

---

**PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E BIODIVERSIDADE**

---

**VARIAÇÃO HISTÓRICA E ESPACIAL DO REGIME DE FOGO DO CERRADO  
PAULISTA**

**DHEMERSON ESTEVÃO CONCIANI**

---

**PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E BIODIVERSIDADE**

---

**VARIAÇÃO HISTÓRICA E ESPACIAL DO REGIME DE FOGO DO CERRADO  
PAULISTA**

**DHEMERSON ESTEVÃO CONCIANI**

Dissertação apresentada ao Instituto de Biociências do Câmpus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Mestre em Ecologia e Biodiversidade.

**Orientador:** Thiago Sanna Freire Silva  
**Co-orientadora:** Swanni Tatiana Alvarado

C744v

Conciani, Dhemerson Estevão

Varição histórica e espacial do regime de fogo do Cerrado paulista  
/ Dhemerson Estevão Conciani. -- Rio Claro, 2021

90 p. : il., tabs., mapas

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),  
Instituto de Biociências, Rio Claro

Orientador: Thiago Sanna Freire Silva

Coorientador: Swanni Tatiana Alvarado

1. Ecologia. 2. Ecologia do fogo. 3. Machine learning. 4.  
Sensoriamento remoto. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de  
Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: **VARIAÇÃO HISTÓRICA E ESPACIAL DO REGIME DE FOGO DO CERRADO PAULISTA**

**AUTOR: DHEMERSON ESTEVÃO CONCIANI DA COSTA**

**ORIENTADOR: THIAGO SANNA FREIRE SILVA**

**COORIENTADORA: SWANNI TATIANA ALVARADO ROMERO**

Aprovado como parte das exigências para obtenção do Título de Mestre em ECOLOGIA E BIODIVERSIDADE, área: Biodiversidade pela Comissão Examinadora:

Prof. Dr. THIAGO SANNA FREIRE SILVA (Participação Virtual)  
Biological and Environmental Sciences / University of Stirling - Escócia



Profa. Dra. ANE AUXILIADORA COSTA ALENCAR (Participação Virtual)  
Instituto de Pesquisa Ambiental da Amazônia / Brasília - DF

Prof. Dr. DANIEL BORINI ALVES (Participação Virtual)  
Pós Doutorando do Departamento de Biodiversidade / UNESP - Instituto de Biociências de Rio Claro - SP



Rio Claro, 03 de março de 2021

*Dedico este trabalho ao sistema público de ensino*

## **AGRADECIMENTOS**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento recebido (CNPq - 380522/2019-5). Ao Instituto Florestal do Estado de São Paulo (IF) e à Fundação Florestal Para Conservação e Produção Florestal do Estado de São Paulo (FF) pelo fornecimento de informações georreferenciadas das Unidades de Conservação, em especial à gestão, servidores e amigos da Estação Ecológica de Itirapina pelo intercâmbio profissional e hospitalidade. Agradeço ao United States Geological Survey – Center Science Processing Architecture (USGS-ESPA) pelo acervo de imagens orbitais e ao Observatório de Dinâmicas Ecosistêmicas (Ecodyn) pela troca de experiências e infraestrutura computacional.

Por fim, agradeço a minha família, orientadores, amigos, colegas, conhecidos, desconhecidos; afetos e desafetos; por contribuírem de diferentes formas em minha formação.

*“ [...] Apenas que, busquem conhecimento!”*

E.T. Bilu

## RESUMO GERAL

O Cerrado ocupa 25% do território brasileiro, sendo considerado o segundo maior bioma do país, atrás apenas da floresta Amazônica. Nas últimas décadas o Cerrado experimentou um rápido declínio em sua vegetação nativa, sendo substituído principalmente por agricultura e pastagem. Hoje, menos de 8% do Cerrado encontra-se legalmente protegido no Brasil. Considerando o contexto do estado de São Paulo (SP), esta situação é ainda pior e menos de 1% da cobertura original do Cerrado encontra-se protegida. Neste cenário, os poucos remanescentes de vegetação nativa que restaram no estado de SP estão abrigados no interior de Unidades de Conservação (UC). Por sua vez, essas UCs possuem áreas que variam desde 200 hectares (ha) até 9000 ha, sendo consideradas relativamente pequenas quando comparadas à UCs de Cerrado em outros estados do Brasil. Considerando a raridade de áreas protegidas no Cerrado paulista, a conservação efetiva dessas UCs é fundamental para garantir a provisão de serviços ecossistêmicos e a preservação da biodiversidade. Contudo, as atividades desenvolvidas nas zonas de amortecimento (ZA), isto é, no entorno imediato dessas UCs, impactam negativamente sobre a conservação, contribuindo para disseminação de espécies invasoras e na ignição de queimas acidentais e criminosas. Buscando contribuir no entendimento dessa dinâmica, nós desenvolvemos um algoritmo, geramos e validamos um produto de áreas queimadas adaptado para o contexto do Cerrado paulista entre 1985 e 2018 (acurácia= 79%, erro de omissão= 16%, erro de comissão= 9%). Através de uma análise combinada entre os padrões de área queimada e as mudanças no uso e cobertura do solo nas últimas três décadas, nós identificamos que o padrão de queimadas no Cerrado paulista pode ser explicado pelo tipo de uso do solo. De um modo geral, o regime de fogo nesta região pode ser caracterizado como antrópico, ocorrendo principalmente em áreas de pastagem e cultivos de cana-de-açúcar. As UCs com cobertura predominantemente florestal não queimaram ou pouco queimaram ao longo da série temporal analisada. Por outro lado, picos de área queimada foram identificados a cada 8-9 anos para as UCs campestres enquanto ciclos de queima a cada 2-3 anos foram identificados nas UCs com densa invasão por gramíneas africanas. Novas estratégias de manejo que deem autonomia e



segurança jurídica para os gestores prescreverem queimas controladas precisam ser implementadas no Cerrado paulista, especialmente nas UCs campestres. Por outro lado, as UCs densamente invadidas por gramíneas africanas coincidem com áreas de conflito fundiário e/ou forte pressão urbana. Nesse contexto, é preciso buscar alternativas de restauração para essas áreas e fomentar a inclusão das comunidades locais em um modelo de gestão participativa como forma de mitigar os efeitos das pressões antrópicas e aumentar a efetividade de conservação dessas áreas.

**Palavras-chave:** Ecologia do fogo; Ecologia Vegetal; Unidades de Conservação; Machine Learning; Sensoriamento Remoto; Landsat.

## ABSTRACT

Cerrado cover 25% of the Brazilian territory, being considered the second-largest biome in the country, only behind the Amazon rainforest. In recent decades, the Cerrado has experienced a rapid decline in its native vegetation, being replaced mainly by agriculture and pasture. Recently, less than 8% of the Cerrado is legally protected in Brazil. Considering the context of the state of São Paulo (SP), this situation is even worse and less than 1% of the original Cerrado area is protected. In this scenario, the fewer remnants of native vegetation that currently occurs in the state of SP are sheltered inside Protected Areas (PA). Area of these PAs ranges from 200 hectares (ha) to 9000 ha, being considered relatively small when compared to Cerrado PAs in other Brazilian regions. Considering the rarity and context of PAs in the São Paulo's Cerrado, the effective conservation of these PAs is essential to guarantee the provision of ecosystem services and the biodiversity conservation. However, the activities carried out in the buffer zones (BZ), that is, in the immediate surroundings of these PAs, negatively impact conservation, contributing to the spread of alien species and the ignition of accidental and arson fires. To contribute to the understanding of this dynamic, we developed an algorithm, generated and validated a product of burned areas adapted to the context of the São Paulo's Cerrado between 1985 and 2018 (accuracy = 79%, omission error = 16%, commission error = 9%). Through a combined analysis of burnt area patterns and changes in land use and land cover over the last three decades, we identified that the burned area pattern in the São Paulo's Cerrado can be explained by the type of land use. In general, the fire regime in this region can be considered as anthropogenic, occurring mainly in pasture areas and sugarcane crops. PAs with predominantly forest cover did not burn or burned few times over the analyzed time series. On the other hand, peaks of burned area were identified every 8-9 years within grassland PAs, while burning cycles every 2-3 years were identified in PAs with dense invasion by African grasses. New management strategies that provide autonomy and legal security to PA managers to execute prescribed fires needs to be implemented in São Paulo's Cerrado, mainly in grassland PAs. On the other hand, PAs densely invaded by African grasses coincide with areas of land conflict and/or strong urban pressure. In this context, it's necessary to seek restoration alternatives for

these degraded PAs as well encourage the inclusion of local communities in a participative management model as a way to mitigate the effects of human pressures and increase the effectiveness of the São Paulo's Cerrado conservation.

**Key-words:** Fire Ecology; Plant Ecology; Protected Areas; Machine Learning; Remote Sensing; Landsat.

## SUMÁRIO

<b>Introdução geral</b> .....	10
Objetivo geral .....	12
Objetivos específicos .....	12
Referências bibliográficas .....	12
<b>Developing a machine learning based algorithm for regional time-series burned area mapping: The highly anthropized Cerrado challenge</b> .....	15
Abstract .....	16
Introduction .....	17
Material and Methods .....	19
Study area .....	19
Algorithm workflow .....	20
Building spectral library .....	21
Pre-processing .....	25
Model training and testing .....	26
Burned area validation .....	28
Post-processing .....	31
Final product compilation .....	32
Results and Discussion .....	32
Hyperparameters tuning and model selection .....	32
Multivariate Adaptive Regression Splines – MARS .....	32
Random Forest – RF .....	33
Extreme Gradient Boosting – XGB .....	35
Final model selection .....	37
Predictor’s importance .....	38
Burned area validation .....	40
Final product and data access .....	48
Known issues and future development .....	48
Conclusion .....	50
Acknowledgements .....	51
References .....	51
Supplementary .....	57

<b>The conservation paradox on highly anthropized Cerrado: Protect to burn or burn to protect?</b> .....	60
Abstract .....	61
Introduction .....	62
Methods .....	64
Study area .....	64
Management context .....	67
Landscape structure .....	67
Data collection .....	68
Protected areas and buffer zones .....	68
Burned area .....	69
Land cover and land use changes .....	69
Data processing and analyses .....	69
Pre-processing burned area data .....	69
Burned area by land cover and land use .....	70
Fire regime metrics .....	70
Results and Discussion .....	71
General burned area pattern from the highly anthropized Cerrado .....	71
Fire frequency within protected areas .....	74
Spatio-temporal variation of burned areas across PAs and its buffer zones .....	77
Management implications and future perspectives .....	80
Conclusion .....	82
Acknowledgments .....	83
References .....	83
Supplementary .....	87
<b>Considerações finais</b> .....	91

## INTRODUÇÃO GERAL

O Cerrado é um conjunto de ecossistemas cuja classificação formal suscita divergências ainda nos dias atuais, sendo apresentado como um único bioma (IBGE, 2019), um conjunto de diversos biomas (Batalha, 2011) ou um domínio fitogeográfico (Leopoldo Magno Coutinho, 2006). O Cerrado ocupa uma área de aproximadamente 2 milhões de km<sup>2</sup> (25% do território brasileiro), estendendo-se do norte do estado do Paraná até o litoral do Maranhão, pelo qual possui interfaces de transição com a Amazônia, Caatinga, Mata Atlântica e Pantanal (Durigan & Ratter, 2016; IBGE, 2019). Caracterizado pelo mosaico de fitofisionomias que variam desde formações campestres (campo limpo), formações savânicas (campo sujo, campo cerrado, cerrado típico) até as formações florestais (cerradão e mata de galeria), é considerado a savana mais biodiversa do planeta em termos de espécies de plantas (Oliveira & Marquis, 2002; Overbeck et al., 2015).

Os primeiros trabalhos sobre a distribuição de ecossistemas no Cerrado buscavam explicar a variação de tipos vegetacionais através da gênese geomorfológica (Eiten, 1972) e dos componentes edáficos como a fertilidade do solo (Goodland & Pollard, 1973). Posteriormente, trabalhos na área de ecologia do fogo demonstraram que as fisionomias do Cerrado e sua composição florística possuem estreita relação com a frequência e intensidade de distúrbios, sobretudo das queimas (L. M. Coutinho, 1982, 1990; Mistry, 1998). Desde então, trabalhos-chave têm apresentado evidências cada vez mais robustas a respeito da função ecológica e evolutiva do fogo nos ecossistemas de Cerrado (Bond et al., 2004; Pivello, 2011; Simon et al., 2009; Simon & Pennington, 2012).

Caracterizado pela alternância entre estações secas e chuvosas bem definidas nos últimos 2 milhões de anos, as espécies de plantas do Cerrado evoluíram adaptadas a ocorrência de queimas naturais ocasionadas por raios durante as tempestades de transição entre estações, especialmente no final da seca (William J. Bond & Keeley, 2005; Ramos-Neto & Pivello, 2000). Contudo, a domesticação do fogo pelo ser humano e seu emprego histórico com as mais diversas finalidades representaram para as savanas antropizadas uma nova fonte de ignição que não a natural, pelo qual o ser humano assumiu o papel de principal causador e supressor de queimas nestes ecossistemas (Goldammer & G., 1993). Atualmente o Cerrado encontra-se criticamente fragmentado e densamente povoado,

sendo que apenas 7% de sua área total encontra-se legalmente protegida (Soares-Filho et al., 2014). Este mosaico formado entre remanescentes naturais e os mais variados tipos de uso do solo agrava-se em locais como o estado de São Paulo onde apenas 0,84% do Cerrado está protegido e inserido em um contexto de forte pressão agrícola, pastoril e urbana (Alencar et al., 2020; Kronka et al., 2005), alterando o regime de fogo destas áreas (Archibald, 2016; Conciani et al., 2021).

Atualmente a reconstrução do histórico de queimas constitui base fundamental na decodificação de processos ecológicos, subsidiando a tomada de decisões e a elaboração de políticas públicas que levem em conta a história natural em nível ecossistêmico. Neste sentido, o sensoriamento remoto e a ciência da computação têm contribuído sobremaneira na consolidação de metodologias e na geração de produtos de áreas queimadas cada vez mais adequados aos diferentes contextos e aplicações (Bastarrika et al., 2014; Hawbaker et al., 2017; Ramo & Chuvieco, 2017).

Considerando o contexto do estado de São Paulo, os últimos remanescentes de vegetação nativa do Cerrado encontram-se protegidos em esparsas Unidades de Conservação (UC) sobre gestão do Instituto Florestal e da Fundação Florestal, ambos vinculados à Secretaria de Infraestrutura e Meio Ambiente do Estado de São Paulo. Compartilhando a mesma política de exclusão do fogo e brigadas de incêndio, essas UCs experimentam diferentes níveis de pressão antrópica, sendo o fogo relatado como uma das principais ameaças à sua conservação (Durigan et al., 2007), especialmente nas UCs de formações campestres.

Nesse sentido, através de uma abordagem inovadora, nós buscamos combinar os mais recentes avanços em classificação de imagens utilizando aprendizagem de máquina e o vasto acervo de imagens em média resolução espacial (30 metros) e temporal (16 dias) da série Landsat para criar um algoritmo, gerar e validar um produto de áreas queimadas para todo o Cerrado paulista entre 1985 e 2018 (capítulo 1). Esta abordagem permitiu-nos considerar as condições regionais de alta variação no uso e cobertura do solo, resolvendo alguns problemas detectados em outros produtos de áreas queimadas (como por exemplo os erros de comissão em áreas agrícolas e infraestrutura urbana e erros de omissão em áreas pequenas). No segundo capítulo, nós combinamos o nosso produto de áreas queimadas com os mapas anuais de uso e cobertura do solo (MapBiomas) e buscamos entender, como subsídio à conservação, os padrões temporais e

espaciais da ocorrência de fogo no Cerrado paulista, nas Unidades Conservação e em suas Zonas de Amortecimento.

### **OBJETIVO GERAL**

1. Determinar o regime de queimas no Cerrado do estado de São Paulo e analisar sua relação com o histórico e mudanças no uso e cobertura do solo.

### **OBJETIVOS ESPECÍFICOS**

1. Desenvolver um algoritmo para mapeamento de áreas queimadas.
2. Gerar e validar um produto de áreas queimadas para o Cerrado paulista.
3. Analisar os padrões espaciais e temporais do regime de fogo do Cerrado paulista, Unidades de Conservação e Zonas de Amortecimento.

### **REFERÊNCIAS BIBLIOGRÁFICAS**

- Alencar, A., Z. Shimbo, J., Lenti, F., Balzani Marques, C., Zimbres, B., Rosa, M., Arruda, V., Castro, I., Fernandes Márcico Ribeiro, J. P., Varela, V., Alencar, I., Piontekowski, V., Ribeiro, V., M. C. Bustamante, M., Eyji Sano, E., & Barroso, M. (2020). Mapping Three Decades of Changes in the Brazilian Savanna Native Vegetation Using Landsat Data Processed in the Google Earth Engine Platform. *Remote Sensing*, 12(6), 924. <https://doi.org/10.3390/rs12060924>
- Archibald, S. (2016). Managing the human component of fire regimes: lessons from Africa. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1696), 20150346. <https://doi.org/10.1098/rstb.2015.0346>
- Bastarrika, A., Alvarado, M., Artano, K., Martinez, M., Mesanza, A., Torre, L., Ramo, R., Chuvieco, E., Bastarrika, A., Alvarado, M., Artano, K., Martinez, M. P., Mesanza, A., Torre, L., Ramo, R., & Chuvieco, E. (2014). BAMS: A Tool for Supervised Burned Area Mapping Using Landsat Data. *Remote Sensing*, 6(12), 12360–12380. <https://doi.org/10.3390/rs61212360>
- Batalha, M. A. (2011). O cerrado não é um bioma. *Biota Neotropica*, 11(1), 21–24. <https://doi.org/10.1590/S1676-06032011000100001>
- Bond, W. J., Woodward, F. I., & Midgley, G. F. (2004). The global distribution of ecosystems in a world without fire. *New Phytologist*, 165(2), 525–538. <https://doi.org/10.1111/j.1469-8137.2004.01252.x>
- Bond, William J., & Keeley, J. E. (2005). Fire as a global ‘herbivore’: the ecology and evolution of flammable ecosystems. *Trends in Ecology & Evolution*, 20(7), 387–394.



<https://doi.org/10.1016/J.TREE.2005.04.025>

Conciani, D., Santos, L., Silva, T. S. F., Durigan, G., & Alvarado, S. T. (2021). Human-Climate Interactions Shape Fire Regimes in the Cerrado of São Paulo state, Brazil. *Journal for Nature Conservation*.

Coutinho, L. M. (1982). *Ecological Effects of Fire in Brazilian Cerrado* (pp. 273–291).

[https://doi.org/10.1007/978-3-642-68786-0\\_13](https://doi.org/10.1007/978-3-642-68786-0_13)

Coutinho, L. M. (1990). *Fire in the Ecology of the Brazilian Cerrado* (pp. 82–105). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-75395-4\\_6](https://doi.org/10.1007/978-3-642-75395-4_6)

Coutinho, Leopoldo Magno. (2006). O conceito de bioma. In *Acta Botanica Brasilica* (Vol. 20, Issue 1, pp. 13–23). <https://doi.org/10.1590/s0102-33062006000100002>

Durigan, G., & Ratter, J. A. (2016). The need for a consistent fire policy for Cerrado conservation. *Journal of Applied Ecology*, 53(1), 11–15. <https://doi.org/10.1111/1365-2664.12559>

Durigan, G., Siqueira, M. F. de, & Franco, G. A. D. C. (2007). Threats to the Cerrado remnants of the state of São Paulo, Brazil. *Scientia Agricola*, 64(4), 355–363. <https://doi.org/10.1590/S0103-90162007000400006>

Eiten, G. (1972). The cerrado vegetation of Brazil. *The Botanical Review*, 38(2), 201–341.

<https://doi.org/10.1007/BF02859158>

Goldammer, & G., J. (1993). Historical biogeography of fire: tropical and subtropical. *Fire in the Environment : The Ecological, Atmospheric, and Climatic Importance of Vegetation Fires.*, 297–314. <https://ci.nii.ac.jp/naid/10007236290/>

Goodland, R., & Pollard, R. (1973). The Brazilian Cerrado Vegetation: A Fertility Gradient. *The Journal of Ecology*, 61(1), 219. <https://doi.org/10.2307/2258929>

Hawbaker, T. J., Vanderhoof, M. K., Beal, Y. J., Takacs, J. D., Schmidt, G. L., Falgout, J. T., Williams, B., Fairaux, N. M., Caldwell, M. K., Picotte, J. J., Howard, S. M., Stitt, S., & Dwyer, J. L. (2017). Mapping burned areas using dense time-series of Landsat data. *Remote Sensing of Environment*, 198, 504–522. <https://doi.org/10.1016/j.rse.2017.06.027>

IBGE. (2019). *Mapa de Biomas do Brasil*.

<http://www.terrabrasil.org.br/ecotecadigital/index.php/estantes/mapas/563-mapa-de-biomas-do-brasil>

Kronka, F. J. N., Nalon, M. A., Matsukuma, C. K., Kanashiro, M. M., Ywane, M. S. S., Lima, L., Guillaumon, J. R., Barradas, A. M. F., Pavão, M., & Manetti, L. A. (2005). Monitoramento da

- vegetação natural e do reflorestamento no Estado de São Paulo. *Simpósio Brasileiro de Sensoriamento Remoto*, 12, 16–21.
- Mistry, J. (1998). Fire in the cerrado (savannas) of Brazil: an ecological review. *Progress in Physical Geography: Earth and Environment*, 22(4), 425–448.  
<https://doi.org/10.1177/030913339802200401>
- Oliveira, P. S., & Marquis, R. J. (2002). *The cerrados of Brazil : ecology and natural history of a neotropical savanna*. Columbia University Press. [https://books.google.com.br/books?hl=pt-BR&lr=&id=TXvY\\_kAFPU0C&oi=fnd&pg=PA159&dq=HOFFMANN,+William+A.%3B+MOREIRA,+Adriana+G.+The+role+of+fire+in+population+dynamics+of+woody+plants.+The+Cerrados+of+Brazil.+Ecology+and+Natural+History+of+a+Neotropical+Savanna,+p.+159-177,+2002.&ots=\\_HuLK8HL7q&sig=RjM-0dKdC1RiZYbG8xYOC47gWo4#v=onepage&q&f=false](https://books.google.com.br/books?hl=pt-BR&lr=&id=TXvY_kAFPU0C&oi=fnd&pg=PA159&dq=HOFFMANN,+William+A.%3B+MOREIRA,+Adriana+G.+The+role+of+fire+in+population+dynamics+of+woody+plants.+The+Cerrados+of+Brazil.+Ecology+and+Natural+History+of+a+Neotropical+Savanna,+p.+159-177,+2002.&ots=_HuLK8HL7q&sig=RjM-0dKdC1RiZYbG8xYOC47gWo4#v=onepage&q&f=false)
- Overbeck, G. E., Vélez-Martin, E., Scarano, F. R., Lewinsohn, T. M., Fonseca, C. R., Meyer, S. T., Müller, S. C., Ceotto, P., Dadalt, L., Durigan, G., Ganade, G., Gossner, M. M., Guadagnin, D. L., Lorenzen, K., Jacobi, C. M., Weisser, W. W., & Pillar, V. D. (2015). Conservation in Brazil needs to include non-forest ecosystems. *Diversity and Distributions*, 21(12), 1455–1460.  
<https://doi.org/10.1111/ddi.12380>
- Pivello, V. R. (2011). The Use of Fire in the Cerrado and Amazonian Rainforests of Brazil: Past and Present. *Fire Ecology*, 7(1), 24–39. <https://doi.org/10.4996/fireecology.0701024>
- Ramo, R., & Chuvieco, E. (2017). Developing a Random Forest Algorithm for MODIS Global Burned Area Classification. *Remote Sensing*, 9(11), 1193. <https://doi.org/10.3390/rs9111193>
- Ramos-Neto, M. B., & Pivello, V. R. (2000). Lightning Fires in a Brazilian Savanna National Park: Rethinking Management Strategies. *Environmental Management*, 26(6), 675–684.  
<https://doi.org/10.1007/s002670010124>
- Simon, M. F., Grether, R., Queiroz, L. P. de, Skema, C., Pennington, R. T., & Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences*, pnas.0903410106.  
<https://doi.org/10.1073/PNAS.0903410106>
- Simon, M. F., & Pennington, T. (2012). Evidence for Adaptation to Fire Regimes in the Tropical Savannas of the Brazilian Cerrado. *International Journal of Plant Sciences*, 173(6), 711–723.  
<https://doi.org/10.1086/665973>
- Soares-Filho, B., Rajão, R., Macedo, M., Carneiro, A., Costa, W., Coe, M., Rodrigues, H., & Alencar, A. (2014). Cracking Brazil's forest code. *Science*, 344(6182), 363–364.

