USO DE SÉRIES TEMPORAIS PARA CLASSIFICAÇÕES DE USO E COBERTURA DA TERRA EM PETROLINA, PERNAMBUCO

Pedro V. da Silva Brito[1], *Michel E. D. Chaves*[1], *Herica F. de Sousa Carvalho*[2], *Felipe C. de Souza*[1], *Baggio L. de Castro e Silva*[1], *Karine R. Ferreira*[1], *Rafael D. Coelho dos Santos*[1] *e Gilberto R. de Queiroz*[1],

[1]Instituto Nacional de Pesquisas Espaciais (INPE), Av. dos Astronautas, 1758 - São José dos Campos-SP, Brasil, {pedro.brito;michel.chaves;felipe.carvalho;baggio.silva;karine.ferreira;rafael.santos;gilberto.queiroz}@inpe.br
[2]Universidade Federal de Pernambuco (UFPE), Av. Prof. Moraes Rego 1235, 50670-901, Cidade Universitária, Recife-PE, Brasil, herica.carvalho@ufpe.br

RESUMO

A vasta gama de imagens de satélite disponibilizadas gratuitamente e os avanços de métodos computacionais ampliaram o uso de técnicas de sensoriamento remoto para analisar a superfície terrestre, melhorando análises sobre vegetação. Em regiões como a zona semiárida do bioma Caatinga, isso tem sido essencial para avaliar as condições e a sazonalidade característica da vegetação local. O objetivo deste estudo foi gerar classificações de uso e cobertura da terra e avaliar mudanças entre 2018 e 2020 por meio de duas abordagens: (1) classificação de séries temporais de imagens de satélite e (2) classificação de cada imagem de forma individual. Esse estudo conclui que a primeira abordagem usando o método *Randon Forest* e séries temporais NDVI potencializou a capacidade de diferenciar classes de uso e cobertura da terra em Petrolina-PE.

Palavras-chave – Random Forest, Cubos de dados, Análise temporal.

ABSTRACT

The wide range of freely available satellite images and advances in computational methods have promoted the use of remote sensing techniques to analyze the Earth's surface, improving analysis of vegetation. In regions such as the semi-arid zone of the Caatinga biome, this has been essential to assess the conditions and seasonality characteristic of the local vegetation. The objective of this study was to generate land use and land cover classifications and assess changes between 2018 and 2020 through two approaches: (1) classification of satellite image time series and (2) classification of each satellite image individually. This study concludes that the first approach using the Randon Forest method and NDVI time series enhanced the differentiation of land use and land cover classes in Petrolina-PE.

Key words – *Random Forest, Data cubes, Temporal analysis.*

1. INTRODUÇÃO

Rico em recursos naturais quando comparado a outras regiões semiáridas do mundo, o bioma Caatinga está entre os ecossistemas menos estudados e mais ameaçados no quesito mudanças de uso e cobertura da terra no Brasil [1]. Contudo, a retirada de vegetação nativa em demasia, especialmente por ação antrópica, tem alterado processos físicos, químicos e biológicos em escala local, provocando mudanças no ciclo hidrológico e no clima local [2]. Neste sentido, melhorar a compreensão sobre condições e sazonalidade da vegetação é uma necessidade premente. Sensoriamento remoto é amplamente utilizado no desenvolvimento de metodologias que possam embasar ações voltadas ao planejamento e monitoramento do uso adequado da terra e, consequentemente, da vegetação. Análises de imagens de satélite são úteis para o monitoramento sistemático e contínuo da cobertura vegetal de agroecossistemas [3].

A disseminação gratuita de imagens de satélite possibilitou aumentar o entendimento sobre a superfície terrestre e sua dinâmica [4]. Isso tornou possível o uso de séries temporais de imagens para gerar informações precisas sobre uso e cobertura da terra e monitorar processos naturais e antrópicos. Para análise de séries temporais de imagens, o pacote R SITS (*Satellite Image Time Series Analysis on Earth Observation Data Cubes*) fornece um conjunto de métodos, incluindo aprendizado de máquina e aprendizado profundo, para gerar classificações de uso e cobertura da terra [5].

