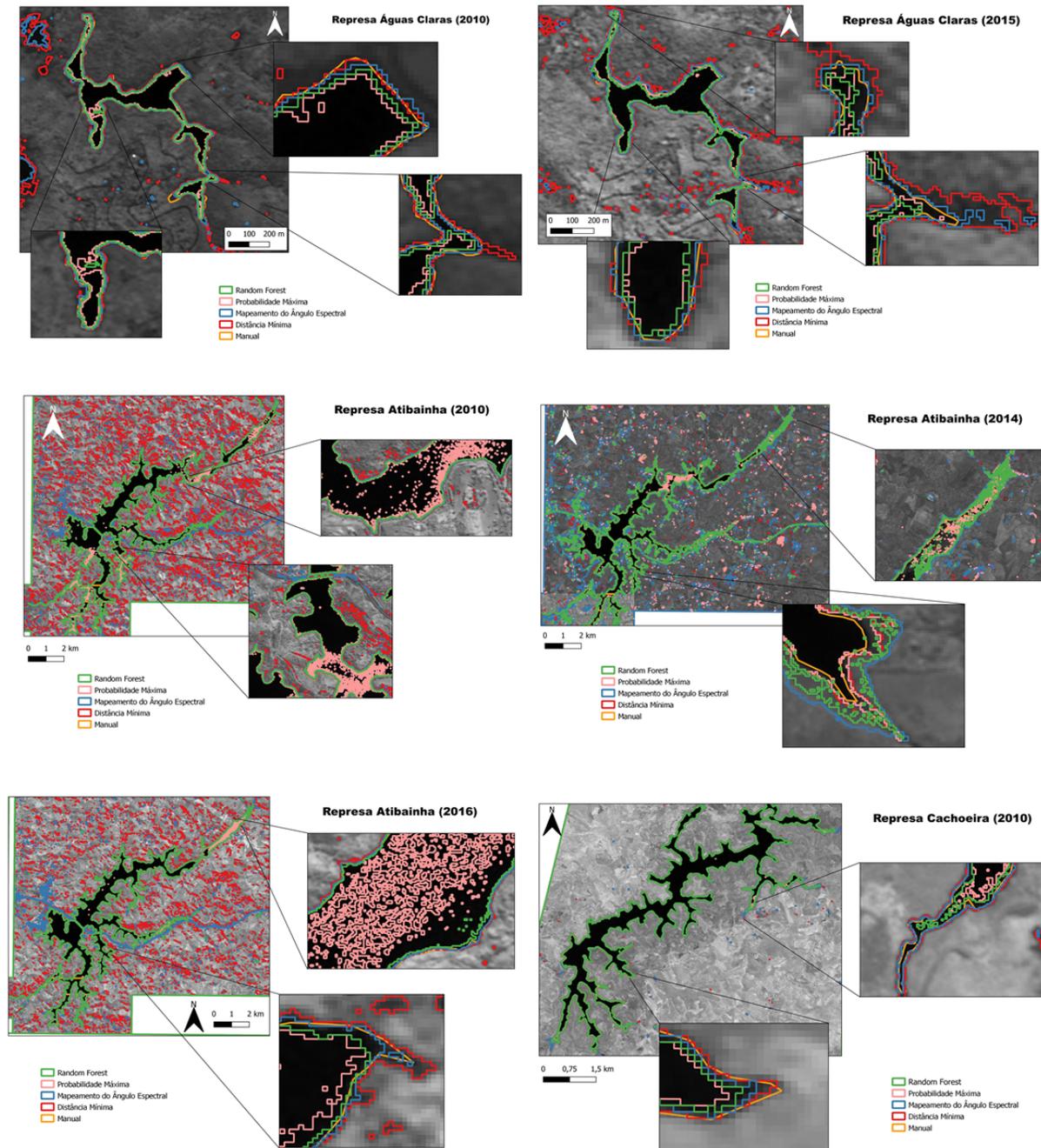
Da classificação feita pelo SCP, utilizando técnicas clássicas, o método de Mapeamento de Ângulo Espectral apresentou os resultados inferiores, com uma diferença significativa em relação aos outros métodos. Sua exatidão geral média ficou em 88,41% e o valor Kappa médio em 0,68, considerado substancial na classificação de Landis e Koch (1977). O método de Distância Mínima teve um bom desempenho, apresentando uma exatidão geral média de 95,91% e valor Kappa de 0,86, sendo classificado como quase perfeito, mas ainda assim não teve o melhor desempenho dentre as técnicas básicas.