## **Developing a machine learning based algorithm for regional time-series burned area mapping: The highly anthropized Cerrado challenge**

Dhemerson E. CONCIANI <sup>1\*</sup>; Swanni T. ALVARADO <sup>2</sup>; Thiago S. SILVA <sup>3</sup>

<sup>1</sup> Universidade Estadual Paulista (UNESP), Instituto de Biociências, Departamento de Ecologia, Avenida 24-A 1515, 13506-900, Rio Claro, Brazil. \*[dhemerson.conciani@unesp.br](mailto:dhemerson.conciani@unesp.br)

<sup>2</sup> Universidade Estadual do Maranhão, Programa de Pós-Graduação em Agricultura e Ambiente, Praça Gonçalves Dias, s/n, 65800-000 Balsas, Brazil

<sup>3</sup> Biological and Environmental Sciences, Faculty of Natural Sciences, Stirling University, Stirling, FK9 4LA, UK

## ABSTRACT

This study aims to develop a regional burned area (BA) algorithm for Landsat surface reflectance (SR) images by testing different machine learning (ML) algorithms. Three ML algorithms (RF, XGB, MARS) were fed and tuned by using more than 1 million of spectral signatures of BA and anthropic land-uses from a balanced dataset. As predictors, we used both SR bands and spectral indexes. Different combinations of hyperparameters were tested, being the optimal values selected by using the largest accuracy. RF overcome XGB and MARS, presenting a balanced accuracy of 98%. Validation was made by using the RF model to predict 59 scenes. RF model alone was not sufficient to generate a BA product with suitable quality ( $\kappa= 0.53$ ), thus, post-processing was implemented. Higher accuracy ( $\kappa= 0.79$ ) was obtained by combining infrastructure and terrain masks with a spatial contiguity filter. Balancing of errors prioritized a higher omission ( $OE= 0.16$ ) than commission ( $0.09$ ), guarantying that this product can be applied to perform regional analysis without overestimating the BA. Finally, this study launches the first Cerrado's collaborative burned area mapping platform, a simple and intuitive way to share the result with the community and take feedbacks to improve the product quality in the future.

**Key- words:** burned area; random forest; Landsat; landscape; land-cover; land-use.

## 1. Introduction

The Cerrado vegetation covers an area of c.a 2 million km<sup>2</sup>, about 25% of Brazilian territory (DURIGAN; RATTER, 2016). The Cerrado, the largest South American savanna, evolved under natural fire regimes (SIMON et al., 2009). However, contemporary fire regimes are highly affected by the established settlements, managing landscape for agriculture and livestock, thus changing the natural fire regimes according with to the local cultural and economic practices (DIAS, 2006). These changes in natural fire regimes can completely alter the ecosystem's structure, composition and functionality. Increasing fire frequency is often related to the conversion of savannas into pastures for cattle grazing and opening of new croplands such as soybean or sugar cane (DALDEGAN et al., 2014), while decreasing fire frequency is specially observed in areas with woody encroachment (ROSAN et al., 2019).

Currently, 93% of the entire Cerrado is under an anthropic matrix (SOARES-FILHO et al., 2014) covered mainly by agriculture, livestock and forestry (ALENCAR et al., 2020). In the state of São Paulo, anthropic land-use is observed in more than 95% of the Cerrado's original cover (KRONKA et al., 2005). The presence of highly populated cities ( $\Sigma$ = 44 million inhabitants (IBGE, 2014) and dense infra-structure (eg. roads, railways, power transmission lines and telecommunication towers) are added to the Cerrado of São Paulo's landscape, increasing its complexity. For this reason, fire has been considered as a destructive force and legally banned (State Law 10 547/ 2001) to decrease fire use and mitigate its impacts on human populations. Despite the Cerrado being a fire prone ecosystem, private companies and government agencies maintain fire brigades to extinguish any fire occurrence, not mattering its cause or location (DURIGAN; RATTER, 2016).

When monitoring fire regime changes it is important to track where, when and how much vegetation is being affected by fire, ensuring that spatial features of the fire regime can be assessed on fire policy reviews and included on future studies of modeling, species distribution, botany, etc. Current Remote Sensing products derived from MODIS have been showing good results in evaluating natural and anthropic processes affecting land surface, allowing the global detection of active fires (Active Fire Data | MCD14DL, 1 x 1 km) and burn scars (LP DAAC - MCD64A1, 500 x 500m) since the 2000's. However, these products are not suitable to analyze regional and

local patterns due to their low spatial resolution, especially in highly fragmented landscapes such as the state of São Paulo. The program “INPE Queimadas” produced burned area products based on Landsat images (AQM30, 30 x 30 m) for the entirety of Cerrado, but these products have not advanced beyond the beta phase and are limited to the 2011-2018 period, leaving a gap for the reconstruction of larger and more reliable fire regime historical series.

Traditional methods for time-series burned area classification by using moderate resolution sensors (like Landsat) are mainly based on reflectance ratios between fire sensitive bands (NIR, SWIR<sub>1</sub>, SWIR<sub>2</sub>) and spectral index variations (NBR, CSI, NDVI) (BASTARRIKA et al., 2014; HAWBAKER et al., 2017; KOUTSIAS; KARTERIS, 2000). These approaches present satisfactory results on homogeneous landscapes, but accuracy errors are not balanced to consider open and heterogeneous landscapes like the Cerrado. When considering highly managed landscapes exposed to constant land-use and land-cover changes (LULCC) the accuracy is highly compromised, making these products not suitable to assess ecosystem fire regime features.

Recent advancements on machine learning algorithms and open source libraries present a new opportunity to explore potential applications on automated and semi-automated burned area mapping (PEREIRA et al., 2017; RAMO; CHUVIECO, 2017). Thus, we tested potential applications of different machine learning algorithms (eXtreme Gradient Boosting, Multivariate Adaptive Regression Spline, Random Forest) to reconstruct the contemporary fire regime of São Paulo’s Cerrado by using Landsat time-series data (TM, ETM+, OLI). We trained and tuned classification models by using spectral signatures of burned areas and LULCC’s from six Cerrado’s protected areas and their respective buffer zones (7 km). By using the largest accuracy value to select the optimal model, we applied it into a dense Landsat time-series and generated a standardized burned area product from 1985 to 2018 for the highly anthropized Cerrado. We performed the validation of this product by considering an independent multi-temporal burned area dataset and an adaptative post-processing routine.

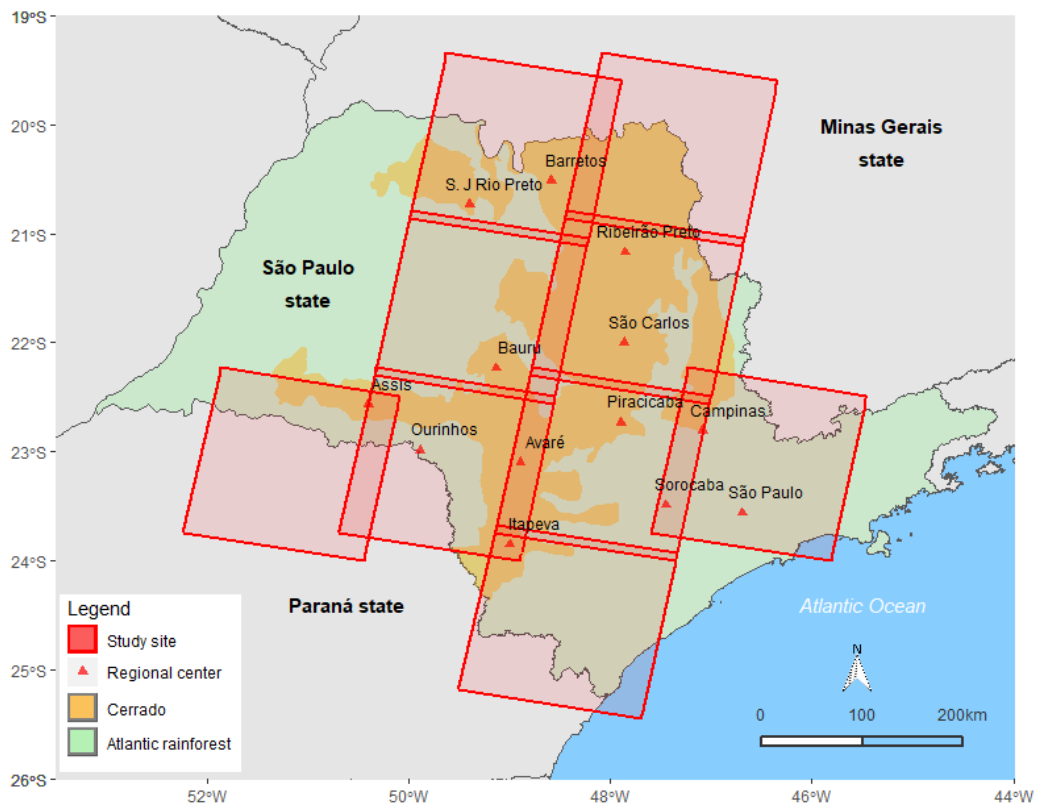
Before beginning, we theorized: i) tuning the hyperparameters will affect the accuracy performance; ii) random forest and extreme gradient boosting will outperform multivariate adaptive regression spline; iii) mask of some LULCCs will be

need to reach an acceptable product quality; iv) the final will be of sufficient quality to carry out environmental analysis on regional scale.

## 2. Methods

### 2.1. Study area

We focused on generating an accurate burned area product for highly anthropized Cerrado considering the São Paulo state land cover context. For this, we selected 9 WRS Landsat paths/rows covering an area of 228 776 km<sup>2</sup> (Figure 1). Parts of other states were included when sharing the same scene as our target sites. Thus, the total covered area by this study can be divided into 69% covering São Paulo state, 15% southern Minas Gerais state, 14% northern Paraná state and 2% Atlantic Ocean, the last covering small islands on São Paulo's coast.



**Figure 1.** Study area covered by Landsat Burned Area product for the Cerrado of São Paulo state

The population for the study area was estimated around 38,3 million inhabitants in the 2010 census, distributed along 487 municipalities, representing almost 87% of São Paulo's state population and 20% of Brazil's population (IBGE, 2014). São Paulo presents the highest GDP (Gross Domestic Product) from Brazil,

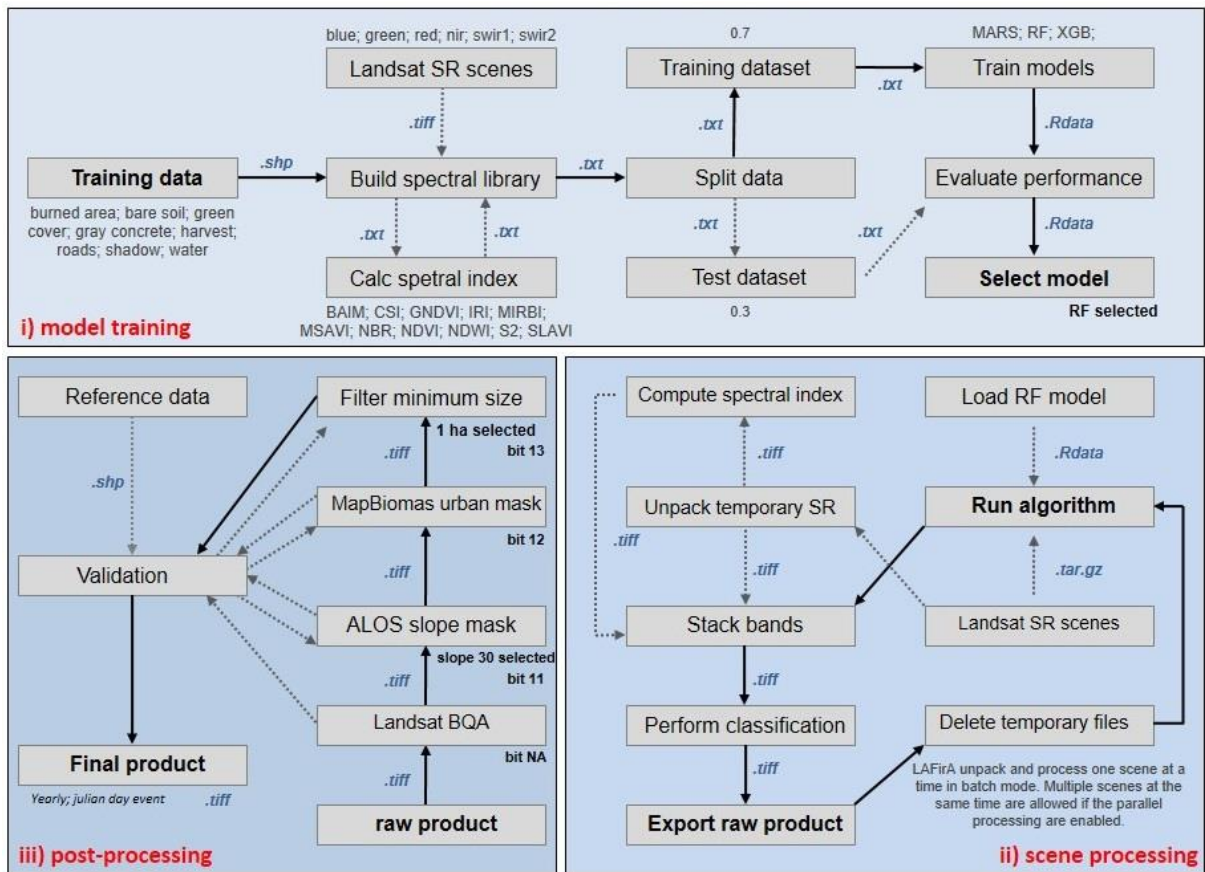
where the countryside is responsible by primary production (mainly sugar-cane, soybean, coffee and forestry) while the main cities rely heavily on services and heavy industries, concentrating a high population compared to neighboring cities (IBGE, 2014).

Remnants of primary native vegetation in São Paulo state are characterized by 15.7% total state area of Atlantic Rainforest and ~1% of Cerrado (ATLÂNTICA, 2017). These native remnants of both vegetation types are highly fragmented and these patches are located mainly in protected areas (KRONKA et al., 2005). The Cerrado remnants are mainly dominated by forest-like formations (locally known as “cerradão”), and only very few areas of open physiognomies (locally known as “campo limpo”, “campo sujo”), which are extremely rare and restricted to small patches into the anthropized matrix of the countryside (VICENTE; SOUZA FILHO; PEREZ FILHO, 2005).

## *2.2. Algorithm workflow*

We developed an automatic algorithm to detect burned areas in the highly anthropized Cerrado. First, we trained different classification models based on machine learning algorithms and assessed prediction performance of each one by using the balanced accuracy and kappa index. Second, the best fitted model was applied to classify a dense Landsat time-series. Finally, we balanced commission and omission errors in the burned area to ensure that the final product can be used to perform regional scale analysis. All the processing steps briefly described in this section were represented in the Figure 2 and will be detailed in the following sections.





**Figure 2.** Algorithm graphical abstract. Blue boxes show a set of one-step processes. Red labels show step titles. Gray boxes represent each individual process. Black bold labels inner gray boxes show start/end processes from a step. Black solid arrow points primary flux of processes, while gray arrow indicates secondary processes that occur in the background and feed primary processes. Gray labels offer a short description in specific boxes. Black bold labels outside gray boxes points selected parameters/setup. Blue labels points file extensions expected as input and exported as output.

All processing steps were parallelized to take advantage of multicore CPUs and made using R (RCT, 2020). Specific key-steps were accomplished by using community packages “caret”, “raster” and “rgdal” (BIVAND et al., 2015; HIJMANS; VAN ETEN, 2012; KUHN; JOHNSON, 2013) and jscript implementations into Google Earth Engine. Source codes are available and can be accessed in GitHub ([https://github.com/musx/FireGIS\\_SP](https://github.com/musx/FireGIS_SP)). The computation infrastructure used was core i7 5820K 3.3 GHz CPU, 64GB RAM and a GTX 1060 6GB GPU.

### 2.3. Building the spectral library

We previously selected training sites that contain representative sample areas of native vegetation remnants and anthropic land uses. These sites correspond to six protected areas being three of full preservation (Assis, Santa Bárbara and Itirapina

Ecological Stations) and three of sustainable use (Assis, Santa Bárbara and Itirapina State Forests). We also considered buffer zones of 7 kilometers around each one of these protected areas. A highly accurate and manual burn scar mapping dataset was already available for these areas from 1984 to 2016 (Conciani et al in press). This previous mapping was performed based on visual detection and manual delineation of every burn scar detected into 805 Landsat surface reflectance Level-2 scenes from *Earth Resources Observation and Science- Center Science Processing Architecture* (EROS-ESPA, <https://espa.cr.usgs.gov/>) for the WRS path/rows 220/75, 221/76, 222/76.

In order to create a diverse spectral library, we mapped samples over time of land uses with similar spectral signature when compared to burn scars (Table 1). Furthermore, spectral signature of generic land covers (e.g. “green cover”, “bare soil”) was also mapped in order to train a landscape classifier with ability to recognize burn scars on highly anthropized areas.

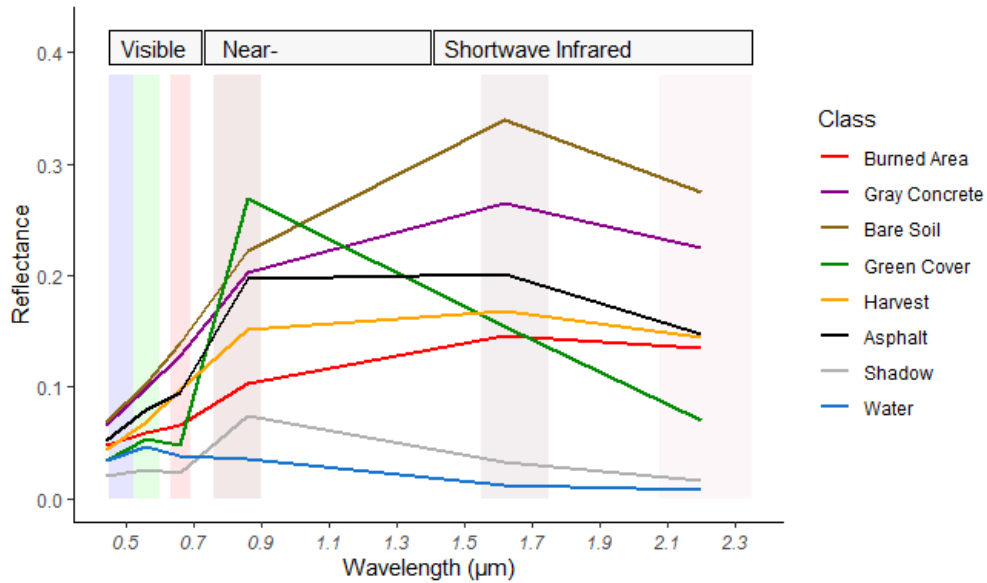
**Table 1.** Mapped classes to train classification models. A byte value was assigned for each class in order to identify these elements in further proceedings.

Byte	Class	Description
1	Burned area	Recently burned area, with ash presence
2	Bare soil	Soil without any type of vegetation cover
3	Green cover	Any type of green cover, forests, agriculture, pastures
4	Gray concrete	Impermeable structures, cities
5	Harvest	Recent harvest with the presence of decomposing organic matter atop the soil
6	Asphalt	Highways and paved streets
7	Shadow	Cloud and relief shadows
8	Water	Natural/ artificial water courses and water masses

We performed the spectral signature extraction from surface reflectance bands (Table 2) that matches between the mapped vectors by class and the Landsat images for each date. The data extracted in the process was compiled and exported as a database, being used to build our spectral library. A graphical summary of the spectral library is presented in the Figure 3 considering the mean reflectance value for each one of the classes.

**Table 2.** Surface reflectance spectral bands used to extract spectral signatures. TM = Thematic Mapper (Landsat 5); ETM+ = Enhanced Thematic Mapper Plus (Landsat 7); OLI = Operational Land Imager (Landsat 8).

Spectral band	Landsat TM and ETM+		Landsat OLI	
	Band number	Bandwidth ( $\mu\text{m}$ )	Band number	Bandwidth ( $\mu\text{m}$ )
Blue	1	0.45 - 0.52	2	0.45 - 0.51
Green	2	0.52 - 0.60	3	0.52 - 0.60
Red	3	0.63 - 0.69	4	0.63 - 0.69
NIR	4	0.77 - 0.90	5	0.77 - 0.90
SWIR <sub>1</sub>	5	1.55 - 1.75	6	1.55 - 1.75
SWIR <sub>2</sub>	7	2.09 - 2.35	7	2.09 - 2.35



**Figure 3.** Mean reflectance (y-axis) over the wavelengths (x-axis) from each class present in our spectral library. Line colors represent different classes (described on legend). Background colors represent Landsat bands ordered by wavelength (Blue, Green, Red, NIR, SWIR<sub>1</sub> and SWIR<sub>2</sub>).

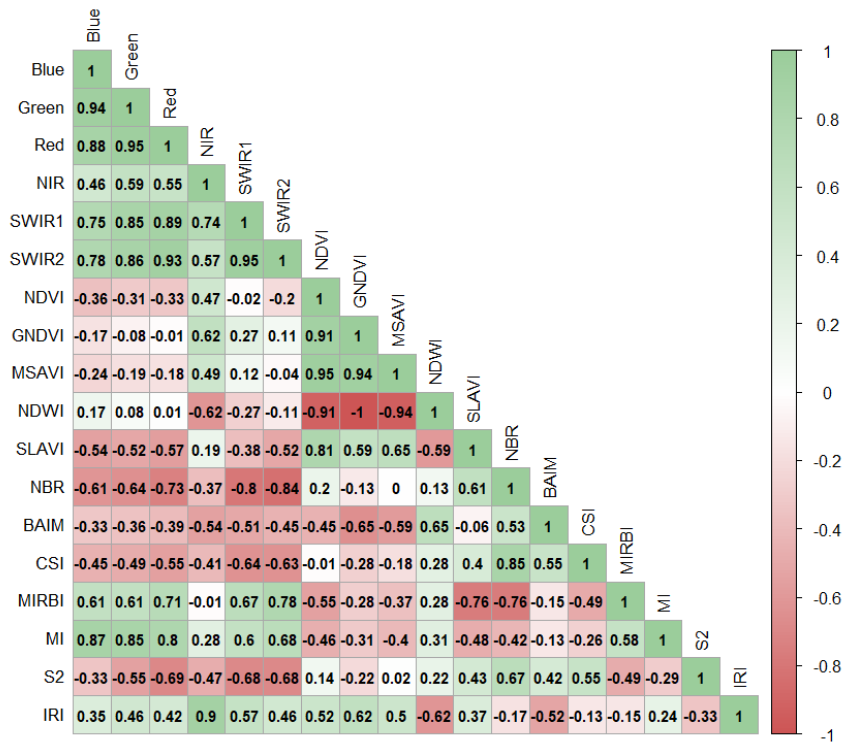
Finally, we used these surface reflectance database as input for the generation of several spectral indices (Table 3). We selected some of the most commonly used indexes in the literature to assess features from the burn scars, vegetation, soil and water, and included them in our spectral library.