O potencial de ferramentas como esta, tem tornado possível melhorar análises acerca da sazonalidade da vegetação em áreas de constante mudanças no uso da terra ao longo de determinados períodos. A zona semiárida da Caatinga tem sofrido intensa degradação nos últimos 400 anos devido ao uso desordenado e predatório [6]. Um exemplo está no município de Petrolina, localizado em Pernambuco, que é responsável pelo desenvolvimento de grandes projetos agrícolas irrigados ao longo de sua extensão, promovendo e intensificando mudanças do uso da terra. Assim, o objetivo deste estudo foi gerar classificações de uso e cobertura da terra em Petrolina a partir de imagens Sentinel-2/MultiSpectral Instrument (S2/MSI) e avaliar mudanças entre 2018 e 2020 por meio de duas abordagens: (1) classificação de séries temporais de imagens e (2) de uma imagem por ano de forma individual.

2. MATERIAL E MÉTODOS

Maior polo de fruticultura do Semiárido brasileiro, Petrolina tem área de $4.561,870\ km^2$ [7]. O clima é Semiárido quente e seco, com períodos definidos de chuva, entre novembro e abril, e estiagem, entre maio e outubro. A temperatura média é acima de $25°C$, e a pluviosidade de cerca de 500 mm/ano [8]. As etapas envolveram a classificação de séries temporais e de imagens individuais anuais (Figura 1). Foram coletadas 3000 amostras de uso e cobertura da terra, 1000/ano, 200 de cada uma das classes: Vegetação de caatinga, Agricultura irrigada, Corpos hídricos, Área urbana e Solo exposto.

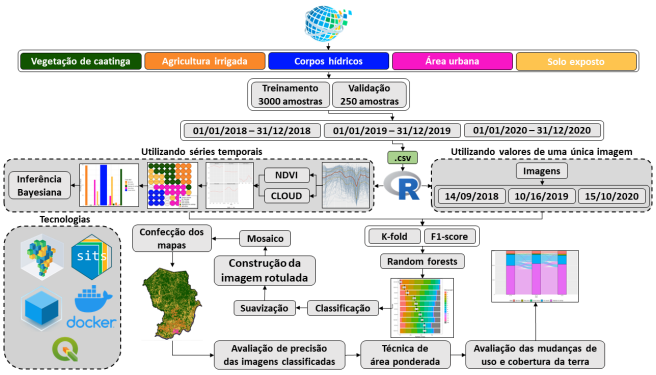


Figura 1: Fluxograma das etapas metodológicas do trabalho.

Para validar, foram coletadas 750 amostras (250/ano, 50/classe) via interpretação visual de imagens S2/MSI, com auxílio de séries temporais NDVI, usando a plataforma *TerraCollect*, desenvolvida no escopo do projeto *Brazil Data Cube* (BDC) [3]. Na classificação de séries temporais, foram usados cubos de dados S2/MSI disponibilizados pelo projeto BDC (S2-SEN2COR_10_16D_STK-1) e séries temporais NDVI extraídas desses cubos [3]. Eles possuem imagens com resolução espacial de 10 metros, projetadas e recortadas na grade BDC, com função de composição temporal de 16 dias usando a abordagem de melhor *pixel*. Também foi usada a banda SCL para mitigar efeitos de nuvens/sombras e aplicada interpolação linear para substituir valores removidos. Padrões espectro-temporais NDVI foram avaliados via *Generalized Additive Model* (GAM). As amostras foram filtradas pelo método proposto por Santos et al. (2021) [9], que usa mapas auto-organizáveis (*Self-Organizing Maps*, SOM) e inferência Bayesiana. Foram gerados 5 experimentos, variando tamanhos da grade de neurônios das amostras, adotando parâmetros propostos [9], que foram $5\frac{\sqrt{N}}{2}$ neurônios, em que, N é o número de amostras, taxa de aprendizado decrescente, de 0,50 a 0,01, e métrica euclidiana. Os tamanhos de grades foram 8x8, 9x9, 10x10, 11x11 e 12x12, com taxa de aprendizado de 0,50 a 0,01, 100 interações e distância euclidiana. Para classificar imagens individuais, foi escolhida uma imagem por ano (com menos nuvens/sombras, datadas de 14/09/2018, 16/10/2019 e 15/10/2020).