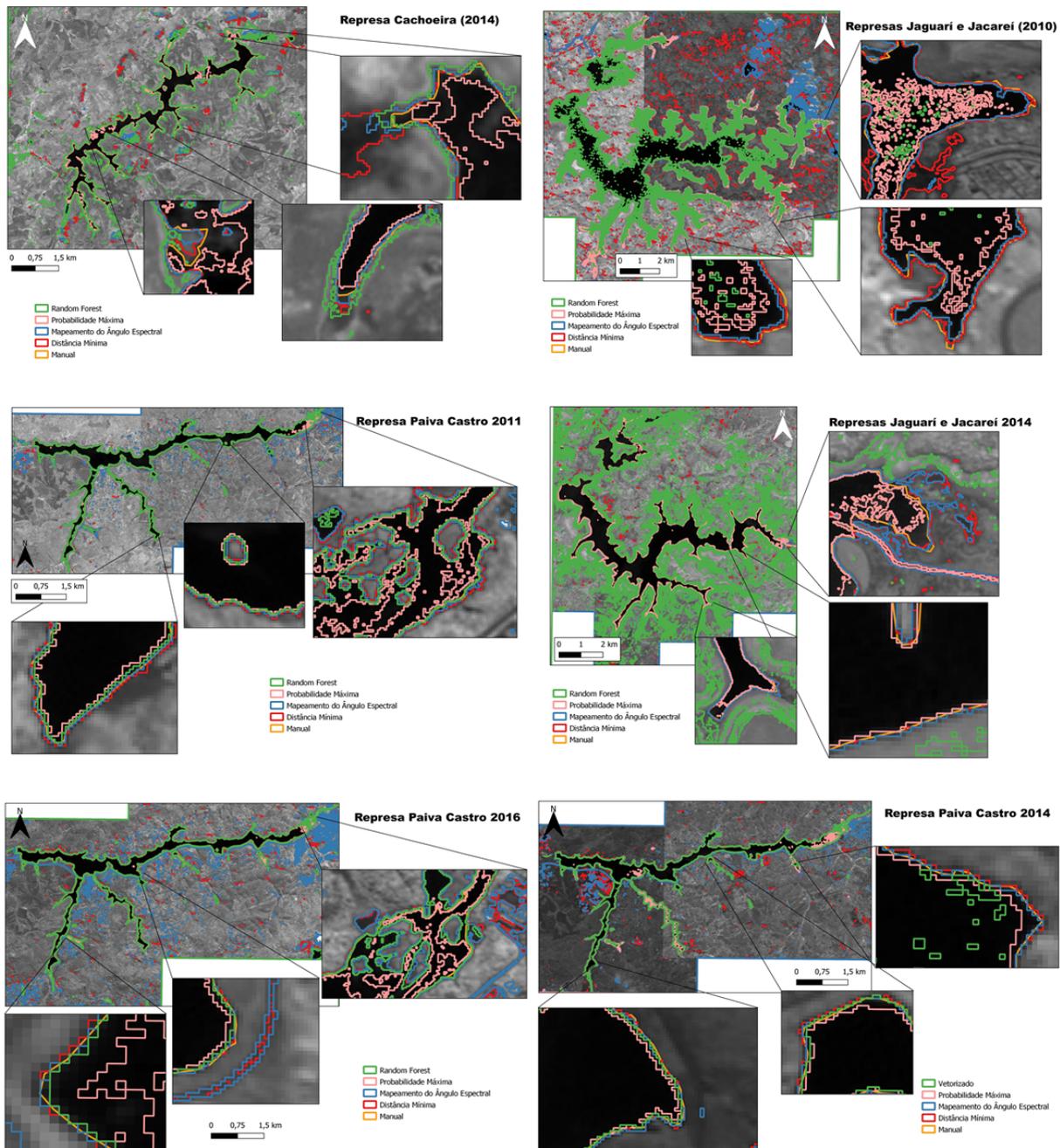
Pereira & Guimarães (2018), em seu estudo testando o SCP para uso e ocupação de solo, sugeriu o uso do método de Máxima Verossimilhança, pois foi o método com resultados superiores nos testes realizados, apresentando exatidão quase perfeita. O comportamento se mostrou idêntico em nosso estudo, com o método de Máxima Verossimilhança apresentando os melhores resultados (exatidão de 97,84% e valor Kappa de 0,90) entre os algoritmos tradicionais, próximos aos valores observados pelo método de *machine learning*.