**Table 3.** Spectral indexes generated to enhance our spectral library. The  $\lambda$  symbol represents the reflectance value of the spectral band.

Spectral Index	Reference	Formula
Burned Area Index (BAIM)	(MARTÍN; CHUVIECO, 2006)	$\frac{1}{(0.05 - \lambda \text{ NIR})^2 + (0.2 - \lambda \text{ SWIR1})^2}$
Char Soil Index (CSI)	(SMITH et al., 2005)	$\frac{\lambda \text{ NIR}}{\lambda \text{ SWIR1}}$
Green Normalized Difference Vegetation Index (GNDVI)	(GITELSON; MERZLYAK; LICHTENTHALER, 1996)	$\frac{\lambda \text{ NIR} - \lambda \text{ Green}}{\lambda \text{ NIR} + \lambda \text{ Green}}$

Infrared Index (IRI)	(Hardisky et al., 1983)	$\sqrt{\frac{\lambda \text{ NIR}^2 + \lambda \text{ SWIR}2}{\lambda \text{ SWIR}1}}$
Mid-Infrared Bispectral Index (MIRBI)	(TRIGG; FLASSE, 2001)	$10 \times \lambda \text{ SWIR}1 - 9.8 \times \lambda \text{ NIR} + 2$
Modified Soil Adjusted Vegetation Index (MSAVI)	(QI et al., 1994)	$\frac{\lambda \text{ NIR} + 0.5 - (0.5}{\times \sqrt{(2 \times \lambda \text{ NIR} + 1)^2 - 8 \times \lambda \text{ NIR} - (2 \times \lambda \text{ Red})}}$
Normalized Burn Ratio (NBR)	(KEY; BENSON, 2006)	$\frac{\lambda \text{ NIR} - \lambda \text{ SWIR}1}{\lambda \text{ NIR} + \lambda \text{ SWIR}1}$
Normalized Difference Vegetation Index (NDVI)	(ROUSE; HAAS; DEERING, 1974)	$\frac{\lambda \text{ NIR} - \lambda \text{ Red}}{\lambda \text{ NIR} + \lambda \text{ Red}}$
Normalized Difference Water Index (NDWI)	(GAO, 1996)	$\frac{\lambda \text{ Green} - \lambda \text{ NIR}}{\lambda \text{ Green} + \lambda \text{ NIR}}$
Salinity Index 2 (S2)	(DOUAOUI; NICOLAS; WALTER, 2006)	$\frac{\lambda \text{ Blue} - \lambda \text{ Red}}{\lambda \text{ Blue} + \lambda \text{ Red}}$
Specific Leaf Area Vegetation Index (SLAVI)	(Lymburner et al., 2000)	$\frac{\lambda \text{ NIR}}{\lambda \text{ Red} + \lambda \text{ SWIR}2}$

A Spearman's correlation map was computed (Figure 4) to inspect the relationships between the surface reflectance bands and the generated spectral indexes.

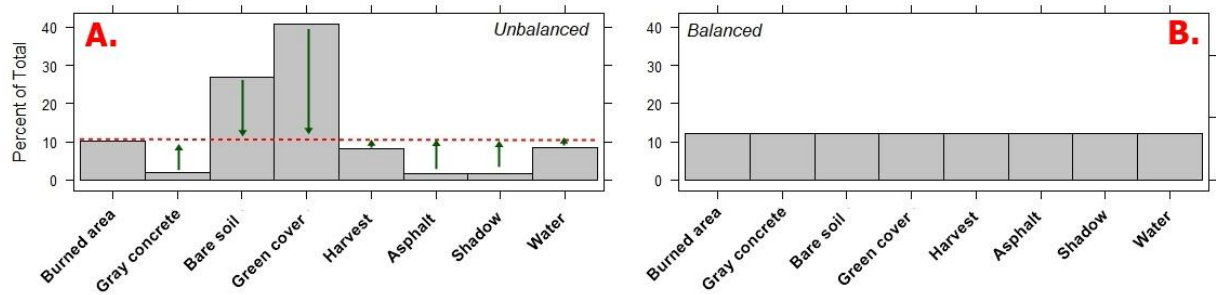


**Figure 4.** Correlation map between the surface reflectance bands and the spectral indexes in our spectral library. Right bar color vary between strong negative correlations (red) to strong positive correlations (green). Black number inside each square shows de Spearman's correlation (positive or negative).

#### 2.4. Pre-processing

We computed a count of spectral signatures for each class in our spectral library and plotted them into a histogram (Figure 5A). We detected unbalanced observations per class on our database. As a remedy to prevent learning bias, we artificially balanced the frequencies by using the burned area as reference and performing a random down-sampling for classes with higher frequency than the reference (ignoring cases from the majority) and an up-sampling for classes with less frequency than the reference (replicating cases from the minority) (Figure 5B). Thus, we generated a balanced dataset containing 1,153,040 spectral signatures, being 144,130 (12.5%) of each class. This balancing strategy is described in the literature as an alternative to prevent the learning overwhelm by the majority classes (GUO et al., 2008; PROVOST, 2000). Furthermore, more accurate performances were reported to classifiers trained with balanced datasets when comparing them to classifiers trained with the original data (BATISTA; PRATI; MONARD, 2004;

JEATRAKUL; WONG; FUNG, 2010; VAN HULSE; KHOSHGOFTAAR; NAPOLITANO, 2007)



**Figure 5.** Histogram of frequencies by class. **A.** Original dataset. Red line indicates burned area frequency used as reference to balance other classes. Green arrows points if an up-sampling (up arrow) or down-sampling (down arrow) were performed to balance each class. **B.** Balanced dataset after up-sampling and down-sampling.

Following the most common proportions reported in the literature (KUHNN; JOHNSON, 2013), we divided the balanced dataset into training dataset by creating a stratified partition with 70% of the data and an test dataset by using the 30% of the remaining data. We centered and scaled the numeric data to take standard deviation one and mean of zero for all the predictors.

### 2.5. Model training and testing

Using the training dataset as input, we implemented machine learning algorithms considering the scope of non-parametric regressions (Multivariate Adaptive Regression Spline - MARS), decision trees (Random Forest – RF) and boosted trees (eXtreme Gradient Boosting – XGB). Each one of these algorithms have specific parameters that affects the model's accuracy and which cannot be estimated by using the dataset (Table 4). Since there is no analytical formula available to calculate an appropriate value, these parameters are referred as *tuning parameters* or *hyperparameters*. Since these hyperparameters control the model complexity, poor choices for the inputted values can result in low accuracy or over-fitting (KUHNN; JOHNSON, 2013). In this way, following the adaptative search method described in Olsson & Nelson, 1975, we defined a set of candidate values for each hyperparameter (Table 4). Finally, to avoid the over-fitting, we used the k-fold cross-validation resampling technique (k= 5, repeats =3) for training and estimating the performance of the models by considering all the possible combinations between the candidate values.

**Table 4.** Hyperparameters description for each algorithm. Numbers following the names of the algorithms refer to the version of the R package that has been implemented. The range column represent the minimum and maximum allowed values for each hyperparameter. The candidate values column represents the set of values that we used as input to train and evaluate different models accuracy.

Algorithm	Hyperparameter	Description	Range	Candidate values
earth 5.1.2	degree	Product degree	1 – Inf	1 – 3
	nprune	Number of terms	1 – Inf	1 – 20
RandomForest 4.6-14	ntree	Number of trees to grow	1 – Inf	1 – 750
	mtry	Number of variables randomly sampled as candidates at each split	1 – $n(\beta)$	2 – 8
xgboost 0.90.0.2	nrounds	Number of boosting iterations	1 – Inf	50 – 150
	max_depth	Max tree depth	0 – Inf	1 – 3
	eta	Shrinkage	0 – 1	0.3 – 0.4
	gamma	Minimum loss reduction	0 – Inf	$D_0$
	subsample	Subsample percentage	0 – 1	0.5 – 1
	colsample_bytree	Subsample ratio of columns	0 – 1	0.6 – 0.8
	min_child_weight	Minimum sum of instance weight	0 – Inf	$D_1$

*Inf = Infinite;  $n(\beta)$  = number of predictors;  $D_{0,1}$  = default hyperparameter value (0 and 1 respectively)*

We computed and used the largest values of overall accuracy (ACC – eq. 1) and the Cohen’s Kappa index (Kappa – eq. 2) obtained in the training by the k-fold cross validation to select the best values for the hyperparameters as well as the optimal model trained by each algorithm. Then, we used these three finalist models (one by algorithm) to predict the test dataset and assessed the performance of each one by computing a confusion matrix comparing the predicted classes vs. reference classes. Once again, we used the largest accuracy value obtained by the test dataset classification to select the final model used in this article.

$$ACC = \frac{\sum TP + \sum TN}{n}$$

eq.1

TP= true positive; TN= true negative; n= total population

Kappa

$$= \frac{\left[ (\sum TP + \sum TN) - ((\sum TP + \sum FN) \times (\sum TP + \sum FP) + (\sum FP + \sum TN) \times (\sum FN + \sum TN)) / n \right]}{\left[ n - ((\sum TP + \sum FN) \times (\sum TP + \sum FP) + (\sum FP + \sum TN) \times (\sum FN + \sum TN)) / n \right]}$$

eq.2

FN= false negative; FP = false positive

## 2.6. Burned area extraction into a dense time-series

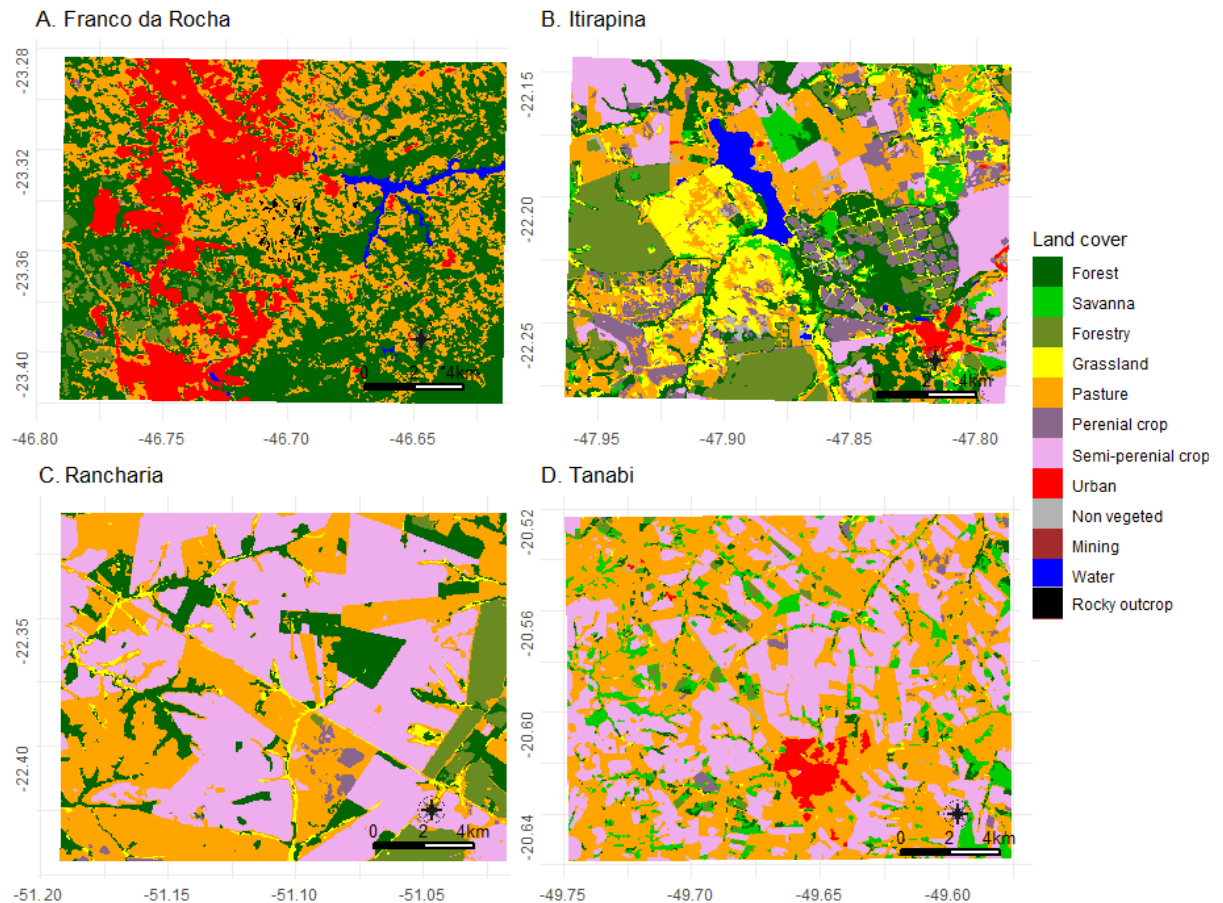
Considering the study site extension, we retrieved the metadata for all the available Landsat scenes from Earth Explorer (<https://earthexplorer.usgs.gov/>). Assessing scene availability according cloud cover percentage (Supplementary Fig S1), we found that a maximum of 75% cloud cover is the more suitable threshold for this study. Thus, we discarded all the scenes with more than 75% of cloud cover, preventing unnecessary processing caused by scenes with high cloud cover (NoData). Considering the metadata of processing level, we discarded scenes classified into L1GT (without precision correction) and L1GS (without terrain correction), using all remaining scenes available in the L1TP level (precision, terrain, geometric and radiometric corrections). We used this filtered list to build a request containing 4180 Level-2 scenes (surface reflectance) from 1985 to 2018 and downloaded them from Earth Resources Observation and Science- Center Science Processing Architecture (EROS-ESPA, <https://espa.cr.usgs.gov/>). Using the images, we calculated the same spectral indexes used to build our spectral library (Table 3) and stacked them as different bands into each one of the downloaded scenes.

For each scene, we used the final model to run per-pixel burned area and LULCC classification, being a value from 1 to 8 associated to each pixel as a result. These values from 1 to 8 correspond to the byte code of the predicted class (Table 1). Finally, using these classified scenes as input, we performed the binarization of the burned area class (1= burned area, 0= unburned – all other classes from 2 to 8). These binarized burned area data were written in new raster files containing the same metadata as the original scenes.

### *2.7. Burned area validation*

Validation is the term used to refer to the process of assessing the accuracy of a product by comparing with an independent reference data (ROY; BOSCHETTI, 2009). In the context of this article, we used the selection of representative places in space and a random design in time to assess the transferability of the classifier to regions outside the training data scope. In this way, we established four plots of 270 km<sup>2</sup> each (15 x 18 km) in different path/rows considering land-cover and land-use variations in São Paulo state (Figure 6, Table 5) and generated an independent validation dataset by performing the manual vectorization of burned areas over 59 cloud-free scenes across four random years (Supplementary Table S2).





**Figure 6.** MapBiomas land cover for the validation plots in the year of 2018. **A.** Franco da Rocha – path 219/ row 76; **B.** Itirapina – 220/75; **C.** Rancharia – path 222/76 and; **D.** Tanabi – path 221/74.

**Table 5.** Landscape description and considered years for validation in each site.

Validation plot	Years	Description
<b>Franco da Rocha</b>	1995	Densely populated suburban area inserted in São Paulo capital city urban zone (> 1 million inhabitants). This area presents highly rugged relief mainly covered by “Serra do Mar” Atlantic rainforest. However, few pasture areas are observed on the landscape and open Cerrado “campo sujo” fragments occur in the Juquery state park.
	2003	
	2017	
	2018	
<b>Itirapina</b>	1985	This area represents the biggest open Cerrado “campo limpo” and “campo sujo” remnants of São Paulo state, located at the Itirapina Ecological Station (~2200 ha). Outside the protected area, the landscape is dominated by cattle grazing, forestry and sugar-cane plantations. Some wetlands divide space with a rich drainage system, small urban zones (< 20 000 inhabitants), highways and railways.
	1988	
	2015	
	2018	

<b>Rancharia</b>	1985	This area corresponds to a transition between Cerrado and Atlantic rainforest. The validation plot includes a rural zone dominated by semi-perennial croplands of bean, soybean and corn. Small fragments of Atlantic rainforest remnants are maintained by farmers as a legal requirement by the National Forest Code (National Law 12 651 / 2012).
	2001	
	2017	
	2018	
<b>Tanabi</b>	1995	Regional hub in sugar and ethanol industrial production. Landscape is dominated by sugar-cane croplands with small rivers and some riparian forests. As well as the Rancharia area, here there are small fragments of “cerradão” and Atlantic rainforest maintained under legal requirement while the Cerrado area has been converted into pastures. Tanabi’s urban zone (< 25 000 inhabitants) was included in the validation plot.
	2006	
	2016	
	2018	

We used the date metadata (yyyy-mm-dd) to match and overlap the manually mapped vectors and the binarized rasters of burned area generated by our algorithm. Then, a confusion matrix was computed to compare each one of the spatial and temporal matches; in other words, we estimated and stored in a database the kappa index (eq. 2), omission error (OE – eq. 3) and commission error (CE – eq. 4) for each one of the comparisons between references vs. predictions. We used this database to calculate the mean value of these metrics for each one of the validation plots (eq. 5) and assumed the result as a representative value of the spatio-temporal product quality. Furthermore, the quality assessment of the product provided by the validation was used to delineate the post-processing routine in order to improve the product accuracy.

$$OE = \frac{\sum FN}{n(R)}$$

eq.3

FN= false negative; n= total population; R= reference

$$CE = \frac{\sum FP}{n(P)}$$

eq.4

FP= false positive; P= predicted

$$\bar{x} = \frac{\sum_{i=1}^n x_i (Kappa; OE; CE)}{n(P \sim VP)}$$

Pr= predicted; ~ in each; VP= validation plot

eq.5

## 2.8. *Post-processing*

As a standard procedure, we used the date metadata to match and mask (into NoData) any pixels detected as radiometric saturation, cloud, cloud shadow and water in the binarized burned area product by using the Landsat Quality Assessment Band (QA). For precaution, we applied restrictive thresholds in the QA parameters, being masked all the pixels (and also their adjacencies) that presented any of the previous described anomalies, independently of the confidence level.

Considering that our gray concrete spectral signatures is mainly composed by urban-zones and sparse buildings, since our aim is to generate a product to assess the ecosystem fire patterns, we decided to mask all the urban-zones. In this way, the MapBiomas Brasil project (collection 4.1) offers an accurate yearly classification of land-cover for the Cerrado by performing Landsat scenes classification (ALENCAR et al., 2020). Thus, we used the "urban-infrastructure" class from the MapBiomas products to mask our binarized burned areas. Besides that, we also used the "mining", "beach" and "rocky-outcrop" MapBiomas classes (employing our empirical knowledge that these classes don't burn) to mask our data.

Due the earth's movements (e.g. rotation, translation) the sun-earth inclination angle change across the seasons, and so the extent of the mountain shadows projected over the land surface accompany this variation (GILES, 2001). Previous studies focused in scene classifications have reported that the spectral mixture caused by the projection of mountain shadows over the surface of highly sloped areas can induce several misclassifications (CHEN et al., 2018; GIGLIO et al., 2015; PAUL, 1997). In this way, we obtained the AWD3D30 v1.1 ALOS Digital Surface Model (Jaxa, 2020) from the Google Earth Engine library and derived the terrain slope for São Paulo state. We binarized slope rasters (1= slope greater than  $x$ , 0= slope less than  $x$ ) by considering different slope thresholds ( $x= 10^\circ, 20^\circ, 30^\circ$  and  $40^\circ$ ) and tested how the terrain masking can improve or degrade the product quality of the burned area product in the context of this study.

Finally, to improve the consistency and the product quality, we assume that isolated pixels classified as burned area (without neighbor pixels classified as burned area) have a great chance to be misclassifications. To test this premise we implemented a "minimum spatial contiguity" filter based in the count of pixels classified as burned area that share their borders. Thus, we tested the effects in the

product accuracy by masking burned area pixel aggregations less than 5 pixels (0.45 ha), 11 pixels (0.98 ha) and 16 pixels (1.44 ha) and compared them to the product without spatial contiguity filter (considering alone pixels as valid burned areas).

### 2.9. *Final product compilation*

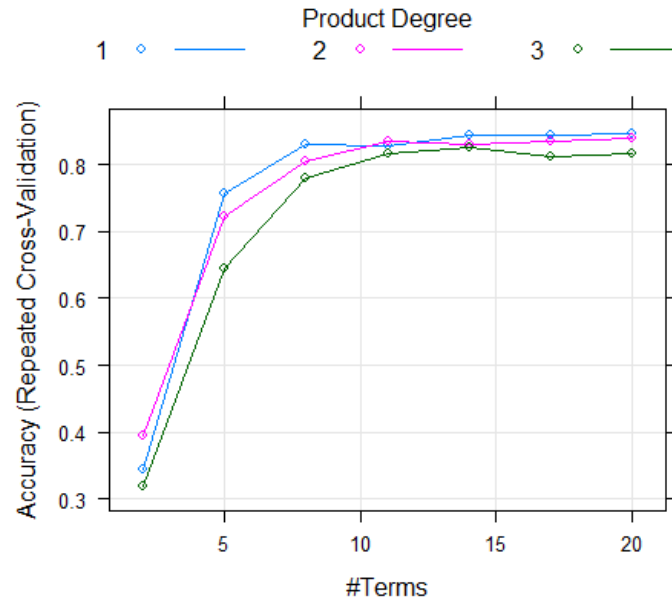
After applying the post-processing steps and finding the best parameters for the masks and filters by balancing the omission and commission errors, we retrieved the gregorian date (yyyy-mm-dd) for each year. Final product was serialized by year ~ path/row and resulted in a library of 309 raster files in .tif format. The file names were built to store the burned area product metadata as text strings, being: i) the WRS-2 path [path] and WRS-2 row [row] as spatial descriptors inside the same separator [path+row]; ii) the gregorian year as temporal descriptor [yyyy] and; iii) the abbreviation of a short product description and the version number [jdba1], equivalent to "julian day of burned area detection, version 1". This metadata were compiled so that the final file names presented the format "pathrow\_yyyy\_jdba1.tif" (e.g. 22075\_1985\_jdba1.tif, 21976\_2018\_jdba1.tif, etc.).

## 3. Results and Discussion

### 3.1. *Hyperparameters tuning and model selection*

#### 3.1.1. *Multivariate Adaptive Regression Spline – MARS*

We detected a positive effect of the hyperparameter maximum number of terms (nprune) in the models accuracy. Lower accuracies were observed by using low nprune values (ACC ranging from 0.324 to 0.402 when nprune = 2), independently of the product degree (degree). As new MARS models have been trained by increasing the number of terms, we observed strong gains in the accuracy until the nprune value = 15 (ACC ranging from 0.813 to 0.845) (Figure 7). This result points that despite the number of terms having largely contributed to the accuracy gain, this gain has a tendency to saturate since a threshold, being all the addition of complexity since this point responsible for a overwhelm of the model. This saturation pattern was statistically and empirically demonstrated by Kuhn & Johnson, 2013 and also reported as result from other studies that have tuned MARS models to make forecasts (FERLITO; ADINOLFI; GRADITI, 2017; LI et al., 2019).



**Figure 7.** Multivariate Adaptive Regression Spline (MARS) hyperparameters tuning. Colored lines represent models considering different product degrees. The number of terms hyperparameter (*nprune*) was represented in the x-axis while the y-axis points the models accuracy.