Nas duas etapas, foi aplicada a validação cruzada *k-fold* ($k = 5$) para selecionar os melhores modelos de classificação. Os modelos *Random Forest* (RF) avaliados diferenciam-se pelo número de árvores (100, 500, 1000, 1500 e 2000). Após a seleção do modelo mais preciso por ano, foi aplicada a avaliação *F1-score*. Os passos seguintes envolveram treinamento dos modelos, classificação e suavização bayesiana, para refiná-las. A precisão das classificações foi avaliada via técnica de área ponderada [10]. Mudanças de uso e cobertura da terra foram avaliadas via diagrama *alluvial*, analisando 3000 amostras aleatórias.

3. RESULTADOS E DISCUSSÃO

Os padrões espectro-temporais NDVI (Figura 2) mostram como as classes possuem diferenças. Apenas Área urbana e Solo exposto apresentaram perfis semelhantes. Em relação aos agrupamentos das séries temporais via SOM (Figura 3), em 2019 e 2020, houve apenas um neurônio *outlier*, com classes divergentes das predominantes na vizinhança. O agrupamento de 2018 não apresentou *outliers*.

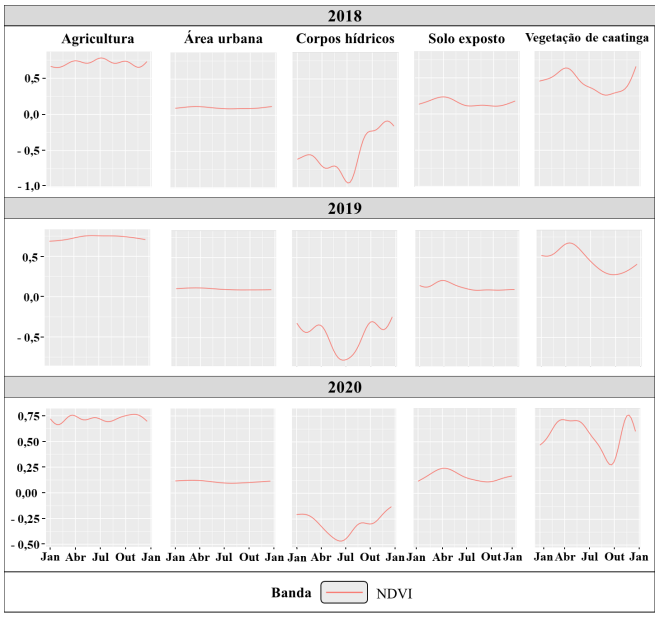


Figura 2: Padrões espectro-temporais NDVI de cada classe para o NDVI obtido via GAM para os anos de 2018, 2019 e 2020.

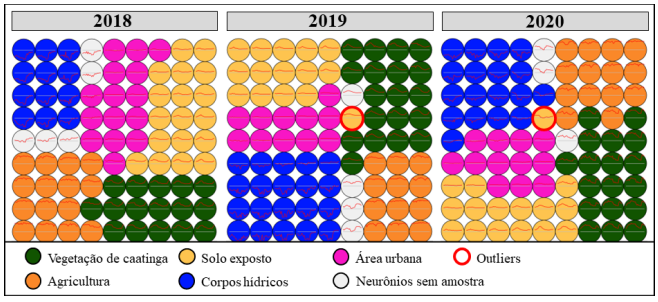


Figura 3: Agrupamentos das séries temporais de NDVI na grade 9x9 para os anos 2018, 2019 e 2020.

A aplicação do SOM mostra, em todo o período, agrupamentos com graus de pureza acima de 86% (Figura 4).

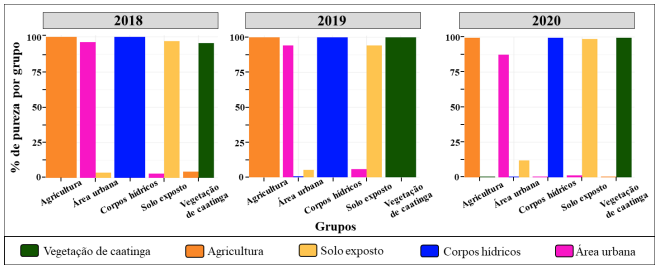


Figura 4: Porcentagem de confusão entre os grupos do agrupamento SOM com grade 9x9 para 2018, 2019 e 2020.