O RF apresentou os melhores resultados, levemente melhor que o método de Máxima Verossimilhança. O RF apresentou uma exatidão geral de 98,06% e valor Kappa de 0,94. A imagem gerada pela classificação em cada um dos quatro métodos foi comparada com a delimitação feita de forma manual (Figura 2).



Figura 2 - Comparação da classificação utilizando os métodos tradicionais de sensoriamento remoto e aprendizado de máquina com a delimitação manual.





Fonte: Elaborado pelo Autor.

Visualmente é difícil dizer qual método se mostra mais confiável e próximo da realidade. Há pontos em que as delimitações automáticas feitas pelos métodos se mostram bem próximas à delimitação feita de forma manual, porém há pontos em que as delimitações automáticas variam bastante. Um exemplo claro pode ser visto na represa Águas Claras, na imagem de 2010, onde em um ponto mais acima do método de Distância Mínima está bem



próximo aos limites manuais, porém em um outro ponto mais abaixo ele superestima os limites do corpo hídrico indo além da delimitação manual, assim como das outras delimitações feitas pelos demais métodos.

Comparar técnicas clássicas de sensoriamento remoto com técnicas de aprendizado de máquina foi interessante porque a maioria dos trabalhos compara as técnicas clássicas apenas entre si, como o já citado estudo de Pereira & Guimarães (2018) que comparou tais técnicas aplicadas ao mapeamento de solo considerando diferentes base de dados e estratégias amostrais. De forma semelhante, as técnicas de aprendizado de máquina são comparadas entre elas, como por exemplo o estudo de Kampichler et. al (2010) que compara cinco métodos de aprendizado de máquina aplicados em ecologia e conservação biológica. Mais recentemente, há trabalhos onde houve a comparação de técnicas de aprendizado de máquina com técnicas de *deep learning*. Por exemplo, Fang et. al (2019) utilizaram métodos de *deep learning* aplicados em imagens Landsat 8 para o reconhecimento de reservatórios de água, comparando-os com métodos tradicionais de aprendizado de máquina. Os autores verificaram que os métodos de *deep learning* se mostraram superiores, apresentando maior exatidão.

A comparação de técnicas tradicionais com aquelas baseadas em aprendizagem de máquina, assim como a comparação de técnicas de aprendizado de máquina com técnicas de *deep learning*, mostram que técnicas mais complexas tendem a apresentar resultados mais precisos. Isso não significa que técnicas básicas são incapazes de fornecer resultados com alta confiabilidade, conforme verificamos no presente trabalho.

CONSIDERAÇÕES FINAIS

O presente trabalho teve por objetivo avaliar se as técnicas de sensoriamento remoto e aprendizado de máquina se mostram confiáveis para a classificação e análise de recursos hídricos em imagens óticas orbitais no contexto do Sistema Cantareira. Investigamos técnicas de classificação de imagens baseadas em técnicas clássicas de sensoriamento remoto (Distância Mínima, Mapeamento do Ângulo Espectral e Máxima Verossimilhança), comparando-as com uma técnica de *machine learning* (*Random Forest*).

Concluimos que os métodos clássicos de Distância Mínima e Máxima Verossimilhança fornecem resultados confiáveis (95,91% e 97,84%, respectivamente), mas que a aplicação do RF é capaz de nos entregar uma classificação ainda mais próxima dos 100% (98,06%).