On the other hand, although in small proportion in relation to the total number of terms, the product degree has affected the models accuracy. While the addition of terms has induced accuracy gains, the addition of product degrees has negatively affected the models by degrading the accuracy. In other words, independently of the number of terms in the model, greater values of accuracy were observed when only first order interactions between variables are allowed (degree = 1), while less accuracy was found when increasing the product degree. The combined effects of both hyperparameters can be easily observed in Figure 8, since the results showed a graphical ordered pattern. Finally, the largest accuracy value (ACC= 0.845, Kappa = 0.823) was used to select the optimal MARS model (*nprune* = 20 and degree = 1).

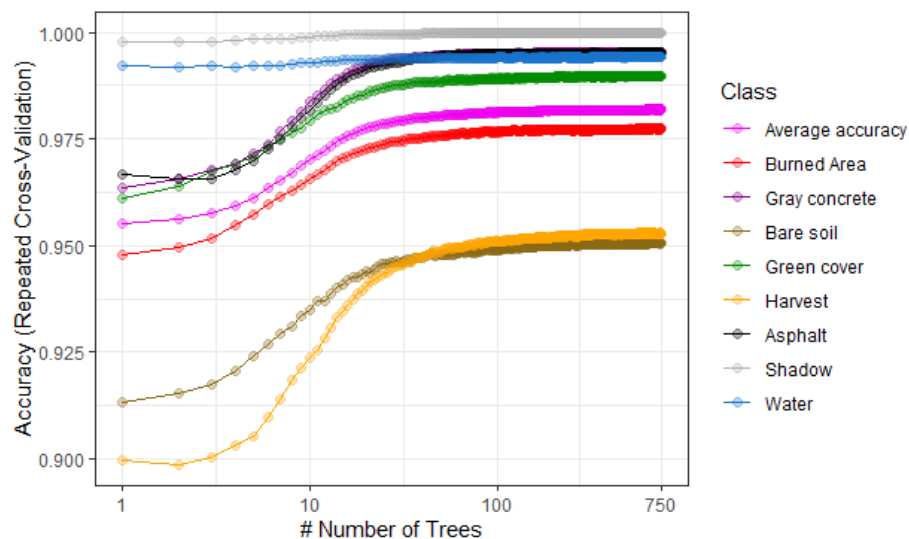
### 3.1.2. Random Forest – RF

We started by training an exploratory RF model using an approximation of the default value for the number of variables randomly sampled as candidates at each tree split ( $mtry = \sqrt{n. of predictors} = \sqrt{6 SR bands + 11 spectral index} = 4.12$ , rounded to 4). On one hand if a low number of trees are related to poor final classifiers, on the other a large number of trees are generally related to a model overwhelm and to unnecessary computation (OSHIRO; PEREZ; BARANAUSKAS, 2012). Since no

default value is available to the ntree hyperparameter, we rely on literature and set our initial candidate value as 750, an approximation of the optimal ntree value reported by Ramo & Chuvieco, 2017 to perform MODIS burned area classification.

Interestingly, a high value of accuracy (ACC= 0.976, Kappa= 0.971) was obtained from this exploratory model. Even when the model considered only a unique tree, a relatively high value of accuracy was obtained (ACC= 0.952). However, increasing in the accuracy of all the classes was observed in the window by ensembling from 5 to 100 trees (Figure 8). From this point on, an average accuracy of 0.975 was reached and a tendency of stabilization was detected for all the classes, except for the harvest class, where a small increase of accuracy (< 0.02) occurred until the 750<sup>th</sup> tree.

Since we reached a global threshold of accuracy gain (and also individual for all the classes), we assume that the input of more trees in the training is computationally unnecessary, since no more significative accuracy gains shall be observed. Thus, despite the optimal value for the number of trees can be any value since the accuracy stabilization threshold (around the 100<sup>th</sup> tree), considering the computational context of having already trained a stabilized model, we used this classifier and assumed 750 trees as our optimal ntree value in the context of this study.

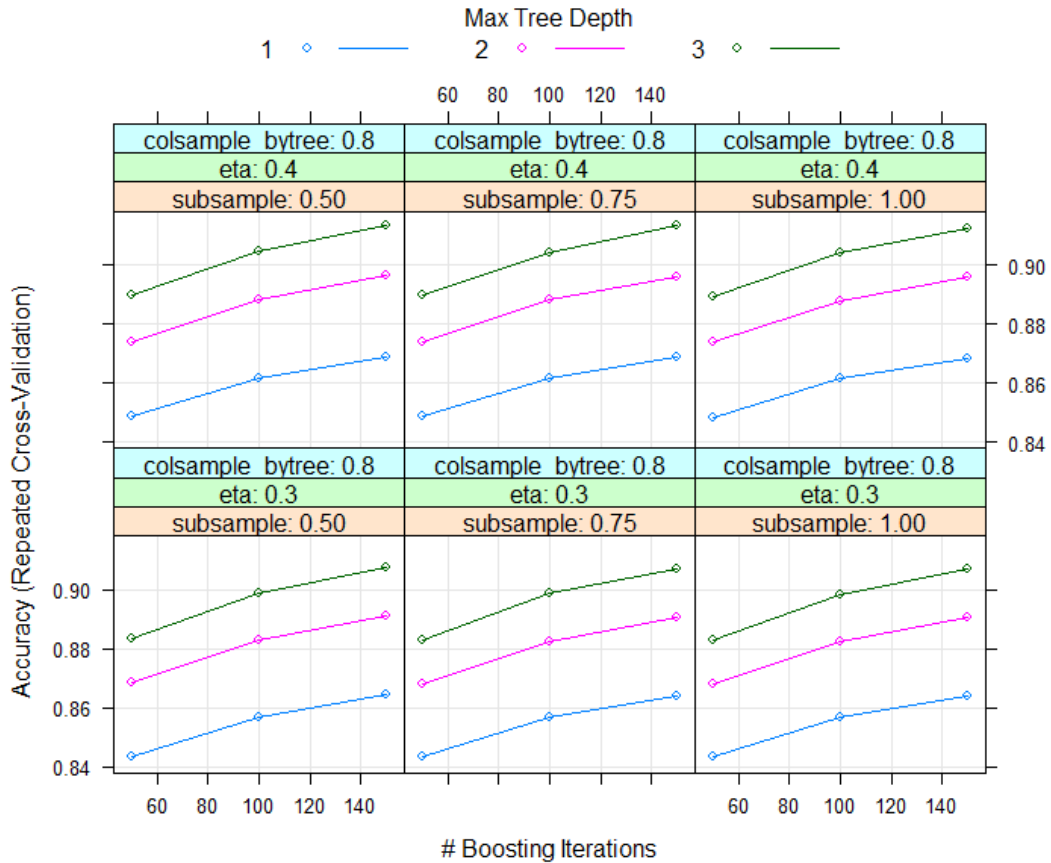


**Figure 8.** Random Forest (RF) number of trees tuning. Colored lines represent the accuracy error for each one of the classes. Pink line named “Average accuracy” represent the model overall accuracy by considering all the classes. The hyperparameter ntree was represented in the x-axis (log10 adjusted) while the y-axis points the accuracy value.

From this point on, we started to evaluate how the set of different values for the `mtry` hyperparameter affects the accuracy. For this, we tested values that corresponds to the half (2) and twice (8) of the default value (4). Considering the half value of the default, we found that the accuracy presented an insignificant drop ( $< 0.001$ ), such as none accuracy gain was observed by twice the default value (ACC was held constant in 0.976). Absence of influence of the number of variables randomly sampled as candidates at each tree split goes on the opposite direction to that reported by studies that have classified optical and radar remote sensing data by using the random forest algorithm (PAL, 2005; RAMO; CHUVIECO, 2017). Furthermore, other studies have related that the accuracy in the random forest was more sensitive to the `mtry` tuning than to the number of trees (BELGIU; DRĂGU, 2016; GHOSH; JOSHI, 2014; TOPOUZELIS; PSYLLOS, 2012). However, Catal & Diri, 2009 have reported that the behavior of the accuracy in RF can be dependent from the dataset size and feature selection methods (based on predictors relationships). Thus, our results point that the structural particularities from our dataset can be induced by the prioritization of the `n tree` in relation to `mtry`.

### 3.1.3. *eXtreme Gradient Boosting – XGB*

We report that the subsample ratio of columns (`colsample_bytree`), shrinkage (`eta`) and subsample (`subsample`) did not showed determinant effects on the accuracy values, having only small variations ( $< 0.01$ ) being observed by ranging the set of values proposed in the search grid (see Table 4). As reported by Joharestani et al., 2019, the number of allowed boosting iterations (`nrounds`) and the maximum tree depth (`max_depth`) appears to be the more sensitive to XGB hyperparameters. This way, we detected a positive relationship with the accuracy by increasing the values into each one of these parameters (Figure 9).



**Figure 9.** EXtreme Gradient Boosting (XGB) hyperparameters tuning. Colored lines represent models considering different maximum tree depth ("max\_depth"). The labels inside colored upper boxes on each plot refers to the values provided to the subsample ratio of columns (colsample\_bytree), shrinkage (eta) and subsample percentage (subsample). The x-axis represents the number of boosting iterations (nrounds) while the y-axis points the models accuracy.

Previous studies that have assessed the performance of classification algorithms by comparing XGB vs. RF models pointed that an optimal parameterized XGB tends to outperform RF models (Georganos et al., 2018; Naghibi et al., 2020; Joharestani et al., 2019). However, tuning the hyperparameters into the XGB algorithm is more difficult than MARS and RF algorithms for the simple reason that the first has 3.5 times more parameters to be set. Since the number of possible combinations between candidate values is a function of the number of candidate values in each parameter (NCV) raised to the total number of hyperparameters to be set (7) ( $NCV^7$ ), the delineation of a detailed search grid can demand the training of thousands of models. For example: a search design that considers the input of 5 candidate values for each one of the parameters needs to train 78,125 models,



making the search for the optimal hyperparameters a heuristic process that depends of computational resources availability for a long period of time.

Given that, we report that despite the optimal XGB model having reached less accuracy than the optimal RF model (- 0.063) (Table 6), this result is biased by our decision to end the XGB parameterization before the stabilization of the accuracy gain. Since a small range of eta and colsample\_bytree values were tested and no saturations in the accuracy were detected by increasing the nrounds and the max\_depth until the tested limits, a new set of values could have been supplied as new candidate values by following the conceptions of the Nelder-Mead method (OLSSON; NELSON, 1975). However, since this heuristic search process would consume much more processing time and we had already reached an accuracy that we considered satisfactory in the context of this study by using the RF algorithm, we decided to end the tuning of hyperparameters and make a better use of the research time by processing and validating a high quality final product.

**Table 6.** Optimal hyperparameters and accuracy measurements for each algorithm. Size (GB) refers to the size (expressed in gigabytes) of trained models when exported as .RData files. Gray shadow points the best model obtained from the k-fold cross validation.

Algorithm	Hyperparameters	Accuracy	Kappa	Size (GB)
RF	mtry = 4; ntree= 750	0.976	0.971	0.536
XGB	nrounds= 150; max_depth= 3; eta = 0.4; gamma =0; colsample_bytree= 0.8; min_child_weight= 1; subsample= 0.75	0.913	0.900	0.310
MARS	nprune= 20; degree= 1	0.845	0.823	19.198

#### 3.1.4. Final model selection

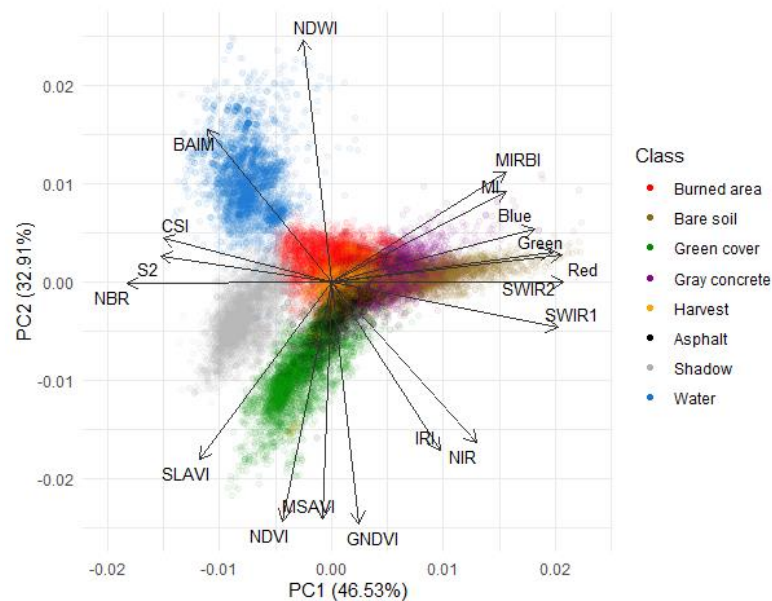
Given the previous steps, we performed the classification of our test dataset (30% of the spectral library) by using the optimal model of each algorithm. Only small variations were observed when comparing the accuracy results from the training k-fold cross validation (see Table 6) and test dataset (Table 7). Thus, we confirmed our hypothesis number ii (RF and XGB > MARS) and used the largest value of accuracy (ACC= 0.982) to select the RF as the final model to be applied in the classification.

**Table 7.** Accuracy measurements of the test dataset classification for each algorithm. 95% C.I refers to the accuracy 95% confidence interval. Grey shadow points the selected final model by considering the largest values of accuracy and Kappa.

Algorithm	Accuracy	95% C.I	Kappa
RF	0.982	0.981 – 0.983	0.979
XGB	0.913	0.912 – 0.914	0.900

### 3.2. Predictors importance

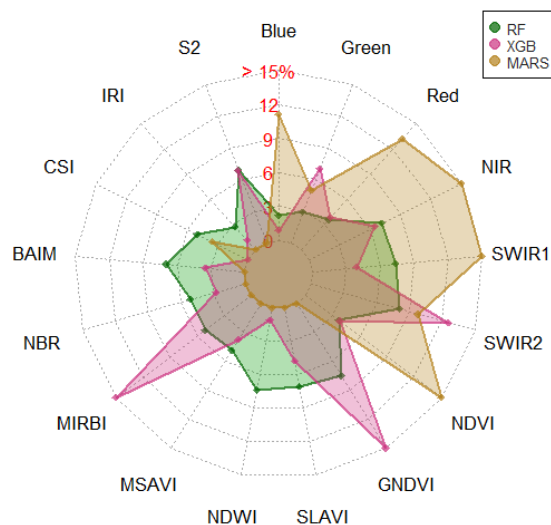
Inspecting the spectral library ordination by performing a principal component analysis (PCA – Figure 10), we detected that the loadings of surface reflectance bands have largely influenced the scores from bare soil, and in a lesser extent, from the gray concrete. As expected, spectral indexes sensible to vegetation features (NDVI, GNDVI, SLAVI, MSAVI) largely influenced the ordination of green cover, such as the burned area indexes (BAIM, CSI, NBR) were largely responsible for the burned area class scores. Highly divergent classes (green cover, shadow, water) were easily separated by the PCA, however, we detected that spectral traits were shared by some of the other classes (burned area, harvest, asphalt, gray concrete, bare soil).



**Figure 10.** Principal Component Analysis (PCA) of our spectral library. Black labels represent variable names. Length of black arrows represents the loadings of each variable.

When comparing the predictors importance inside each one of the optimal models obtained from the cross validation, we detected deep differences in which and how each model has used the predictors (Figure 11). First, the MARS model used only 9 of the 17 predictors, prioritizing the use of all the surface reflectance bands (76.6% of total importance) and using only 2 of the 11 spectral indexes (NDVI and CSI). On the other hand, XGB and RF have used all the predictors, but emphasizing the importance of spectral indexes (XGB= 63.1%, RF= 64.8% of total importance).

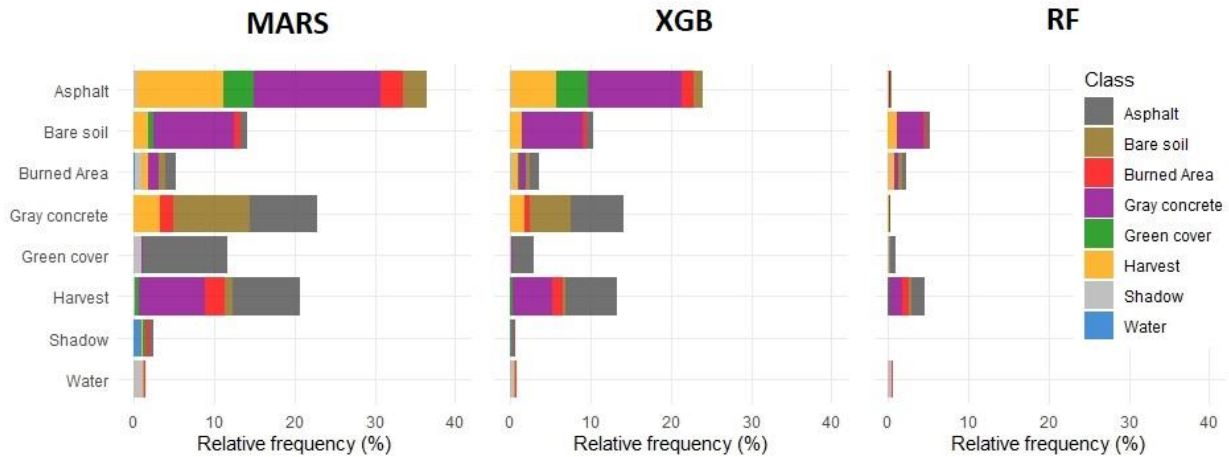
Furthermore, we detected a more homogeneous tendency of predictors importance distribution by using the RF (standard deviation  $\pm 2.29\%$ ) when compared to XGB ( $\pm 5.23\%$ ) and MARS ( $\pm 7.76\%$ ). This result highlights the ability of the RF algorithm in considering the combined effects from predictors with high collinearity by dividing the importance between them (GENUER; POGGI; TULEAU-MALOT, 2010), contrary to the non-parametric variable selection performed by the MARS (DOKSUM; TANG; TSUI, 2008) and the feature selection performed by the XGB (CHEN; GUESTRIN, 2016).



**Figure 11.** Predictors' importance radar plot. Black labels around the plot represent each one of the predictors used in the models training. Starting from the center (0%) and going to the outer edge (> 15%), the red labels inside the plot shows the relative importance for each predictor. Colored lines/polygons represent the implemented algorithms: Random Forest (dark green), eXtreme Gradient Boosting (magenta) and Multivariate Adaptive Regression Spline (brown).

Inspecting the confusion matrix of the test dataset by using the optimal models, we found a poor performance of MARS and XGB to make correct predictions of classes with high spectral mixture (asphalt, gray concrete and harvest), being these classes frequently misclassified among themselves or in burned area (Figure 12). Despite the spectral reflectance bands being often used to perform LULCC and burned area classification by other studies, better results were reported by adding spectral indexes (BASTARRIKA; CHUVIECO; MARTÍN, 2011; HAYES; MILLER; MURPHY, 2014). Many studies have reported that the NDVI and CSI enhances the classification of vegetation and burned area, respectively (Jia et al., 2014; Shao et al., 2016; Smith et al., 2007; Stroppiana et al., 2012). However, the non-consideration of the other spectral indexes and the emphasis in the surface reflectance bands

probably have induced the several observed classification errors for the MARS algorithm.



**Figure 12.** Distribution of total errors by predicting the test dataset. Black labels upper each plot represents the employed algorithm. The y-axis indicates the reference classes while the x-axis points the relative frequency of error per class. The error bar of each reference class is colored to represent the frequencies of wrong predicted classes

Although the XGB has attributed 63.1% of the variables importance to spectral indexes, these value is biased by the high importance lead by MIRBI (18.6%) and GNDVI (15.8%) while the average importance of all the other spectral indexes was around 3.2%. By comparing MARS and XGB, this behavior represents an importance reduction of the surface reflectance bands and an explicit replacement from CSI to MIRBI and from NDVI to GNDVI. Thus, XGB have reduced the classification errors from all the classes, but remained maintaining the same tendencies and biases that MARS.

On the other hand, the balanced use of surface reflectance bands and spectral indexes predictors in the RF practically zeroed the asphalt and gray concrete errors. Furthermore, misclassifications were largely reduced in bare soil (error= 6%), harvest (error= 5%) and burned area (error= 4%). Thus, our result suggests that the distribution of importance among highly collinear predictors have improved the machine learning capabilities in the RF, making possible that this algorithm reached a higher accuracy (98%).

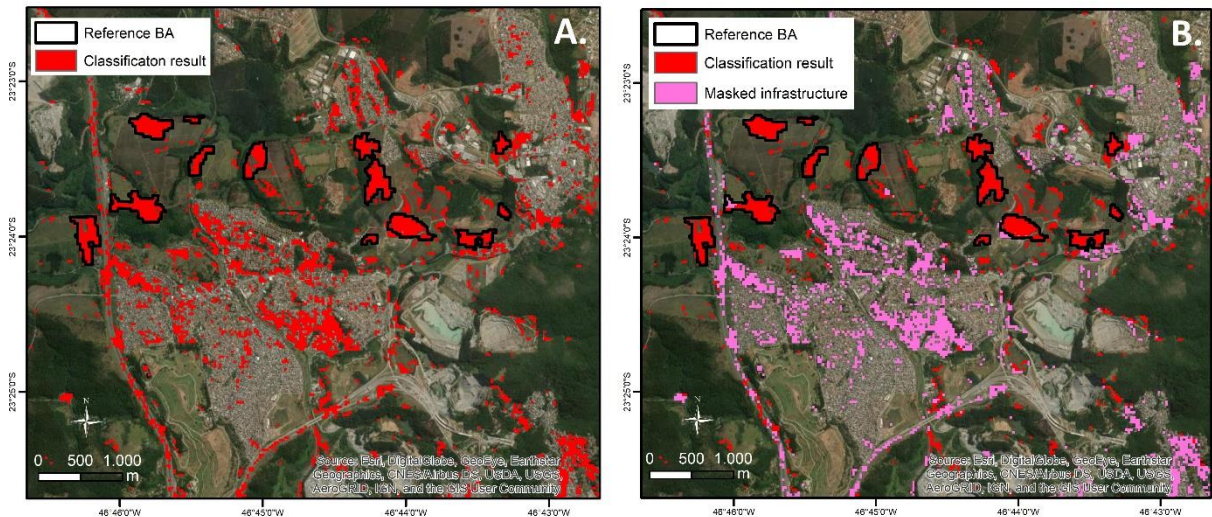
### 3.3. *Burned area validation*

We applied the final RF model (Table 6) to perform the landscape classification of the selected validation scenes (Supplementary Table S2). We

performed the burned area class binarization (1= burned, 2= unburned, all the classes from 2 to 8, see Table 1) and masked these results by applying the BQA band.

Considering the average performance of the RF model to predict burned areas into all the validation plots, first insights about the quality of this raw product showed an average kappa index of 0.53, being the average commission error (CE= 0.50) and average omission error (OE= 0.12). Despite the mean omission error being low in all the validation plots (OE - Franco da Rocha= 0.15, Itirapina= 0.12, Tanabi= 0.14, Rancharia = 0.05), high values of commission errors were observed in all the validation plots (CE – Franco da Rocha= 0.76, Itirapina= 0.50, Tanabi= 0.44, Rancharia= 0.30). We assumed that a mean omission error of 0.12 (lower= 0.05, upper= 0.15) is acceptable in the regional scope of this study and so we centered our efforts in delineating a post-processing protocol to promote the CE reduction. Inspecting our results, we found that pixels that correspond to buildings and roads were not always properly classified by our algorithm, being frequently mapped as burned area (Figure 13A). Thus, we confirmed our hypothesis iii and used this impressions as a starting point to delineate the post-processing.