Os agrupamentos de 2018 e 2019 apresentaram maior separabilidade, com 100% de pureza em Agricultura irrigada e Corpos hídricos. Em particular, 2019 apresentou 100% de pureza para Vegetação de caatinga. Os grupos de Área urbana e Solo exposto apresentaram maior confusão, com exceção de 2018, em que, Vegetação de caatinga mostrou confusão

com Agricultura irrigada. A classe Corpos hídricos não foi homogênea em 2020.

Os resultados obtidos com a aplicação da técnica *k-fold* ($k = 5$) (Tabela 1) mostram maiores valores de acurácia utilizando séries temporais, e modelos com acurácia satisfatória nas duas abordagens. Em 2018, as maiores acurácias ocorreram no modelo com 500 árvores ($97,9\%$ e $77,6\%$ para séries temporais e única imagem, respectivamente). Por outro lado, o modelo com 100 árvores teve acurácia de $98,0\%$ e $98,5\%$ em 2019 e 2020 usando séries temporais e de $76,9\%$ e $74,9\%$ com a imagem anual.

Utilizando séries temporais						
		Acurácia %				
Nº árvores		**100**	**500**	**1000**	**1500**	**2000**
	2018	97,6	97,9	97,5	97,8	97,4
	2019	98,0	97,8	97,7	97,9	97,7
	2020	98,5	98,5	98,0	98,0	98,0
Utilizando uma única imagem por ano						
Nº árvores		**100**	**500**	**1000**	**1500**	**2000**
	2018	76,9	77,6	75,3	76,3	76,0
	2019	76,9	75,9	75,7	76,9	76,4
	2020	74,6	73,2	72,6	73,7	73,8

Tabela 1: Validação *k-fold* no conjunto de treinamento com o modelo RF utilizando séries temporais e uma única imagem.

Os resultados do *F1-score* (Tabela 2) apresentam, para cada classe, maiores valores usando séries temporais (acima de 90% de precisão). Com séries temporais, os menores valores ocorreram para Vegetação de caatinga ($93,8\%$, $94,7\%$ e $95,2\%$). Para Corpos hídricos, maior exatidão (100% e $99,7\%$ para 2018 e 2019). Para Solo exposto, $99,5\%$ para todos os conjuntos. Utilizando apenas uma imagem, os maiores valores foram das classes Solo exposto, e Corpos hídricos com $97,3\%$ e $94,3\%$, 94%, respectivamente. Semelhante aos resultados com séries temporais, Vegetação de caatinga teve menor acurácia usando uma única imagem.

Utilizando séries temporais			
	Acurácia %		
Classes	**2018**	**2019**	**2020**
Agricultura	99,2	99,2	99,2
Área urbana	95,1	95,5	94,5
Corpos hídricos	100,0	99,7	99,0
Solo exposto	99,5	99,5	99,5
Vegetação de caatinga	94,7	95,2	93,8
Utilizando uma única imagem por ano			
Agricultura	86,5	61,4	91,0
Área urbana	63,8	88,7	56,9
Corpos hídricos	93,8	94,0	93,9
Solo exposto	97,3	94,3	77,2
Vegetação de caatinga	45,4	53,3	54,0

Tabela 2: *F1-scores* no conjunto de treinamento com o modelo RF utilizando séries temporais e uma única imagem por ano.

A classificação de uso e cobertura da terra utilizando séries temporais e modelo RF (Figura 5) apresenta melhor identificação de Áreas agrícolas e do rio São Francisco. Por outro lado, observou-se algumas confusões em regiões destacadas (região a) em que, locais de Solo exposto foram classificados como Área urbana. Em 2020, detectou-se mais Corpos hídricos (região b), o que pode ter relação com a precipitação. Por interpretação visual e conhecimento da região, a classificação de 2019 foi mais precisa.