Vale destacar a importância do monitoramento de recursos hídricos por meio de sensoriamento remoto. O monitoramento feito por técnicos, ainda que eficaz, pode exigir uma alta demanda de tempo e custo para sua realização. Por meio do sensoriamento remoto há a possibilidade de cobrir a área com uma visão geral da situação, em um curto período. Há vários programas que disponibilizam suas imagens orbitais gratuitamente, o que não gera custos para a realização desse monitoramento. Por fim, o presente trabalho apresenta também contribuição para a geração de um conjunto de dados rotulados que será relevante para o desenvolvimento de pesquisas futuras, principalmente envolvendo técnicas de *deep learning*.



AGRADECIMENTOS

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

REFERÊNCIAS

- ADAM, Elhadi et al. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, v. 35, n. 10, p. 3440-3458, 2014.
- AHMAD, Iftikhar et al. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE access*, v. 6, p. 33789-33795, 2018.
- ANACHE, Jamil AA et al. Hydrological trade-offs due to different land covers and land uses in the Brazilian Cerrado. *Hydrology & Earth System Sciences*, v. 23, n. 3, p. 1263-1279, 2019.
- ANTUNES, M. A. H; SIQUEIRA, J. C. S. Características das imagens RapidEye para mapeamento e monitoramento e agrícola e ambiental. *Simpósio Brasileiro de Sensoriamento Remoto*, v. 16, p. 547-554, 2013.
- BALLARD, Grant et al. Fine-scale oceanographic features characterizing successful Adélie penguin foraging in the SW Ross Sea. *Marine Ecology Progress Series*, v. 608, p. 263-277, 2019.
- BORSA, Adrian Antal; AGNEW, Duncan Carr; CAYAN, Daniel R. Ongoing drought-induced uplift in the western United States. *Science*, v. 345, n. 6204, p. 1587-1590, 2014.
- CARVALHO, Mairon Anderson Cordeiro Correa de et al. Drought Monitoring Based on Remote Sensing in a Grain-Producing Region in the Cerrado–Amazon Transition, Brazil. *Water*, v. 12, n. 12, p. 3366, 2020.
- CHEN, Y., FAN, R., YANG, X., WANG, J., & LATIF, A. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water*, 10(5), 585. 2018.
- CHOWDHURY, Masuma; HASAN, Mohammad Emran; ABDULLAH-AL-MAMUN, M. M. Land use/land cover change assessment of Halda watershed using remote sensing and GIS. *The Egyptian Journal of Remote Sensing and Space Science*, v. 23, n. 1, p. 63-75, 2020.
- CONCEIÇÃO, Adriana Fantinati et al. Balanço de energia da superfície nas bacias hidrográficas dos rios Cachoeira e Atibainha do Sistema Cantareira por meio de sensoriamento remoto. Adriana Fantinati Conceição. – Campinas, SP : [s.n.], 2018.
- CONGEDO, Luca. Semi-automatic classification plugin documentation. Release, v. 4, n. 0.1, p. 29, 2016.
- COUTINHO, Renato M.; KRAENKEL, Roberto A.; PRADO, Paulo I. Catastrophic regime shift in water reservoirs and São Paulo water supply crisis. *PloS one*, v. 10, n. 9, p. e0138278, 2015.
- CUNHA, Ana Paula et al. Extreme drought events over Brazil from 2011 to 2019. *Atmosphere*, v. 10, n. 11, p. 642, 2019.
- DE OLIVEIRA, Juarez Ventura et al. Differences in precipitation and evapotranspiration between forested and deforested areas in the Amazon rainforest using remote sensing data. *Environmental Earth Sciences*, v. 77, n. 6, p. 1-14, 2018.
- DE OLIVEIRA, Lucas Yuri Dutra et al. Análise de armazenamento hídrico na Bacia do Paraná utilizando dados Grace, associado a dados Mod16 e TRMM. *Anuário do Instituto de Geociências – UFRJ*. Vol. 42-3. p. 624-630, 2019.
- DE OLIVEIRA, Ulisses Costa et al. Avaliação do uso do solo e cobertura vegetal na zona de amortecimento do Parque Nacional de Ubajara, em Ubajara, Ceará. In: *X Congresso Brasileiro de Gestão Ambiental, Fortaleza, Ceará* 2019.