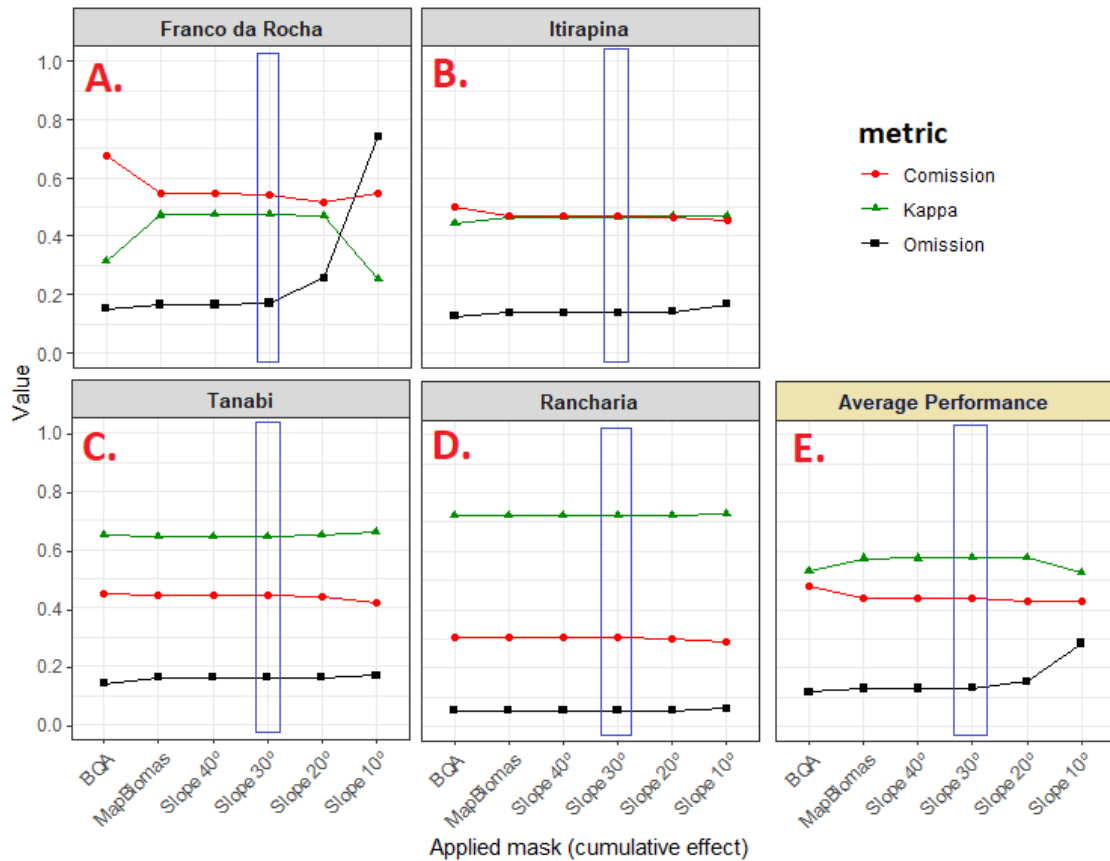
First, we decided to mask the infrastructures, cities and roads due this high commission error rate. In this way, a highly accurate urban zones classification (IRS – 5 m/pixel) is available into DATAGEO (São Paulo state geospatial data repository - <http://datageo.ambiente.sp.gov.br/>). However, these official data refers only to a static snapshot from the time (2005). Since urban infrastructure is constantly changing, we need to include urban-zones data that considers these changes. For this, the MapBiomas Brasil project offers a multi-temporal collection of land cover and land use changes (LULCC) for the entire Cerrado in Landsat resolution (ALENCAR et al., 2020). We used the urban-infrastructure MapBiomas class to mask our burned area product (Figure 13B). Besides that, considering that “rocky outcrop”, “mining” and “beach” are available into MapBiomas product (and we know that these classes don’t burn), we also masked these classes in our product since much of these is frequently related to commission errors in burned area classifications (KOUTSIAS; KARTERIS, 2000; MITRI; GITAS, 2004; OECHSLE; CLARK, 2008).



**Figure 13.** Franco da Rocha, 2018-08-30. **A)** Raw burned area product (only BQA applied). Black polygons represent our reference burned area dataset. Red colored pixels represent pixels classified as burned area by the RF model. **B)** Burned area product masked using MapBiomass (pink). We used a high resolution scene from ESRI Imagery as background in both figures.

After applying the MapBiomass mask, we observed CE decrease in all the validation plots. Franco da Rocha as shown the biggest drop in CE (-0.22), while Itirapina (-0.03), Tanabi (-0.01) and Rancharia (-0.01) presented quasi-neutral CE reduction, probably due to absence of dense urban-zones in these validation plots. Considering OE increase as a collateral effect of product masking, we detected small increases from 0.01 to 0.02 in all the validation plots. Since CE drop and OE gain were balanced for all the validation plots, the higher CE drop in Franco da Rocha has been driven an average Kappa gain from 0.53 to 0.58.

Applying different derived slopes from ALOS AW3D30 product has not reduced CE more than 0.01 in none of the validation plots. On other hand, we detected that OE is sensitive to the restriction level from slope mask, namely, the more restrictive was the slope mask, bigger was the OE increase (Figure 14), especially in Franco da Rocha. Observing that the CE and OE was balanced ( $\pm 0.01$ ) until intermediate degree slopes, we decided to maintain the slope mask ( $30^\circ$ ) as a post-processing step. We assume that despite the neutral influence in our validation plots, the slope mask can be useful to improve the product quality in other rough relief areas that were not considered in this validation scope but occur in the study area.



**Figure 14.** Cumulative performance of tested masks into each validation plot. The x-axis shows the cumulative effect of the different masks (e.g. “Slope 30°” refers to the result obtained by combining “BQA + MapBiomass + Slope 40° + Slope 30°”). Colored lines represent the CE (red), OE (black) and Kappa index (green). Dark blue rectangle points the selected combination of masks. Average performance (E) was computed by considering the mean from the results into the four validation plots.

When comparing the average results from combined masks (Landsat BQA + MapBiomass + Slope 30°) with our first product (only BQA mask), we reported a CE decrease from 0.50 to 0.43 (-0.07) and a Kappa increase from 0.53 to 0.58 (+ 0.04). We observed stable CE and OE variations across masking for all the validation plots, except on Franco da Rocha, where a high CE reduction was observed from 0.76 to 0.54 (-0.22). However, higher CE than Kappa continued to be observed in Franco da Rocha (Kappa= 0.47, CE= 0.54) and Itirapina (Kappa= 0.46, CE= 0.47).

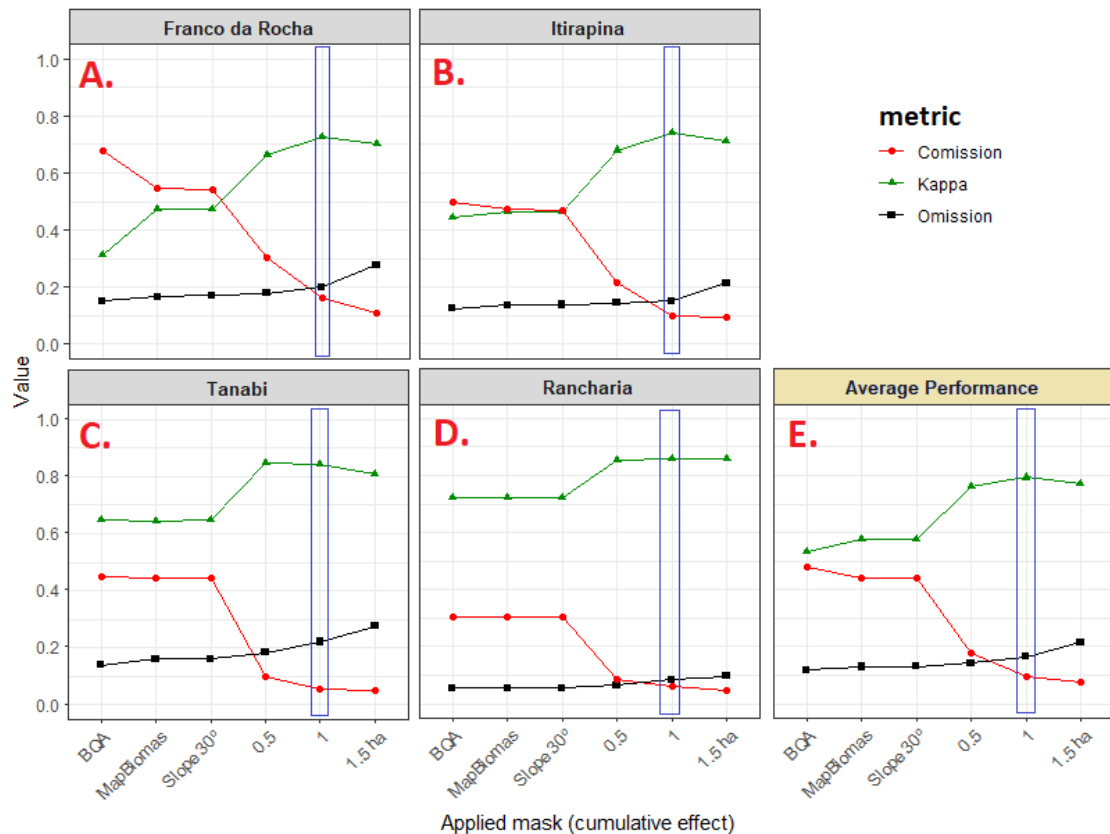
We observed that the remaining errors were distributed in a wide range of contexts. Even applying a restrictive filter to the Landsat BQA mask, we report that some small water masses, sparse clouds and their shadows were not masked. Many of these unmasked pixels corresponded to the spectral mixture between water-vegetation, water-soil, shadow-vegetation and shadow-soil, being frequently

classified as false-positive burned areas. The MapBiomass mask showed a good performance to mask false-positive burned areas in dense urban zones. However, the false-positive burned areas that corresponded to sparse buildings, rural communities and small settlements were not properly masked by MapBiomass. Besides that, São Paulo state has a dense transport infrastructure (railroads and highways) being these also not masked by MapBiomass and related to false-positive burned areas.

Although the error tendencies varied between a wide ranges of contexts, we identified a pattern that joined all these contexts: the absence of spatial contiguity. In other words, while burned areas that have been classified correctly showed high spatial contiguity (higher pixels aggregation), the false-positive burned areas presented low spatial contiguity, occurring much of the times restricted to alone pixels or aggregations less than ten pixels. Thus, we decided to use the spatial contiguity as a parameter and performed a test of filters considering different numbers of minimum pixels in an aggregation to promote the false-positive burned areas masking.

Strong CE drops were detected with the lowest spatial contiguity filter (5 pixels or ~0.5 ha) (Figure 15). Considering the average performance (Figure 15E) we report a reduction in the CE from 0.43 (all previous combined masks) to 0.17 (0.5 ha filter). Contrary to the previously tested masks that have contributed to minimize CE only in specific validation plots, the spatial contiguity filter has decreased CE in all the validation plots (Figure 15A, B, C, and D), pushing up the average performance Kappa from 0.57 to 0.76. When we apply more restrictive parameters to spatial contiguity filter by increasing the minimum number of connected pixels (11 pixels = ~1 ha, 17 pixels = ~1.5 ha), we observed CE reduction in all the validation plots. However, a tendency of CE stabilization was detected since the spatial contiguity of ~1ha (11 pixels). On the other hand, the OE presented tendency to increase in all the validation plots when more restrictive spatial contiguity parameters were provided (Figure 15).



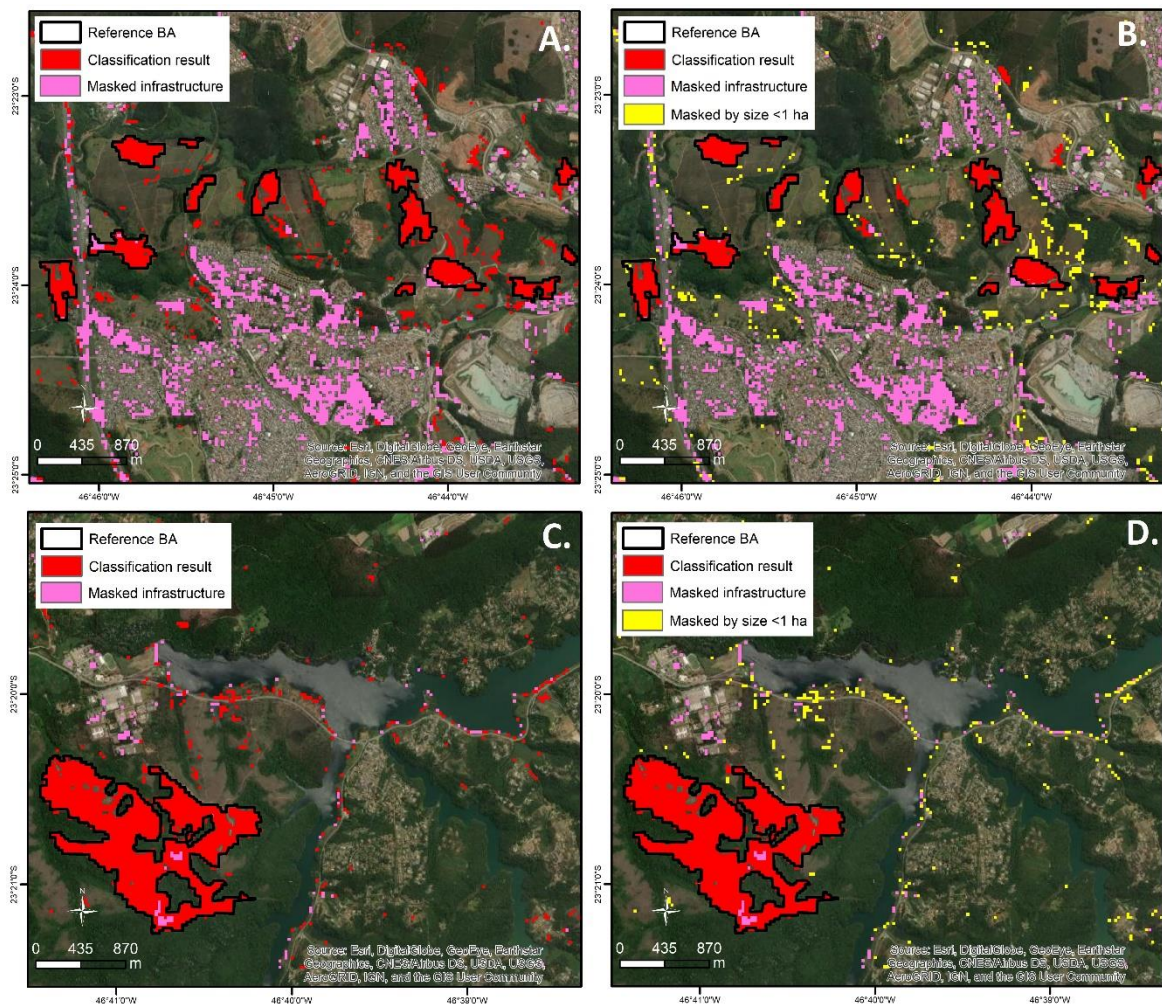


**Figure 15.** Performance of tested masks into each validation plots. The x-axis shows the cumulative effect of the different masks (e.g. “1 ha” refers to the result obtained by combining “BQA + MapBiomass + Slope 30 ° + mask all burns less than 0.5 ha + mask all burns less than 1 ha”). Colored lines represent the CE (red), OE (black) and Kappa index (green). Dark blue rectangle point the selected combination of masks. Average performance (E) was computed by summarizing the mean from the results into the four validation plots.

Inspired in the MODIS burned area products strategy of error balancing (GIGLIO et al., 2015), we assume  $\text{kappa} > \text{OE} > \text{CE}$  as our error balancing strategy. In one hand we guarantee the maximum spatial correspondence by selecting the mask parameters with the highest kappa values. On the other hand, by prioritizing a higher OE than CE we guarantee that our burned area product is ever omitting more than committing, minimizing the risk of poor inferences in regional environmental analysis. Thus, we selected the spatial contiguity filter of ~1 ha (11 pixels of spatial contiguity using rook’s adjacency criterion –  $\text{kappa} = 0.79$ ,  $\text{CE} = 0.09$ ,  $\text{OE} = 0.16$ ) as threshold and all fire scars below this value were excluded.

We detected that many of the previously mentioned error tendencies were corrected by the spatial contiguity filter. First, many of the sparse buildings wrongly classified as burned area by our algorithm were being masked (Figure 16A, B) as well

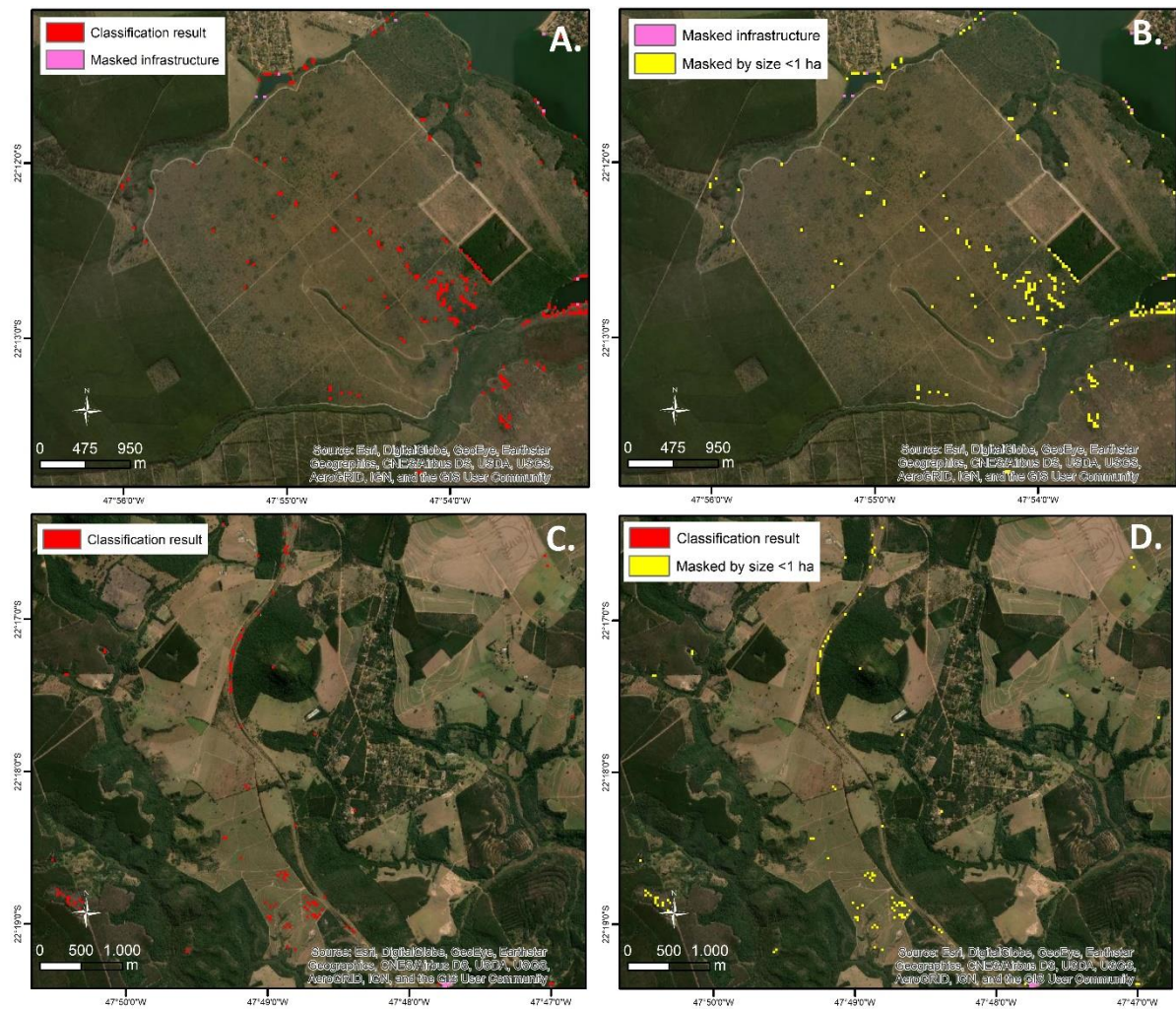
as the spectral mixtures between water-soil and water-vegetation interfaces (Figure 16C, D).



**Figure 16.** Franco da Rocha, 2018-08-30. Black polygons represent our reference burned area dataset. Red colored pixels represent pixels classified as burned area by our algorithm. Pink colored pixels represent the false-positive burned areas masked by MapBiomass. Yellow colored pixels represent the false-positive burned areas masked by minimum contiguity filter. Left boxes (A and C) show representative plots before the run of the minimum contiguity filter. Right boxes (B and D) shows the results of minimum contiguity mask by using the selected value (~ 1ha). We used a high resolution scene from ESRI Imagery as background in both figures.

Second, the native Cerrado grasslands (“campo limpo” and “campo sujo”) shows intensive phenological variations: a greenness peek into the mid wet-season and a high dehydration in the late dry season, making the spectral response of these areas in the late dry season a mixture of dry organic matter and a quartzarenic soil with high reflectance brightness. These pixels were frequently classified as false-positive burned areas (Figure 17A). However, these errors were successfully masked by the minimum spatial contiguity filter (Figure 17B). Like the sparse buildings, other

small anthropic infrastructures (like roads, highways and railways) were also wrongly classified as burned areas (Figure 17C). Since these anthropic infrastructures are small-sized when compared to Landsat scale, all the false-positive errors caused by this pattern were easily removed by applying the minimum spatial contiguity filter (Figure 17D). Finally, we assume that the accuracy ( $\kappa = 0.79$ ) and errors ( $CE = 0.09$ ,  $OE = 0.16$ ) of our burned area product is balanced to make possible future regional scale environmental analysis, validating our hypothesis iv.



**Figure 17.** Itirapina, 2015-08-29. Red colored pixels represent pixels classified as burned area by our algorithm. Pink colored pixels represent the false-positive burned areas masked by MapBiomass mask. Yellow colored pixels represent the false-positive burned areas masked by minimum contiguity filter. Left boxes (A and C) show representative plots before the run of the minimum contiguity filter. Right boxes (B and D) show the results of minimum contiguity mask by using the selected value ( $\sim 1$ ha). We used a high resolution scene from ESRI Imagery as background in both figures.

### 3.4. Final product and data access

We developed an R-Shiny web-application (Figure 18) to provide free access and navigation into our results by municipality or protected area in an interactive map ([https://bit.ly/FireGIS\\_SP](https://bit.ly/FireGIS_SP)) Interactive exploratory graphics were included and are recalculated every time that the end-user change spatial or temporal filters. We also implemented interactive buffer zone filters around protected areas to allow the users to inspect and assess possible human pressures near each protected area. Tools to enable the end-users to report the errors and implement their own improvements in the product are also planned and under development. Thus, by using our product as starting point, we pretend to launch the first collaborative Cerrado's burned area mapping platform.