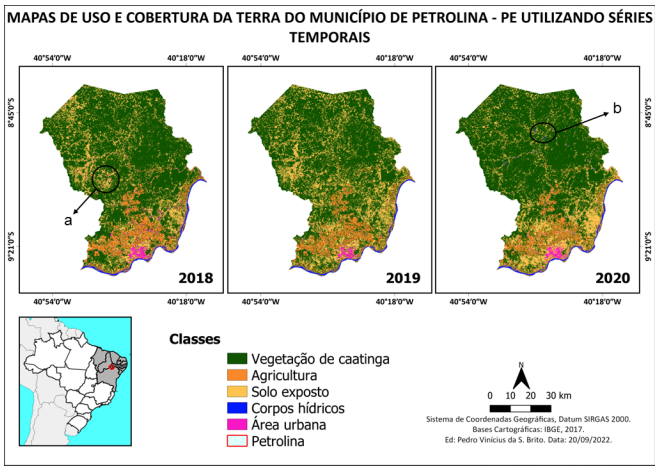


Figura 5: Classificações de uso e cobertura da terra via RF para Petrolina-PE utilizando séries temporais.

A classificação usando uma única imagem anual e modelo RF (Figura 6) mostra maior confusão entre as classes, sobretudo, Solo exposto e Área urbana. As regiões c e d representam locais de Solo exposto que foram classificados como Área urbana. Embora 2018 tenha detectado o leito do rio São Francisco, 2019 e 2020 apresentaram confusões entre Corpos hídricos e Área urbana, predominantemente ao sul.

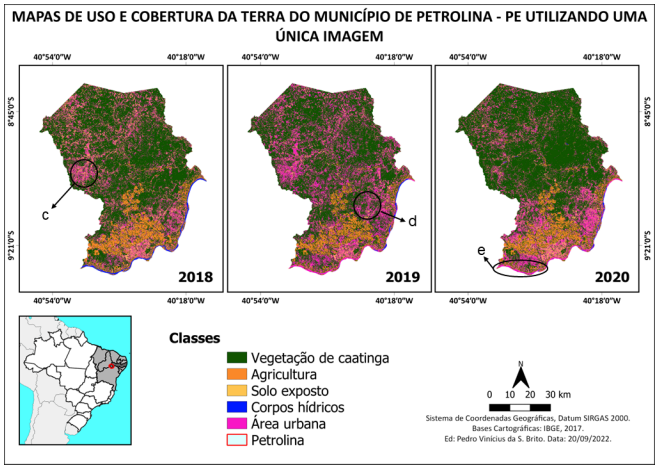


Figura 6: Classificações de uso e cobertura da terra via RF para Petrolina-PE utilizando uma única imagem por ano.

O uso de séries temporais aumentou a acurácia da maioria das classes (Tabela 3). Vegetação de caatinga obteve 100% de acurácia do produtor em todas as situações. No entanto, a classe de Área urbana continuou apresentando confusão. As transições entre classes (Figura 7) mostram, na abordagem de séries temporais, mudanças oriundas da ocupação local, como a transição Vegetação de caatinga-Solo exposto; o que não foi observado com uma imagem por ano.

Os resultados mostraram diferenças entre as abordagens. A abordagem de classificação via séries temporais aperfeiçoou a capacidade de diferenciar classes de uso e cobertura da terra (Tabela 3), diminuindo, consequentemente, confusões

	AP %		AU %		AG %	
2018						
Classes	**ST**	**UI**	**ST**	**UI**	**ST**	**UI**
Agricultura	60	55	100	96	87	77
Área urbana	24	56	82	62		
Corpos hídricos	100	73	100	100		
Solo exposto	92	50	87	62		
Vegetação de caatinga	100	99	84	80		
2019						
Agricultura	100	76	100	100	96	79
Área urbana	22	78	89	45		
Corpos hídricos	100	03	100	100		
Solo exposto	99	55	85	84		
Vegetação de caatinga	100	99	100	91		
2020						
Agricultura	71	66	100	96	94	76
Área urbana	30	59	88	29		
Corpos hídricos	99	03	100	100		
Solo exposto	93	39	88	50		
Vegetação de caatinga	100	100	94	91		

ST = Séries temporais, UI = Uma imagem, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia Global.

Tabela 3: Valor de acurácia das imagens classificadas a partir da técnica de área ponderada.