FANG, W., WANG, C., CHEN, X., WAN, W., LI, H., ZHU, S., ... & HONG, Y. Recognizing Global Reservoirs From Landsat 8 Images: A Deep Learning Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9), 3168-3177. 2019.

FITZ, Paulo Roberto. Considerações sobre o uso de polígonos de amostragem em classificações supervisionadas de imagens de satélite. *Caderno de Geografia*, v. 29, n. 59, p. 1124-1138, 2019.

FURUYA, Danielle Elis Garcia et al. A Machine Learning Approach for Mapping Forest Vegetation in Riparian Zones in an Atlantic Biome Environment Using Sentinel-2 Imagery. *Remote Sensing*, v. 12, n. 24, p. 4086, 2020.

GUSMÃO, Ana CVL et al. Determinação do saldo radiativo na Ilha do Bananal, TO, com imagens orbitais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 16, n. 10, p. 1107-1114, 2012.

GHANBARI, Hamid et al. Gaussian mixture model and Markov random fields for hyperspectral image classification. *European Journal of Remote Sensing*, v. 51, n. 1, p. 889-900, 2018.

IMAGENS LANDSAT. EarthExplorer USGS (United States Geological Survey). 2020. Disponível em: <<https://earthexplorer.usgs.gov>>. Acesso em: 05 de mar. 2020.

IMAGENS RAPIDEYE. Santiago & Cintra Consultoria. 2019. Disponível em:<<https://www.scon.com.br/produtos/imagens-rapideye/>>. Acesso em: 10 de nov. 2019.

ISIKDOGAN, Furkan; BOVIK, Alan C.; PASSALACQUA, Paola. Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 10, n. 11, p. 4909-4918, 2017.

JUNIOR, Emerson Leão; GALO, Maria de Lourdes Bueno Trindade; IVÁNOVÁ, Ivana. Comparação entre abordagens de classificação por pixel e baseada em objeto para o monitoramento de recursos hídricos. *Revista Brasileira de Cartografia*, v. 70, n. 4, p. 1289-1309, 2018.

KAMPICHLER, Christian et al. Classification in conservation biology: a comparison of five machine-learning methods. *Ecological Informatics*, v. 5, n. 6, p. 441-450, 2010.

KUSSUL, Nataliia et al. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, v. 14, n. 5, p. 778-782, 2017.

LANDIS, J. Richard; KOCH, Gary G. The measurement of observer agreement for categorical data. *Biometrics*, p. 159-174, 1977.

MARCATO JUNIOR, J.; MARTINS, E. F. O.; MARCATO, D.; MARINI, L. B.; PARANHOS FILHO, A. C. Avaliação da qualidade geométrica de ortoimagens RapidEye disponibilizadas pelo GeoCatálogo do MMA. *Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR, João Pessoa-PB, Brasil, 25 a 29 de abril de 2015, INPE*.

MARTINS, Iris Amati et al. Determination of the Trophic State Index (TSI) using remote sensing, bathymetric survey and empirical data in a tropical reservoir. *Limnetica*, v. 39, n. 2, p. 539-553, 2020.

MCFEETERS, Stuart 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.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003.

NAMIKAWA, Laércio Massaru. Imagens landsat 8 para monitoramento de volume de água em reservatórios: estudo de caso nas barragens jaguari e jacaréi do sistema cantareira. *Simpósio Brasileiro de Sensoriamento Remoto*, v. 17, p. 4828-4835, 2015.

PAL, Mahesh. Random forest classifier for remote sensing classification. *International journal of remote sensing*, v. 26, n. 1, p. 217-222, 2005.

PEREIRA, Marçal Duarte; SCHETTINI, Carlos Augusto França; OMACHI, Claudia Yuki. Caracterização de feições oceanográficas na plataforma de Santa Catarina através de imagens orbitais. *Revista Brasileira de Geofísica*, v. 27, n. 1, p. 81-93, 2009.