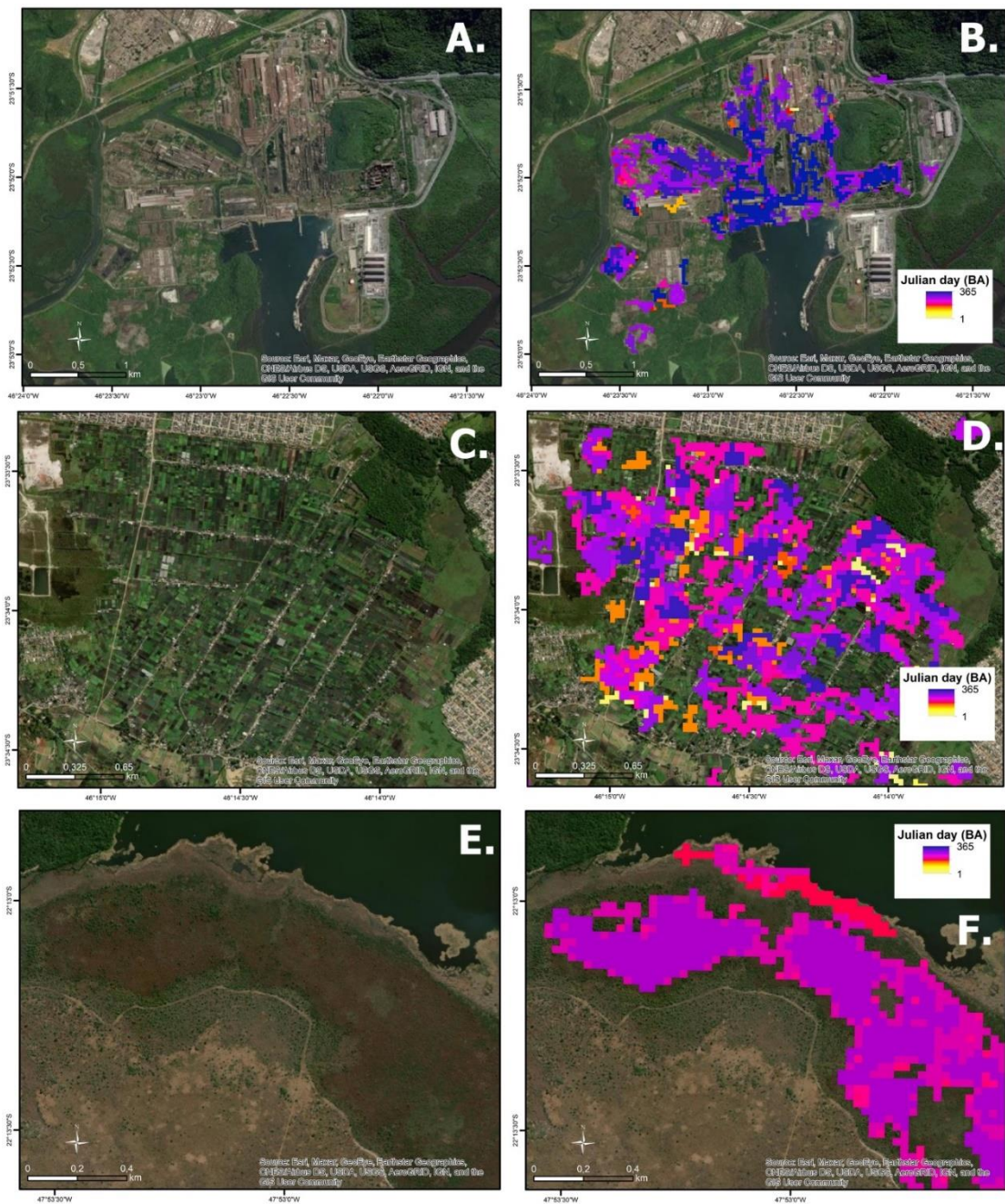


**Figure 18.** Graphical user-interface (UI) to access, visualize and analyze the final product.

### 3.5. Known issues and future development

Despite our efforts to provide a highly accurate product in regional scale, we detected some issues and biases that can affect analyses in local scale. Inspecting the final product outside our validation scope, we found that some large infrastructures (e.g. steel and petroleum industries) were not masked by the post-processing steps, being wrongly classified as burned area (Figure 19A, B). Several commission errors were also observed into managed floodplains for agriculture (e.g. rice, vegetables) (Figure 19C, D) and native floodplains locally known as “campo úmido” or “várzea” (Figure 19E, F). Thus, we report that local applications of this

product by end-users need to be inspected and, if necessary, supervised by performing the necessary improvements.



**Figure 19.** Known issues of the final product. All the figures refer to municipalities from São Paulo state and for burned area classifications from the year of 2018. **A/B.** Cubatão – “Siderúrgica Usiminas”; **C/D.** Mogi das Cruzes – Vegetable garden complex; **E/F.** São Carlos – Native floodplain “campo úmido de várzea”. We used a high resolution scene from ESRI Imagery as background in both figures.

Another important aspect to be mentioned refers to the Landsat temporal resolution and their scenes availability. Post-fire vegetation responses over tropical savannas, like Cerrado, are quick and occur only a few months after the disturbance

(BOWMAN et al., 2009; COUTINHO, 1990). We point that despite the revisit of Landsat imagery occurs every 16 days, the availability of cloud-free scenes for the study area was rare (Supplementary Fig S1). So even if we have used scenes with until 75% of cloud-cover, it is possible that some burn scars have not been imaged before the vegetation recovery due the constant cloud-cover in repeated Landsat scenes, thus being invisible to Landsat sensors when employed alone (Alves et al., 2018; Veraverbeke et al., 2011). Furthermore, our training data of burned area considers a wide range of burn scars with different contents of ash presence. That is, if the ash spectral signal disappear from the terrain surface as a result from the rain and wind, this pixel probably would be misclassified as bare soil (Pereira, 2003; Trigg & Flasse, 2001). Thus, we report that applications of our product to analyze open Cerrado patches (“campo úmido”, “campo limpo”, “campo sujo”) in local contexts need to be conducted with caution, if possible by using field data and empirical knowledge from the locals.

In this way, future versions of this algorithm will be developed by combining harmonized Landsat and Sentinel-2 imagery with the community contributions inside the Cerrado's collaborative burned area mapping platform. Thus, by summing efforts, we believe that the challenge of mapping burned areas in the highly anthropized cerrado can be overcome.

#### **4. Conclusion**

This study presented a reproducible methodology to generate a burned area algorithm by tuning and comparing different machine learning algorithms. In general, machine learning algorithms performed well to classify the LULCC and burned area in highly anthropized landscapes. RF was proved to be a better classifier than XGB and MARS, being used to classify and extract the burned area from a dense Landsat time-series. An adaptative post-processing have been implemented to balance omission (OE) and commission errors (CE) by using the strategy  $OE > CE$ , so that the final product showed to have quality to be employed in regional analysis (Kappa= 0.79).

This study generated the first burned area open dataset for the highly anthropized Cerrado. We recognize that this is only the first step, since some issues were reported in the known issues section and some improvements are necessary. However, we consider that this product represents many solid advancements in the

fire mapping for the highly anthropized Cerrado's since no burned area data are available in regional scale for this context. Besides that, this study launched the first Cerrado's collaborative burned area mapping platform, providing free and instant access to our data and glimpsing that the challenge of mapping burned areas in complex contexts needs to be shared and overcome collectively.

## 5. Acknowledgments

D.E.C. is currently receiving financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) financial code 001 and received technical scholarship support from Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grant#441968/2018-0).

## 6. Referencics

- Active Fire Data | Earthdata*. (n.d.). Retrieved July 29, 2020, from <https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data>
- Alencar, A., Z. Shimbo, J., Lenti, F., Balzani Marques, C., Zimbres, B., Rosa, M., Arruda, V., Castro, I., Fernandes Márcico Ribeiro, J. P., Varela, V., Alencar, I., Piontekowski, V., Ribeiro, V., M. C. Bustamante, M., Eyji Sano, E., & Barroso, M. (2020). Mapping Three Decades of Changes in the Brazilian Savanna Native Vegetation Using Landsat Data Processed in the Google Earth Engine Platform. *Remote Sensing*, 12(6), 924. <https://doi.org/10.3390/rs12060924>
- ALOS Global Digital Surface Model "ALOS World 3D - 30m" (AW3D30)*. (2020). <https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm>
- Área Queimada 30m*. (n.d.). Retrieved July 29, 2020, from <http://queimadas.dgi.inpe.br/queimadas/aq30m/>
- Atlântica, S. O. S. M. (2017). Atlas dos remanescentes florestais da Mata Atlantica período 2015-2016. São Paulo, Brasil. *Fundação SOS Mata Atlantica. Instituto Nacional Das Pesquisas Espaciais*.
- Bastarrika, A., Alvarado, M., Artano, K., Martinez, M., Mesanza, A., Torre, L., Ramo, R., Chuvieco, E., Bastarrika, A., Alvarado, M., Artano, K., Martinez, M. P., Mesanza, A., Torre, L., Ramo, R., & Chuvieco, E. (2014). BAMS: A Tool for Supervised Burned Area Mapping Using Landsat Data. *Remote Sensing*, 6(12), 12360–12380. <https://doi.org/10.3390/rs61212360>
- Bastarrika, A., Chuvieco, E., & Martín, M. P. (2011). Mapping burned areas from landsat TM/ETM+ data with a two-phase algorithm: Balancing omission and commission errors. *Remote Sensing of Environment*, 115(4), 1003–1012. <https://doi.org/10.1016/j.rse.2010.12.005>
- Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6(1), 20–29. <https://doi.org/10.1145/1007730.1007735>
- Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 114, pp. 24–31). Elsevier B.V. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., & Hijmans, R. (2015). rgdal: Bindings for

the Geospatial Data Abstraction Library 2017. *R Package Version 0.8-13*.

- Borini Alves, D., Montorio Llovería, R., Pérez-Cabello, F., & Vlassova, L. (2018). Fusing Landsat and MODIS data to retrieve multispectral information from fire-affected areas over tropical savannah environments in the Brazilian Amazon. *International Journal of Remote Sensing*, 39(22), 7919–7941. <https://doi.org/10.1080/01431161.2018.1479790>
- Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., D'Antonio, C. M., Defries, R. S., Doyle, J. C., Harrison, S. P., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Marston, J. B., Moritz, M. A., Prentice, I. C., Roos, C. I., Scott, A. C., ... Pyne, S. J. (2009). Fire in the Earth system. *Science (New York, N.Y.)*, 324(5926), 481–484. <https://doi.org/10.1126/science.1163886>
- Catal, C., & Diri, B. (2009). Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem. *Information Sciences*, 179(8), 1040–1058. <https://doi.org/10.1016/j.ins.2008.12.001>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August, 785–794*. <https://doi.org/10.1145/2939672.2939785>
- Chen, Y., Fan, R., Yang, X., Wang, J., & Latif, A. (2018). Extraction of Urban Water Bodies from High-Resolution Remote-Sensing Imagery Using Deep Learning. *Water*, 10(5), 585. <https://doi.org/10.3390/w10050585>
- Coutinho, L. M. (1990). *Fire in the Ecology of the Brazilian Cerrado* (pp. 82–105). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-75395-4\\_6](https://doi.org/10.1007/978-3-642-75395-4_6)
- Daldegan, G., de Carvalho, O., Guimarães, R., Gomes, R., Ribeiro, F., McManus, C., Daldegan, G. A., De Carvalho, O. A., Guimarães, R. F., Gomes, R. A. T., Ribeiro, F. D. F., & McManus, C. (2014). Spatial Patterns of Fire Recurrence Using Remote Sensing and GIS in the Brazilian Savanna: Serra do Tombador Nature Reserve, Brazil. *Remote Sensing*, 6(10), 9873–9894. <https://doi.org/10.3390/rs6109873>
- Dias, B. F. (2006). *Degradação ambiental: Os impactos do fogo sobre a diversidade do cerrado*. In I. Garay and B. Becker (Eds.), *Dimensões Humanas da Biodiversidade: O Desafio de Novas Relações Homem-Natureza no Século XXI*. Editora Vozes.
- Doksum, K., Tang, S., & Tsui, K. W. (2008). Nonparametric variable selection: The EARTH algorithm. *Journal of the American Statistical Association*, 103(484), 1609–1620. <https://doi.org/10.1198/016214508000000878>
- Douaoui, A. E. K., Nicolas, H., & Walter, C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma*, 134(1–2), 217–230. <https://doi.org/10.1016/j.geoderma.2005.10.009>
- Durigan, G., & Ratter, J. A. (2016). The need for a consistent fire policy for Cerrado conservation. *Journal of Applied Ecology*, 53(1), 11–15. <https://doi.org/10.1111/1365-2664.12559>
- Ferlito, S., Adinolfi, G., & Graditi, G. (2017). Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. *Applied Energy*, 205, 116–129. <https://doi.org/10.1016/j.apenergy.2017.07.124>
- Gao, B. C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)



- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Georganos, S., Grippa, T., Vanhuyse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very High Resolution Object-Based Land Use-Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geoscience and Remote Sensing Letters*, 15(4), 607–611. <https://doi.org/10.1109/LGRS.2018.2803259>
- Ghosh, A., & Joshi, P. K. (2014). A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26(1), 298–311. <https://doi.org/10.1016/j.jag.2013.08.011>
- Giglio, L., Justice, C. O., Boschetti, L., & Roy, D. P. (2015). MCD64A1 MODIS. *Terra+ Aqua Burned Area Monthly L3 Global 500m SIN Grid V006 MCD64A1* (<https://doi.org/10.5067/MODIS/MCD64A1.006>).
- Giles, P. T. (2001). Remote sensing and cast shadows in mountainous terrain. *Photogrammetric Engineering and Remote Sensing*, 67(7), 833–840.
- Gitelson, A. A., Merzlyak, M. N., & Lichtenthaler, H. K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology*, 148(3–4), 501–508. [https://doi.org/10.1016/S0176-1617\(96\)80285-9](https://doi.org/10.1016/S0176-1617(96)80285-9)
- Guo, X., Yin, Y., Dong, C., Yang, G., & Zhou, G. (2008). On the class imbalance problem. *Proceedings - 4th International Conference on Natural Computation, ICNC 2008*, 4, 192–201. <https://doi.org/10.1109/ICNC.2008.871>
- Hardisky, M. A., Klemas, V., & Smart, R. M. (1983). *The Influence of Soil Salinity, Growth Form, and Leaf Moisture on-the Spectral Radiance of *Paritina alterniflora* Canopies*.
- Hawbaker, T. J., Vanderhoof, M. K., Beal, Y. J., Takacs, J. D., Schmidt, G. L., Falgout, J. T., Williams, B., Fairaux, N. M., Caldwell, M. K., Picotte, J. J., Howard, S. M., Stitt, S., & Dwyer, J. L. (2017). Mapping burned areas using dense time-series of Landsat data. *Remote Sensing of Environment*, 198, 504–522. <https://doi.org/10.1016/j.rse.2017.06.027>
- Hayes, M. M., Miller, S. N., & Murphy, M. A. (2014). High-resolution landcover classification using random forest. *Remote Sensing Letters*, 5(2), 112–121. <https://doi.org/10.1080/2150704X.2014.882526>
- Hijmans, R. J., & van Etten, J. (2012). *raster: Geographic analysis and modeling with raster data. R package version 2.0–12*.
- IBGE. (2014). *Cidades do Brasil*. <https://cidades.ibge.gov.br/>
- IBGE | Censo 2010. (n.d.). Retrieved August 4, 2020, from <https://censo2010.ibge.gov.br/>
- Jeatrakul, P., Wong, K. W., & Fung, C. C. (2010). Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6444 LNCS(PART 2), 152–159. [https://doi.org/10.1007/978-3-642-17534-3\\_19](https://doi.org/10.1007/978-3-642-17534-3_19)
- Jia, K., Liang, S., Zhang, L., Wei, X., Yao, Y., & Xie, X. (2014). Forest cover classification using Landsat ETM+ data and time series MODIS NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, 33(1), 32–38. <https://doi.org/10.1016/j.jag.2014.04.015>
- Key, C. H., & Benson, N. C. (2006). Landscape assessment (LA). In: *Lutes, Duncan C.; Keane, Robert E.; Caratti, John F.; Key, Carl H.; Benson, Nathan C.; Sutherland, Steve; Gangi, Larry J. 2006*.

*FIREMON: Fire Effects Monitoring and Inventory System. Gen. Tech. Rep. RMRS-GTR-164-CD. Fort Collins, CO: US Department Of , 164.*

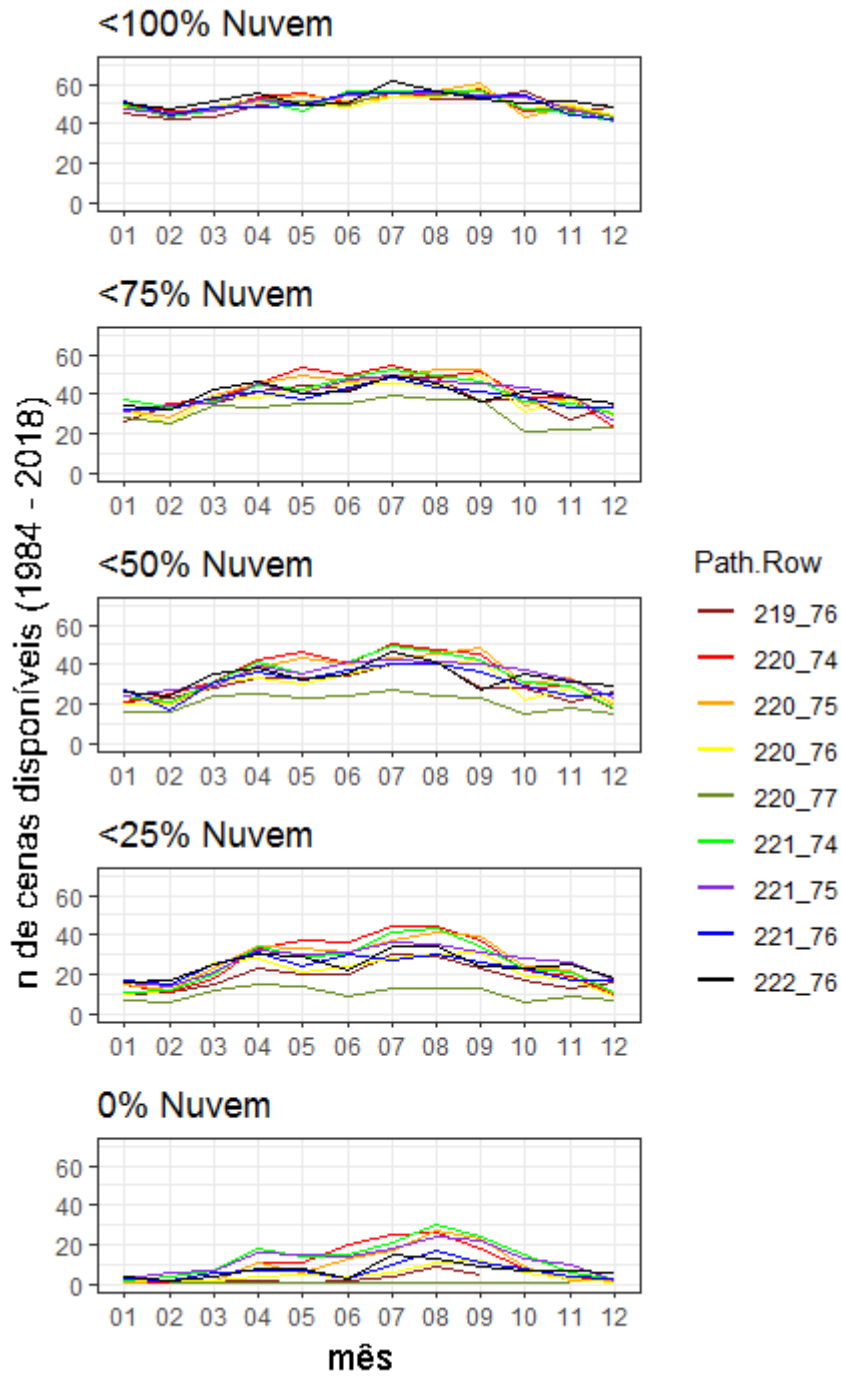
- Koutsias, N., & Karteris, M. (2000). Burned area mapping using logistic regression modeling of a single post-fire Landsat-5 Thematic Mapper image. *International Journal of Remote Sensing*, 21(4), 673–687. <https://doi.org/10.1080/014311600210506>
- Kronka, F. J. N., Nalon, M. A., Matsukuma, C. K., Kanashiro, M. M., Ywane, M. S. S., Lima, L., Guillaumon, J. R., Barradas, A. M. F., Pavão, M., & Manetti, L. A. (2005). Monitoramento da vegetação natural e do reflorestamento no Estado de São Paulo. *Simpósio Brasileiro de Sensoriamento Remoto*, 12, 16–21.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. <https://doi.org/10.1007/978-1-4614-6849-3>
- Li, D. H. W., Chen, W., Li, S., & Lou, S. (2019). Estimation of hourly global solar radiation using Multivariate Adaptive Regression Spline (MARS) – A case study of Hong Kong. *Energy*, 186, 115857. <https://doi.org/10.1016/j.energy.2019.115857>
- LP DAAC - MCD64A1. (n.d.). Retrieved July 29, 2020, from <https://lpdaac.usgs.gov/products/mcd64a1v006/>
- Lymburner, L., Beggs, P. J., & Jacobson, C. R. (2000). *Estimation of Canopy-Average Surface-Specific Leaf Area Using Landsat TM Data*.
- Martín, M. P., & Chuvieco, E. (2006). Burnt Area Index (BAIM) for burned area discrimination at regional scale using MODIS data. *Article in Forest Ecology and Management*. <https://doi.org/10.1016/j.foreco.2006.08.248>
- Mitri, G. H., & Gitas, I. Z. (2004). A semi-automated object-oriented model for burned area mapping in the Mediterranean region using Landsat-TM imagery. *International Journal of Wildland Fire*, 13(3), 367. <https://doi.org/10.1071/WF03079>
- Naghibi, S. A., Hashemi, H., Berndtsson, R., & Lee, S. (2020). Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. *Journal of Hydrology*, 589, 125197. <https://doi.org/10.1016/j.jhydrol.2020.125197>
- Oechsle, O., & Clark, A. F. (2008). Feature extraction and classification by genetic programming. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5008 LNCS, 131–140. [https://doi.org/10.1007/978-3-540-79547-6\\_13](https://doi.org/10.1007/978-3-540-79547-6_13)
- Olsson, D. M., & Nelson, L. S. (1975). The nelder-mead simplex procedure for function minimization. *Technometrics*, 17(1), 45–51. <https://doi.org/10.1080/00401706.1975.10489269>
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7376 LNAI, 154–168. [https://doi.org/10.1007/978-3-642-31537-4\\_13](https://doi.org/10.1007/978-3-642-31537-4_13)
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. <https://doi.org/10.1080/01431160412331269698>
- Paul, A. (1997). Rule-based classification of water in Landsat MSS images using the variance filter. *Photogrammetric Engineering & Remote Sensing*, 63(5), 485491.
- Pereira, A., Pereira, J., Libonati, R., Oom, D., Setzer, A., Morelli, F., Machado-Silva, F., & de Carvalho,

- L. (2017). Burned Area Mapping in the Brazilian Savanna Using a One-Class Support Vector Machine Trained by Active Fires. *Remote Sensing*, 9(11), 1161. <https://doi.org/10.3390/rs9111161>
- Pereira, J. M. C. (2003). Remote sensing of burned areas in tropical savannas. *International Journal of Wildland Fire*, 12(3–4), 259–270. <https://doi.org/10.1071/wf03028>
- Provost, F. (2000). *Machine Learning from Imbalanced Data Sets 101*. www.aaii.org
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119–126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- Ramo, R., & Chuvieco, E. (2017). Developing a Random Forest Algorithm for MODIS Global Burned Area Classification. *Remote Sensing*, 9(11), 1193. <https://doi.org/10.3390/rs9111193>
- Rosan, T. M., Aragão, L. E. O. C., Oliveras, I., Phillips, O. L., Malhi, Y., Gloor, E., & Wagner, F. H. (2019). Extensive 21st-Century Woody Encroachment in South America's Savanna. *Geophysical Research Letters*, 46(12), 6594–6603. <https://doi.org/10.1029/2019GL082327>
- Rouse, R. W. H., Haas, J. A. W., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resources Technology Satellite (ERTS) Symposium*, 309–317.
- Roy, D. P., & Boschetti, L. (2009). Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned-area products. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4), 1032–1044. <https://doi.org/10.1109/TGRS.2008.2009000>
- Shao, Y., Lunetta, R. S., Wheeler, B., Liames, J. S., & Campbell, J. B. (2016). An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. *Remote Sensing of Environment*, 174, 258–265. <https://doi.org/10.1016/j.rse.2015.12.023>
- Simon, M. F., Grether, R., Queiroz, L. P. de, Skema, C., Pennington, R. T., & Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences*, pnas.0903410106. <https://doi.org/10.1073/PNAS.0903410106>
- Smith, A. M. S., Drake, N. A., Wooster, M. J., Hudak, A. T., Holden, Z. A., & Gibbons, C. J. (2007). Production of Landsat ETM+ reference imagery of burned areas within Southern African savannahs: comparison of methods and application to MODIS. *International Journal of Remote Sensing*, 28(12), 2753–2775. <https://doi.org/10.1080/01431160600954704>
- Smith, Alistair M.S., Wooster, M. J., Drake, N. A., Dipotso, F. M., Falkowski, M. J., & Hudak, A. T. (2005). Testing the potential of multi-spectral remote sensing for retrospectively estimating fire severity in African Savannahs. *Remote Sensing of Environment*, 97(1), 92–115. <https://doi.org/10.1016/j.rse.2005.04.014>
- Soares-Filho, B., Rajão, R., Macedo, M., Carneiro, A., Costa, W., Coe, M., Rodrigues, H., & Alencar, A. (2014). Cracking Brazil's forest code. *Science*, 344(6182), 363–364.
- Stroppiana, D., Bordogna, G., Carrara, P., Boschetti, M., Boschetti, L., & Brivio, P. A. (2012). A method for extracting burned areas from Landsat TM/ETM+ images by soft aggregation of multiple Spectral Indices and a region growing algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 69, 88–102. <https://doi.org/10.1016/j.isprsjprs.2012.03.001>
- Team, R. C. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Topouzelis, K., & Psyllos, A. (2012). Oil spill feature selection and classification using decision tree

- forest on SAR image data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68(1), 135–143. <https://doi.org/10.1016/j.isprsjprs.2012.01.005>
- Trigg, S., & Flasse, S. (2001). An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. *International Journal of Remote Sensing*, 22(13), 2641–2647. <https://doi.org/10.1080/01431160110053185>
- Van Hulse, J., Khoshgoftaar, T. M., & Napolitano, A. (2007). Experimental perspectives on learning from imbalanced data. *ACM International Conference Proceeding Series*, 227, 935–942. <https://doi.org/10.1145/1273496.1273614>
- Veraverbeke, S., Lhermitte, S., Verstraeten, W., & Goossens, R. (2011). A time-integrated MODIS burn severity assessment using the multi-temporal differenced normalized burn ratio (dNBRMT). *International Journal of Applied Earth Observation and Geoinformation*, 13(1), 52–58. <https://doi.org/10.1016/j.jag.2010.06.006>
- Vicente, L. E., Souza Filho, C. R., & Perez Filho, A. (2005). Mapeamento de formações arenosas em fragmentos de Cerrado utilizando dados e produtos do sensor ASTER. *XII Simpósio Brasileiro de Sensoriamento Remoto. INPE, Goiânia*, 3419–3426.
- Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere*, 10(7), 373. <https://doi.org/10.3390/atmos10070373>

## Supplementary

Suplemmentary Fig S1. Landsat scenes availability using different cloud cover thresholds.