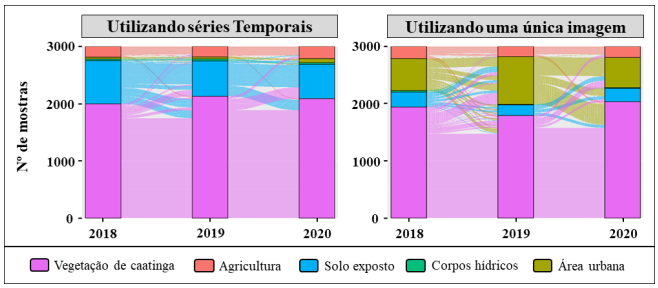


Figura 7: Mudanças de uso e cobertura da terra dos mapas classificados utilizando séries temporais e uma única imagem.

espectrais. Previamente, verificou-se similaridade do perfil espectro-temporal entre as classes Solo exposto e Área urbana (Figura 2), apresentando forte confusão entre as mesmas, devido à similaridade de suas assinaturas espectro-temporais.

A presença de neurônio *outlier* na Figura 3 não representa rotulagem errada das amostras coletadas [9]. Isso pode ocorrer devido às amostras com diferentes padrões espaço temporais para a mesma classe ou não-separáveis com o uso de séries temporais NDVI. A heterogeneidade apresentada na vizinhança de alguns agrupamentos pode representar a variação natural entre as amostras da mesma classe e não devem ser removidas do conjunto de amostras.

Fazendo uso do *k-fold* utilizando o mesmo número de árvores nos dois testes, a abordagem de séries temporais obteve maior acurácia (Tabela 1). Assim, foi possível identificar aumento na acurácia com mais amostras representativas. Já nos resultados do *F1-score* (Tabela 2) percebe-se pequena variabilidade no percentual para Vegetação nativa (caatinga), o que pode estar associado à alta heterogeneidade espacial e variabilidade da cobertura vegetal associada à sazonalidade pluviométrica da região. Ademais, observou-se similaridade entre Solo exposto e Corpos hídricos, alvos pouco alterados durante o ano.

4. CONCLUSÕES

O uso de séries temporais NDVI melhorou a identificação de classes de uso e cobertura da terra em comparação a abordagem de classificação de uma imagem anual e favoreceu a identificação de mudanças na área de estudo. Como o município de Petrolina apresenta dinâmica de conversão envolvendo diferentes classes de vegetação, recomenda-se o uso da abordagem de séries temporais para detectar mudanças relacionadas à sua sazonalidade e dinâmica para o caso de áreas de estudo com características semelhantes.

5. AGRADECIMENTOS

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Código de Financiamento 001). Os autores também agradecem à FAPESP (processo 2021/07382-2 - Michel E. D. Chaves) e ao projeto Brazil Data Cube.

6. REFERÊNCIAS

[1] Z. Fernández, J. A. Santos Júnior, R. Amaral, E. França, and et al. Pernambuco caatinga: relevance of soil chemical composition for biodiversity conservation. *Chemistry and Ecology*, 38:108–121, 2022.

[2] J. Q Chambers and P. Artaxo. Deforestation size influences rainfall. *Nature Climate Change*, 7:175–176, 2017.

[3] K. Ferreira, G. Queiroz, L. Vinhas, R. Marujo, and et al. Earth observation data cubes for brazil: Requirements, methodology and products. *Remote Sensing*, 12(24):4033, 2020.

[4] M. Chaves, M. Picoli, and I. Sanches. Recent applications of landsat 8/OLI and sentinel-2/MSI for land use and land cover mapping: A systematic review. *Remote Sensing*, 12(18):3062, 2020.

[5] R. Simoes, G. Camara, G. Queiroz, and et al. Satellite image time series analysis for big earth observation data. *Remote Sensing*, 13(13):2428, 2021.

[6] J. Bezerra, G. Moura, B. Silva, P. Lopes, and Ê. Silva. Parâmetros biofísicos obtidos por sensoriamento remoto em região semiárida do estado do rio grande do norte, brasil. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 18:73–84, 2014.

[7] IBGE. Instituto brasileiro de geografia e estatística, 2022. Disponível em: <https://cidades.ibge.gov.br/brasil/pe/petrolina/panorama>. Acesso em: 21/10/2022.

[8] J. D. Galvíncio, S. M. Mendes, Y. C. B. Morais, R. Q. Miranda, and et al. Correlação linear entre a precipitação e o índice de área foliar do bioma caatinga. *Revista Brasileira de Geografia Física*, 13:3304–3314, 2020.

[9] L. Santos, K. Ferreira, and et al. Quality control and class noise reduction of satellite image time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 177:75–88, 2021.

[10] P. Olofsson and et al. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129:122–131, 2013.