PEREIRA, Luís Flávio; GUIMARÃES, Ricardo Morato Fiúza. Mapeamento multicategórico do uso/cobertura da terra em escalas detalhadas usando Semi-automatic Classification Plugin. *Journal of Environmental Analysis and Progress*, v. 3, n. 4, p. 379-385, 2018.

PERMUTER, Haim; FRANCOS, Joseph; JERMYN, Ian. A study of Gaussian mixture models of color and texture features for image classification and segmentation. *Pattern Recognition*, v. 39, n. 4, p. 695-706, 2006.

Portal dos Mananciais. Dados dos Sistemas Produtores. Sabesp. 2020. Disponível em: < <http://mananciais.sabesp.com.br/HistoricoSistemas?SistemaId=0> >. Acesso em: 07 de dez. 2020.

QGIS, Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Disponível em: < https://www.qgis.org/pt_BR/site/ >. Acesso em: 05 de nov. 2019.

RAPIDEYE, A. G. Satellite imagery product specifications. Satellite imagery product specifications: Version 6.1, 2016. Sistema Cantareira. Sabesp. 2020. Disponível em: < <http://site.sabesp.com.br/site/interna/Default.aspx?secaoId=132> >. Acesso em: 07 dez. 2020.

RODRIGUES, Wendel Barbosa; DE MORAIS, Fernando; PASCHOAL, Leticia Giuliana. Índice de diferença normalizada da água (NDWI) calculado para estações chuvosas e secas na bacia do Córrego Barreiro, Lagoa da Confusão–TO. *Os Desafios da Geografia Física na Fronteira do Conhecimento*, v. 1, p. 254-263, 2017.

Sala de Situação. Agência Nacional de Águas. 2019. Disponível em: < <https://www.ana.gov.br/sala-de-situacao/sistema-cantareira/sistema-cantareira-saiba-mais> >. Acesso em: 03 de nov. 2019.

SÃO PAULO: população do estado. Biblioteca Virtual, Governo do Estado de São Paulo. 2018. Disponível em: < <http://www.bibliotecavirtual.sp.gov.br/temas/sao-paulo/sao-paulo-populacao-do-estado.php> >. Acesso em: 07 de nov. 2019.

SEADE População. Governo do Estado de São Paulo. 2022. Disponível em: < <https://populacao.seade.gov.br/> >. Acesso em: 05 de ago. 2022.

SILVA, Rebeca de Souza; PAES, Angela Tavares. Por Dentro da Estatística: teste de concordância de Kappa. *Educ Contin Saúde Einstein*, v. 10, n. 4, p. 165-6, 2012.

Sistema Cantareira. Sabesp. 2020. Disponível em: < <http://site.sabesp.com.br/site/interna/Default.aspx?secaoId=132> >. Acesso em: 07 dez. 2020.

SOUZA, João Paulo Sena et. al. Comparação dos métodos de classificação por ângulo espectral e distância euclidiana no mapeamento das formas de terreno. *Revista Brasileira de Geomorfologia, [S. l.]*, v. 17, n. 3, p. 591-613, 2016.

TIAN, Shaohong et al. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sensing*, v. 8, n. 11, p. 954, 2016.

THANH NOI, Phan; KAPPAS, Martin. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, v. 18, n. 1, p. 18, 2018.

VOGEL, Ivan; BLANSHARD, Robert C.; HOFFMANN, Eva R. SureTypeSC—a Random Forest and Gaussian mixture predictor of high confidence genotypes in single-cell data. *Bioinformatics*, v. 35, n. 23, p. 5055-5062, 2019.

COMO CITAR ESTE TRABALHO

OLIVEIRA, Lucas. OSKO, Lucas Prado. JR, José Marcato. RAMOS, Ana Paula Marques. Souza, Maurício. Sensoriamento remoto e aprendizado de máquina aplicados no mapeamento hídrico do Sistema Cantareira. *Revista Tamoios, São Gonçalo*, v. 19, n. 1, p. 193-206, 2023. Disponível em: <https://doi.org/10.12957/tamoios.2023.63354> . Acesso em: DD MM. AAAA.