**Supplementary Table S2.** List of scenes used in the product validation

Validation plot	Scene list
Franco da Rocha	LT05_L1TP_219076_19951002_20170106_01_T1 LT05_L1TP_219076_20030922_20161204_01_T1 LT05_L1TP_219076_20031024_20161203_01_T1 LC08_L1TP_219076_20170624_20170713_01_T1 LC08_L1TP_219076_20170726_20170810_01_T1 LC08_L1TP_219076_20170827_20170914_01_T1 LC08_L1TP_219076_20170912_20170928_01_T1 LC08_L1TP_219076_20180424_20180502_01_T1 LC08_L1TP_219076_20180510_20180517_01_T1 LC08_L1TP_219076_20180830_20180911_01_T1
Itirapina	LT05_L1TP_220075_19850319_20170219_01_T1 LT05_L1TP_220075_19850420_20170219_01_T1 LT05_L1TP_220075_19850506_20170219_01_T1 LT05_L1TP_220075_19850709_20170219_01_T1 LT05_L1TP_220075_19850810_20170219_01_T1 LT05_L1TP_220075_19850911_20170218_01_T1 LT05_L1TP_220075_19880224_20170209_01_T1 LT05_L1TP_220075_19880327_20170209_01_T1 LT05_L1TP_220075_19880701_20170208_01_T1 LT05_L1TP_220075_19880717_20170208_01_T1 LT05_L1TP_220075_19880802_20170207_01_T1 LT05_L1TP_220075_19880919_20170206_01_T1 LT05_L1TP_220075_19881106_20170205_01_T1 LT05_L1TP_220075_19881208_20170205_01_T1 LC08_L1TP_220075_20150117_20180528_01_T1 LC08_L1TP_220075_20150202_20170413_01_T1 LC08_L1TP_220075_20150423_20170409_01_T1 LC08_L1TP_220075_20150509_20170409_01_T1 LC08_L1TP_220075_20150525_20170408_01_T1 LC08_L1TP_220075_20150610_20170408_01_T1 LC08_L1TP_220075_20150728_20170406_01_T1 LC08_L1TP_220075_20150813_20170406_01_T1 LC08_L1TP_220075_20150829_20170405_01_T1 LC08_L1TP_220075_20151016_20170403_01_T1 LC08_L1TP_220075_20180314_20180320_01_T1 LC08_L1TP_220075_20180517_20180604_01_T1 LC08_L1TP_220075_20180618_20180703_01_T1
Tanabi	LT05_L1TP_221074_20060726_20161120_01_T1

	LT05_L1TP_221074_20060811_20161119_01_T1 LT05_L1TP_221074_20060912_20161118_01_T1 LC08_L1TP_221074_20160416_20170326_01_T1 LC08_L1TP_221074_20160502_20170325_01_T1 LC08_L1TP_221074_20160721_20170323_01_T1 LC08_L1TP_221074_20160923_20170321_01_T1 LC08_L1TP_221074_20161009_20170320_01_T1 LC08_L1TP_221074_20180422_20180502_01_T1 LC08_L1TP_221074_20180508_20180517_01_T1 LC08_L1TP_221074_20180625_20180704_01_T1 LC08_L1TP_221074_20180727_20180731_01_T1 LC08_L1TP_221074_20180913_20180928_01_T1
Rancharia	LT05_L1TP_222076_19850128_20170219_01_T1 LT05_L1TP_222076_19850605_20170219_01_T2 LT05_L1TP_222076_19850824_20170218_01_T1 LT05_L1TP_222076_19850909_20170218_01_T1 LT05_L1TP_222076_19851112_20170218_01_T1 LT05_L1TP_222076_20010905_20161211_01_T1 LT05_L1TP_222076_20011007_20161210_01_T1 LT05_L1TP_222076_20011226_20161210_01_T1 LC08_L1TP_222076_20170512_20170525_01_T1 LC08_L1TP_222076_20170715_20170727_01_T1 LC08_L1TP_222076_20170901_20170915_01_T1 LC08_L1TP_222076_20180515_20180604_01_T1 LC08_L1TP_222076_20180531_20180614_01_T1 LC08_L1TP_222076_20180718_20180731_01_T1

## **The conservation paradox on highly anthropized Cerrado: Protect to burn or burn to protect?**

Dhemerson E. CONCIANI<sup>1\*</sup>; Swanni T. ALVARADO<sup>2,3</sup>; Thiago S.F SILVA<sup>4</sup>

<sup>1</sup> Universidade Estadual Paulista (UNESP), Instituto de Biociências, Departamento de Biodiversidade, Avenida 24-A 1515, 13506-900, Rio Claro, Brazil.  
[\\*dhemerson.conciani@unesp.br](mailto:dhemerson.conciani@unesp.br)

<sup>2</sup> Universidade Estadual do Maranhão, Programa de Pós-Graduação em Agricultura e Ambiente, Praça Gonçalves Dias, s/n, 65800-000 Balsas, Brazil

<sup>3</sup> Universidade Estadual do Maranhão (UEMA), Programa de Pós-graduação em Geografia, Natureza e Dinâmica do Espaço, São Luís, Maranhão, Brazil.

<sup>4</sup> Biological and Environmental Sciences, Faculty of Natural Sciences, Stirling University, Stirling, FK9 4LA, UK



## ABSTRACT

Cerrado is the most diverse savanna in the world and the second largest biome in Brazil's, behind only of the Amazon. Because its close evolutionary relationship with the fire, Cerrado vegetation has different adaptations for this disturbance, and is considered as a fire prone ecosystem. In this study, we analyzed regionally calibrated remote sensing datasets of burned area (BA) and land cover and land use changes (LCLUC) to propose a novel approach to track fire regime patterns. Our study focused on the highly anthropized Cerrado in Sao Paulo State. To measure impacts on conservation, we assessed fire regimes within all the protected areas (PAs) covered by Cerrado vegetation in São Paulo's state and compared to the fire regimes observed in a 10 km buffer zone. We found that most of burned areas in the last three decades occurs on sugar-cane croplands and pastures. BA patterns have followed tendencies of LCLUC, with a decrease in total burned area in the last decades due to the replacement of pastures by croplands. PAs with forest formations have rarely burned between 1985 and 2018, while all the PAs dominated by grasslands or with African grasses invasion had higher fire frequencies and larger burned areas. Contradictorily, all the PAs with the highest fire occurrence were categorized into the most restrictive protection level. We conclude that the fire exclusion policy is not suitable to manage PAs on grasslands and savannas formation, since these areas are burning periodically due to the high fuel build-up and accidental or arson fires. This research highlight the need of new management strategies that allow prescribed fires and include local communities as actors in the conservation process.

**Key- words:** fire regime; burned area; protected area; land cover change

## Introduction

Cerrado vegetation cover an area of c.a 2 million of km<sup>2</sup>, about a quarter of the Brazilian territory (Durigan & Ratter, 2016). Composed by a wide-range of vegetation types, Cerrado physiognomies includes open grasslands (locally known as “campo limpo”), savannas (“campo sujo”, “campo cerrado” and “cerrado sensu stricto”) and forest-like formations (“cerradão”) (Coutinho, 1990; Ribeiro & Walter, 2017), and can be considered the most diverse savanna in the world (Murphy et al., 2016).

Cerrado have historically evolved under a natural fire regime since the last 4 Myr, being this considered one of the most important drivers of its ecosystem dynamics (Coutinho, 1990; Simon et al., 2009). Natural fire occurs in the season transitions (later dry season and early rainy season) occasioned mainly by lightnings (Dias, 2006). Natural fire dynamic can be explained by climatic conditions, where the amount of accumulated rainfall over 18-24 months is responsible to drive fuel build-up, while fire season rainfall is responsible for the regulation of the ignition and fire spread probabilities (Alvarado et al., 2020; Bradstock, 2010). In highly anthropogenic landscapes like São Paulo’s Cerrado, climate influence can be overcome by human influence, and fire regimes can be often driven by socio-cultural factors, land use and land cover (Archibald, 2016; Conciani et al., 2021). In addition, archeological evidences highlight that fire has been used as a tool by indigenous peoples in the Cerrado for c.a 3 kyr ago, mainly to manage small areas for agriculture and hunting (Dias, 2006; Guidon, 1992; Mistry et al., 2005).

Considering the high demographic Brazilian growth during the last decades, until 2017 at least 45% of the Cerrado area have been suppressed and replaced by agriculture (i.e. soybean, sugar-cane) and livestock farming (Alencar et al., 2020), modifying the natural fire regime and threatening the conservation of this ecosystem (Durigan et al., 2007). Thus, although fire is a natural disturbance of the Cerrado, most of fires setting up in São Paulo state are arson/accidental and occurs near towns or lands with economic use. For this reason, fire has been historically considered by the media, population and lawmakers as a threat, being completely banned since 2001 by the state law 10547, independently of the origin and the affected land cover.

About the conservation status, only 7% of the remnant Cerrado are under legal protection in Brazil (Soares-Filho et al., 2014). Considering only São Paulo state, the

Cerrado remaining vegetation is restricted to small fragments that covers less than 6% of pre-colonization area, of which only 12% are legally protected (~250 km<sup>2</sup>) (Fiori & Fioravanti, 2001; Kronka et al., 2005). All these remaining patches are inserted into a complex anthropic matrix (Alencar et al., 2020). Besides, the presence of highly populated cities (c.a 44 million inhabitants) and a robust infrastructure network (e.g highways, railways, factories and power plants) (IBGE, 2014) complete the São Paulo state's landscape mosaic, modeling the socio-environmental dynamic and creating a complex scenario.

The available standardized Landsat imagery and remote sensing tools allow us now to monitor more than three decades of land use changes and reconstruct the ecological dynamic shaping landscapes, such as fire occurrence (Alencar et al., 2020; Alvarado et al., 2017; Lentile et al., 2006). A previous study developed in three protected areas of São Paulo's Cerrado showed that fire exclusion policy was an inefficient strategy to manage these areas, increasing its vulnerability and making them dependent of their surrounding areas fire regimes, trapping them into a fire paradox (Arévalo & Naranjo-Cigala, 2018; Conciani et al., 2021). These changes in natural fire regimes directly affect the effectiveness of conservation, especially when considering that the last Cerrado remnants are small patches that needs to be managed to ensure its adequate protection.

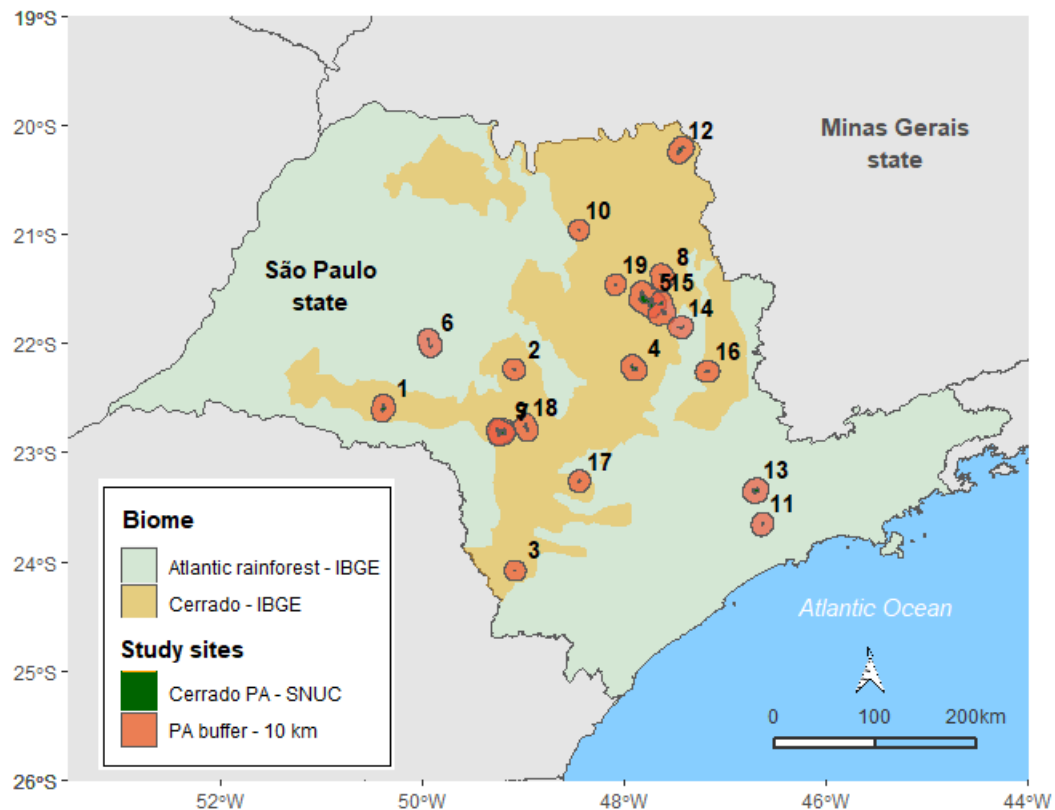
In this way, this study combines more than three decades (from 1985 to 2018) of moderate spatial resolution (pixel= 30 meters) burned area data (Conciani et al. in prep) and land cover changes (MapBiomas collection 5) to assess the contemporary fire regime within and around 19 protected areas in São Paulo state located on Cerrado ecosystem. This innovative analysis considers the context of the highly anthropized Cerrado providing a “big picture” to understand the past, the current and delineate future perspectives to ensure the conservation of this threatened ecosystem. We aim to answer the following questions: i) how is the fire pattern on the different land cover across the last three decades? ii) how is the fire frequency within the protected areas when compared with its surrounding areas? iii) Is there any difference in the spatio-temporal variation of burned area between the protected areas and its surrounding areas?

## Methods

### *Study area*

We considered the Cerrado delineation provided by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística – IBGE). However, once the Cerrado vegetation occurs bordering with Atlantic rainforest biome in an ecotonal zone, we revised every management plan (or equivalent) of each protected areas (PA) registered in São Paulo state to verify the vegetation cover on each PA. Thus, we apply a first filter of the potential study areas by eliminating the PAs presents on the IBGE Cerrado's delineation but covered by Atlantic rainforest vegetation. We also included the PAs with Cerrado vegetation located in Sao Paulo state but outside the delineation of the IBGE Cerrado's biome.

In total, this study included nineteen protected areas in the highly anthropized Cerrado in São Paulo state, which correspond to all the Cerrado remaining patches protected by the Brazilian National Conservation System (*Sistema Nacional de Unidades de Conservação* – SNUC) under full preservation status (Ecological Station, State Park or Biological Reserve) and sustainable use (State Forest, Private Reserve). To assess anthropogenic effects on PA fire regimes, we also considered a buffer zone of 10 kilometers around each PA (Figure 1). Thus, our study covered in total 28 573 hectares of PA and 957 251 hectares of buffer zones (Table 1), where 25 764 hectares of the PA were classified under full preservation category (90%) and 2 809 hectares under sustainable use (10%).



**Figure 1.** Study site considering all the protected areas (PA) of the Brazilian National Conservation System (*Sistema Nacional de Unidades de Conservação – SNUC*). Dark green polygons show the PA. Red polygons represents the estimated buffer of 10 kilometers around each PA. Numbers upper each PA/Buffer indicate the ID of each PA (see table 1). Background colors in São Paulo state represents the distribution of biomes (light green= Atlantic rainforest; orange= Cerrado) estimated by the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística – IBGE*) and used as reference in this study.

**Table 1.** Characteristics of the selected protected areas (PA). ID is the identification number of each PA. SNUC Category is the conservation category following the categories of the Brazilian National Conservation System (ES= Ecological Station; SF= State Forest, SP= State Park, BR= Biological Reserve and PR= Private Reserve). Conservation status: FP= Full Preservation and SU= Sustainable Use. Creation year (yr.) indicates the year that each PA was officially instituted.

ID	SNUC Category <sup>1</sup>	Conservation status <sup>2</sup>	Protected Area	Creation year	Area (ha)	Predominantly vegetation
1	ES	FP	Assis	1984	1768	Cerradão
2	ES	FP	Bauru	1987	310	Atlantic rainforest and savanna
3	ES	FP	Itapeva	1985	100	Atlantic rainforest and cerradão

4	ES	FP	Itirapina	1984	2209	Open grasslands and savanna
5	ES	FP	Jataí	1982	9011	Savanna and cerradão
6	ES	FP	Marília	2010	579	Atlantic rainforest and cerradão
7	ES	FP	Santa Bárbara	1984	2807	Open grasslands and savanna
8	ES	FP	Santa Maria	1985	1313	Open grasslands and savanna
9	SF	SU	Águas de Santa Bárbara	1964	1600	Savanna and forestry
10	SF	SU	Bebedouro	1927	103	Cerradão and forestry
11	SP	FP	Fontes do Ipiranga	1969	552	Atlantic rainforest and cerradão
12	SP	FP	Furnas do Bom Jesus	1989	2065	Atlantic rainforest and savanna
13	SP	FP	Juquery	1993	1998	Open grasslands and savanna
14	SP	FP	Porto Ferreira	1987	611	Atlantic rainforest and cerradão
15	SP	FP	Vassununga	1970	2079	Atlantic rainforest and cerradão
16	BR	FP	Mogi-Guaçu	1942	383	Atlantic rainforest and cerradão
17	PR	SU	Entre Rios	2010	304	Atlantic rainforest and savanna
18	PR	SU	Olavo Egydio Setúbal	2008	617	Atlantic rainforest and cerradão
19	PR	SU	Toca do Paca	2008	185	Atlantic rainforest and savanna

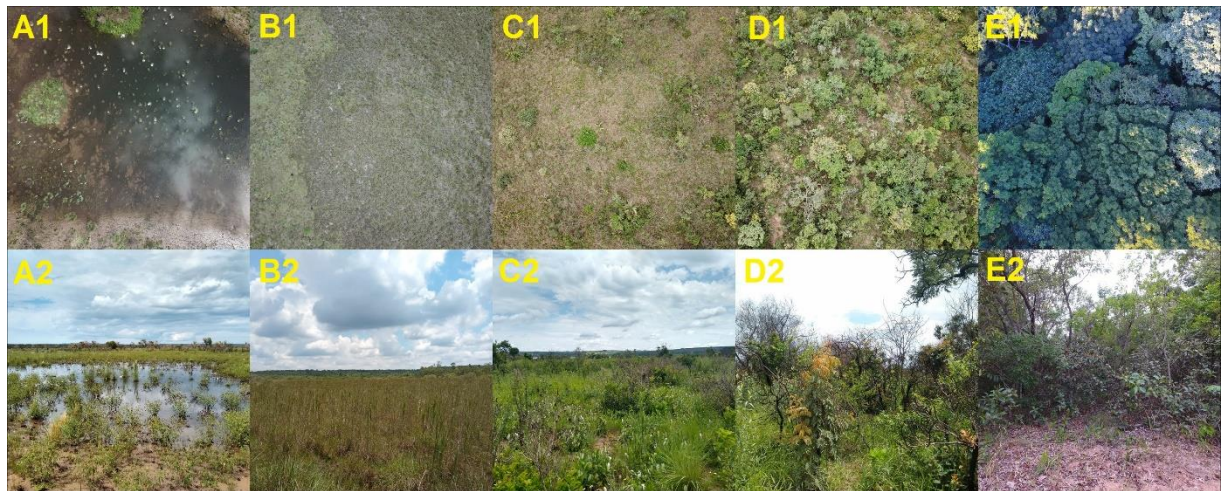
### *Management context*

The 19 protected areas were classified into two groups according with its SNUC conservation status: 14 were classed under full preservation (FP – 74%) and 5 under sustainable use (SU – 26%). The first group included the Ecological Station (ES), State Park (SP) and Biological Reserve (BR) SNUC categories, while the second group included State Forest (SF) and Private Reserve (PR) categories. It's important to notice that most of full preservation PA were created between 1970 – 1980's, when national political initiatives prioritized the protection of the last (and relatively small) fragments of the remaining Cerrado vegetation. Considering the high land cover and land use changes (LCLUC) around these PA and following the “untouchable nature” myth (Diegues, 2001), most of these PAs were categorized into high restrictive conservation categories (such as Ecological Station and Biological Reserve), in which only scientific research and guided environmental education are allowed.

Sustainable use areas are divided into two categories: 1) State Forests are old areas created by the government to foster forestry research in the region. Across time, woody exotic species (like *Eucaliptus* spp. and *Pinus* spp.) have been gradually removed and Cerrado vegetation has recovered on these areas. 2) Private Reserve are recently created PA, managed by private landowners (in contrast to the previous ones managed by the government). All the PAs share the same fire exclusion policy since the creation, independently of its conservation status (Full preservation or sustainable use) or its management (public or private).

### *Landscape structure*

The remaining native vegetation inside the PA includes a gradient of Cerrado physiognomies, ranging from open grasslands (“campo limpo”) to forest-like formations (“cerradão”) (Figure 2). Besides, around 60% of the PA selected in this study are located in the border of the ecotonal zone between Cerrado and seasonal tropical forest (locally know as Atlantic rainforest) and around 10% have a mix of native and woody exotic species.



**Figure 2.** Vegetation types in the study sites: A) “campo úmido” – wet grassland; B) “campo limpo” – open grassland; C) “campo sujo” – shrubland; D) “cerrado *sensu stricto*” - Savanna; E) “cerradão” – forest-like formation. Top images correspond to aerial image from DJI Phantom 4 (50 meters altitude, 90-degree camera angle), bottom images correspond to field photography of the same places where the drone images were taken.

In our study, five PA under the full preservation category represent the largest Cerrado patches in São Paulo state (Jataí Ecological Station (JES), Santa Bárbara Ecological Station (SBES), Itirapina Ecological Station (IES), Vassununga State Park (VSP) and Juquery State Park (JSP)). Together, these areas represent more than 64% of the total area of evaluated PA, where 36% of the remaining area is distributed among the other 14 PA.

Despite the large variation on PA size, these largest fragments can be considered small when compared to the anthropogenic matrix around. Currently the buffer zones around PA is characterized by urban zones, infrastructure (roads, highways, railways, power lines), croplands (sugar-cane, soybean, corn), pasture, reforestation (e.g. Pinus and Eucaliptus forestry), industrial zones and small unprotected native Cerrado remnants (Alencar et al., 2020). Fire was deliberately used as a tool in sugar-cane, pasture and deforestation in the São Paulo state up to the year 2000. However, legal constraints have been imposed by lawmakers, thus fire was banned and the political strategies were delineated to gradually reduce ignitions and total burned area (state law 11 241/ 2002).

#### *Data collection*

##### *Protected areas and buffer zones*

We obtained the georeferenced delineation of São Paulo's protected areas on vector format directly from the institutions responsible for its managing (e.g.



“*Instituto Florestal*” and “*Fundação Florestal*”), all under administration of the state government of São Paulo. Official buffer zone data were not available for all the PAs. Nevertheless, the National Environment Council (*Conselho Nacional do Meio Ambiente* – CONAMA) establishes an automatic radius of 3 km as legal buffer zone (Resolução CONAMA 428/ 2010). But to consider the landscape context, we used the recommendations of the national law 99 274/ 2002 that standardize a linear ring of 10 kilometers around each PA as buffer zone.

#### *Burned area*

We used the yearly burned area product from Conciani et al. (In prep., [https://bit.ly/FireGIS\\_SP](https://bit.ly/FireGIS_SP)). This product was generated by using machine learning and offers a raster dataset of burned areas with 30 meters of spatial resolution and 16 days of temporal resolution from 1985 to 2018, based on Landsat 5, 7 and 8 images. Due to the highly complexity to map burn scars in highly anthropized Cerrado areas, this product was optimized and validated by considering regional aspects (e.g. anthropogenic LCLUC and fire dynamic) ensuring a high accuracy (79%) and quality to perform ecological and environmental analyses.

#### *Land cover and land use change*

Yearly land cover and land use change (LCLUC) was obtained from MapBiomias project – collection 5 (Souza et al., 2020, <https://mapbiomas.org/>). The MapBiomias Project is a multi-institutional initiative to generate annual maps of land cover and use from automatic classification processes applied to Landsat 5, 7 and 8 imagery. The land cover classes considered of this study were: forest formation, savanna formation, grassland formation, wetland, forestry, pasture, temporary crop, soybean, sugar cane and perennial crop.

#### *Data processing and analysis*

##### *Pre-processing burned area data*

Burned area data was provided by year (34 years) and by Landsat WRS-2 path/row (9 path/row) scenes, totalizing 309 raster files. First, we mosaic to a new raster all the individual scenes for each year. We obtained 34 raster files (1 from each year) where pixel values were range from 0 to 365 (0= unburned, 1 to 365= Julian

day of the burn scar detection). An additional binary dataset was generated by reclassifying each yearly burned area data in 0= unburned and 1= burned (1-365 values).

### *Burned area by land cover and land use*

We matched the yearly binary burned area with yearly MapBiomas collection 5 data. For each year, we overwritten pixel value equal to 1 (burned) in the burned area dataset with the LCLUC pixel value (Supplementary Table S1). Thus, we generated a dataset containing the land cover and land use that was burned in each year from 1985 to 2018.

### *Fire regime metrics*

We described the fire regime by using the metrics suggested by Alvarado et al., (2017) and Conciani et al. (2021) to assess the fire regime of the highly anthropized Cerrado in Sao Paulo (Table 2). Each metrics was calculated by considering the IBGE Cerrado's total area, PA and Buffer zones during the period from 1985 to 2018, such as a general sum of these metrics across the time-series.

**Table 2.** Derived metrics used to assess fire regime patterns on São Paulo's Cerrado area, its protected areas and buffer zones.

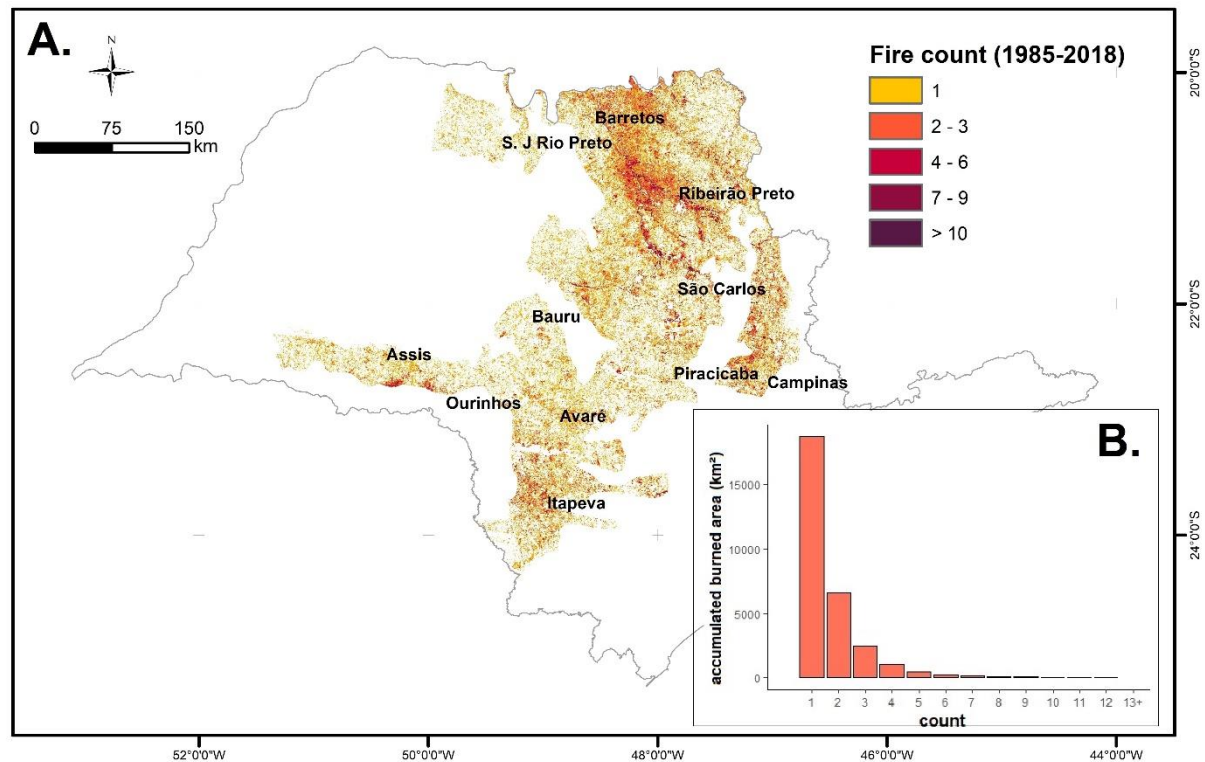
<b>Metric</b>	<b>Unit</b>	<b>Description</b>
Fire count	<i>count</i>	Raster file with the total number of fire events observed over the 34 years of study, obtained from the per-pixel sum of all overlapped binary raster files.
Burned area	<i>km<sup>2</sup></i>	Sum of burned pixels (30 x 30 m of resolution) multiplied by the conversion factor 0.0009 (m to km <sup>2</sup> )
Accumulated burned area	<i>km<sup>2</sup></i>	Sum of annual burned area maps
Relative burned area	%	Proportion between burned area in a given year and the total area (e.g. IBGE Cerrado area, PA area or Buffer zone).
Relative accumulated burned area	%	Sum of relative burned areas over time (previous years + current year)

## *Results and Discussion*

### *General burned area pattern from the highly anthropized Cerrado*

On average, 888 km<sup>2</sup> ( $\pm$  770 km<sup>2</sup>) have burned per year in São Paulo's Cerrado during the last three decades, totalizing 30 198 km<sup>2</sup> of burned areas from 1985 to 2018 (~37.2% of the São Paulo's Cerrado total area, Figure 3A). Of these, 18 644 km<sup>2</sup> was burned only once (equivalent to 61.7% of the total burned area), 9 074 km<sup>2</sup> twice or three times (30%), 1 815 km<sup>2</sup> between four and six times (6.2%), 353 km<sup>2</sup> between seven and nine times (1.1%) and only 309 km<sup>2</sup> were burned more than ten times (Figure 3B). Higher fire frequencies were rare and spatially restricted to the north of São Paulo, specifically to "Ribeirão Preto" and "Barretos" regions, both dominated mainly by sugar-cane croplands.

Our results indicates that most of São Paulo's Cerrado area (62.7%) was not subjected to any fire since 1985, probably as consequence of the fire suppression policies and the expansion of agricultural lands. Low fire frequencies were expected in high managed landscapes, especially in lands with economic use, and higher fire frequencies are usually associated with specific land uses (e.g. cropland and pasture) (Archibald et al., 2010; Mataveli et al., 2018). Considering that a natural fire regime is characterized by a fire occurs every 5 years (Dias, 2006), only 1.1% of the São Paulo's Cerrado have satisfied this criteria, burning 7 to 9 nine times in the last 34 years. However, we cannot consider these 1.1% as a "Cerrado under natural fire regime" because most of these areas are not covered by native remnants vegetation, but mainly covered by pasture and croplands.

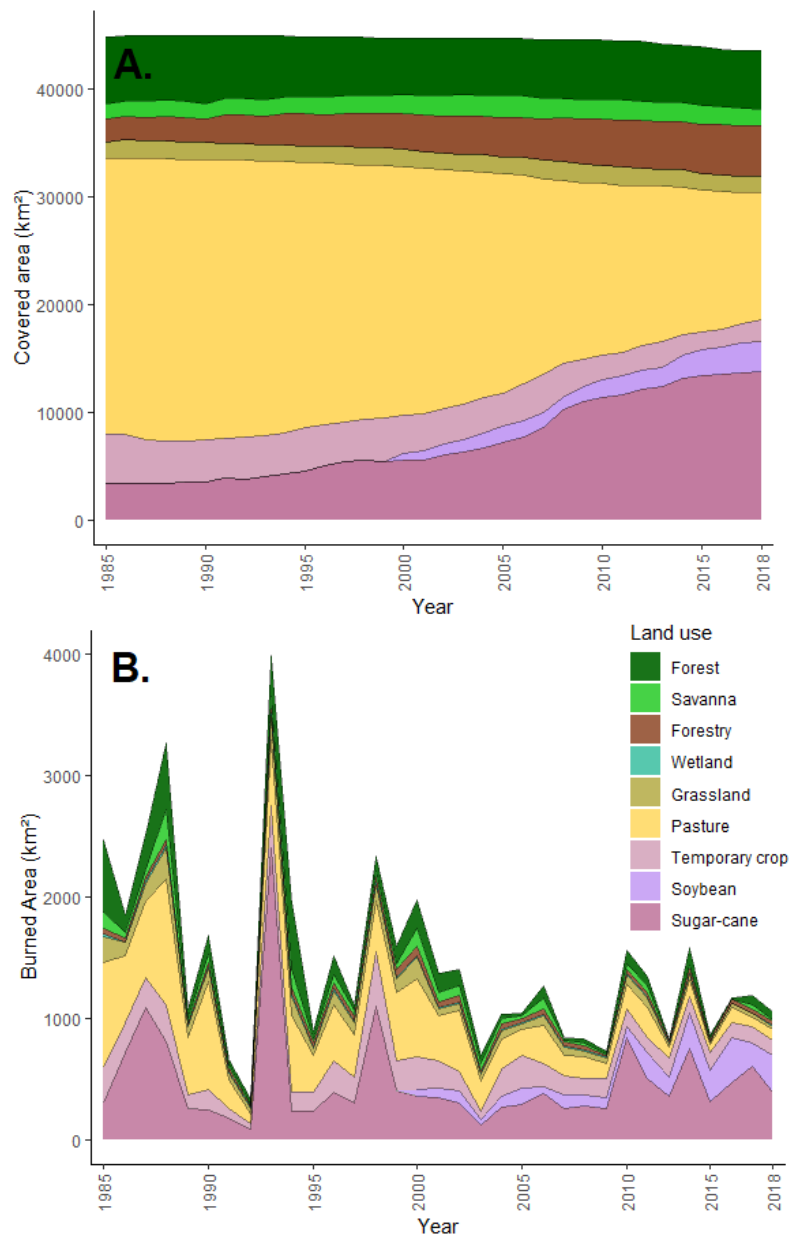


**Figure 3.** General burned area pattern from the highly anthropized Cerrado. **A)** Map of fire count from 1985 to 2018. Gray lines represent the boundaries of São Paulo state. Black labels represent regional towns as spatial reference. Each 30-meter pixel in the São Paulo's Cerrado (IBGE) was colored from yellow (low fire count) to purple (high fire count) considering the number that every pixel burnt. **B)** Histogram of accumulated burned area (km<sup>2</sup>) by fire count.

Considering the temporal distribution of LCLUC in the São Paulo's Cerrado, we detected that native vegetation classes (forest, savanna and grassland) were relatively stable from 1985 to 2018. In contrast, large variations were detected in anthropogenic land uses. We observed a decrease of 54% in the total area occupied by pastures and an increase of 147% in the total area of sugar-cane croplands from 1985 to 2018 (Figure 4A). Aiming to diversify the energy industry and develop the countryside in the 80's, São Paulo's government implemented the "PROALCOOL" – a large tax incentive program for the production of fuel ethanol (Carlos et al., 2006). One of the direct results from this policy was the massive conversion of pasture areas into sugar-cane croplands (Bray et al., 2000), as we observed.

Despite fire has been largely used as a tool to renew pastures in the Cerrado (Coutinho, 1990; Dias, 2006), as we previously described, we found a trend of gradual decrease in total burned area across time (1985= 2473 km<sup>2</sup>, 2000= 1976 km<sup>2</sup> and 2018= 1053 km<sup>2</sup>). In 1992 we observed the lowest burned area (334 km<sup>2</sup>) while 1993 had the highest burned area (3996 km<sup>2</sup>). Interestingly, with the gradual

replacement of pastures by croplands in São Paulo state across time (Figure 4A), we also detected a decrease in yearly burned area (Figure 4B). On the other hand, because fire is often used to facilitate sugar-cane harvesting, we expected the maintain of fire occurrence across time. However, we didn't observe that, suggesting that fire associated to pastures is a most important practice when compared to its agricultural use. Moreover, We observed that despite the sugar-cane cropland area was lower in the 80's and 90's in relation with the current distribution, we detected some fire peaks on this LCLUC in 1988, 1993 and 1998, corresponding to the final years of moderate and strong El Niño events (CPTEC/INPE, 2020).



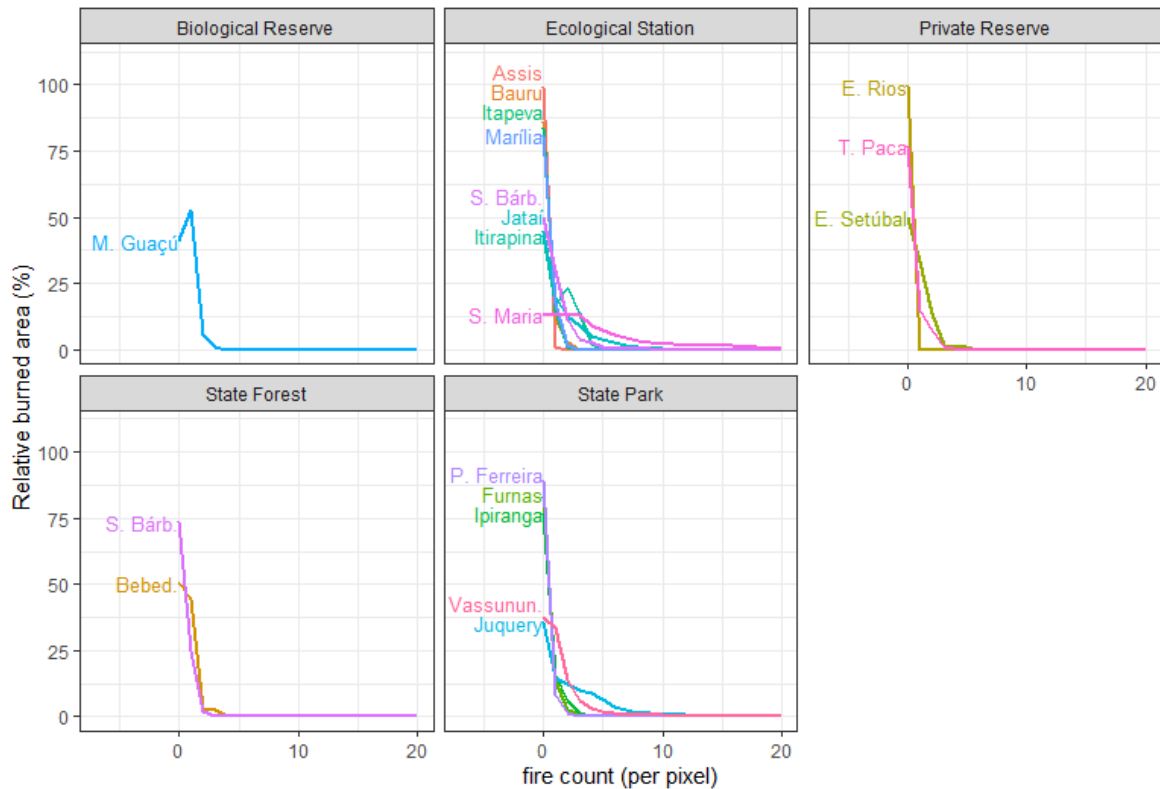
**Figure 4.** Burned area general patterns during the period between 1985 and 2018. **A)** Distribution of the main land cover and land uses in São Paulo's Cerrado. **B)** Burned area by land cover class across time.

To mitigate health and environmental impacts caused by the uncontrolled use of fire in sugar-cane harvesting, the state law 10 547 /2001 introduced legal constraints to reduce fire use as agricultural practice. That could explain why we observed a relative “stability” in total burned area from 2000. However, the uncontrolled increasing of sugar-cane croplands, especially in areas where fire is allowed (e.g. highly slope areas), was probably the cause of the peaks in total burned area in 2010 and 2014.

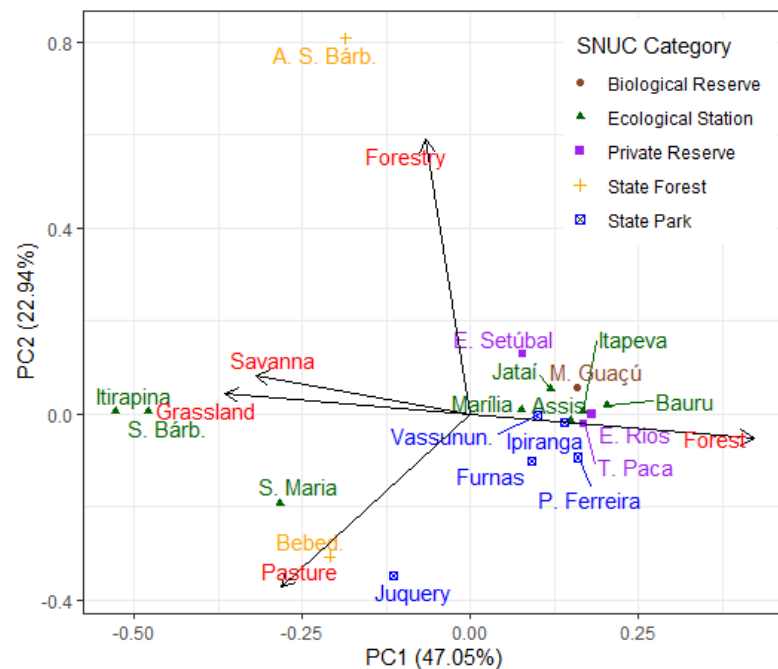
#### *Fire frequency within protected areas*

According to the general patterns identified for the highly anthropized Cerrado in Sao Paulo state, all the protected areas have showed low fire frequencies in the last three decades, independently of its SNUC category of protection (Figure 5). Considering the PAs with the lowest fire frequencies, we constated that the Entre Rios Private Reserve have not been subjected to any fire during the studied period, followed by Assis Ecological Station (99.1% of unburned area), Porto Ferreira State Park (89.5%), Bauru Ecological Station (86.1%) and Itapeva Ecological Station (83.9%). On the other hand, the protected areas with the highest fire frequency were Santa Maria Ecological Station (61% of total area burned three or more times), Juquery State Park (38% burned 3 times or more), Itirapina Ecological Station (17% burned 3 times or more) and Santa Bárbara Ecological Station (11% burned 3 times or more).

We performed two Principal Component Analyses (PCAs) considering the proportion (%) of land cover classes for all the protected areas in 2018 (Figure 6) and 1985 (Supplementary Figure S2). First, we did not detect a significative structural variation in vegetation cover for the PAs from 1985 to 2018, except for the cases of Santa Maria Ecological Station and Juquery State Park where grasslands and savannas were converted into pastures. Second, we found that the vegetation structure of the PAs is highly associated with fire occurrence, thus we found the higher fire frequencies in areas that varied between pasture and grasslands. As expected, we found the lowest fire frequencies in the PAs covered by forest vegetation (Figure 6).



**Figure 5.** Relative burned area (%) across a gradient of fire counts for each protected area (PA). PA were separated in different boxes according with its SNUC protection categories. Labels inside boxes indicate the name of each PA following the table 1 description.



**Figure 6.** Principal Component Analysis (PCA) considering the proportion of predominantly land cover and land use change (LCLUC) by protected area in 2018 (70% of explanation). Black arrows represent the loadings of each variable; Red labels shows the name of each variable (class of land cover); the labels highlight each PA and its colors represent the SNUC category.

Considering the PAs with the highest fire frequencies in São Paulo's Cerrado, we cannot affirm that these areas have a high fire frequency when compared with the results for other Cerrado PAs (Alvarado et al., 2018; Alves & Pérez-Cabello, 2017; Augusto Verola Mataveli et al., 2018). We can only consider the particular cases of Santa Maria Ecological Station and Juquery State Park as a high fire frequency for three reasons: 1) both presented areas that have burned more than 7 times in 34 years; ii) these areas also presented ~20% of their areas covered by pasture, and; iii) these PA are inserted into a region with a context of social conflicts. In Santa Maria, the social conflicts have a land use issue, originated since 1998 when some social movements linked to the agrarian reform agenda have occupied and claimed the use of the PA for agricultural purposes (Povo, 2006). These movements did not succeed to find an agreement with the government, and fire has become a tool for political pressure until present time. On the other hand, Juquery is inserted into a dense urban zone with highways crossing their limits (increasing accidental fires). Besides, a prison complex is located within Juquery State Park, with some related inmate escapes who set the fire in native vegetation as distraction (G1, 2016; R7, 2017).

It is important to highlight that the transitions from grassland/savanna to pasture mapped by using the MapBiomass product for Santa Maria and Juquery is mainly caused by an invasion and very fast colonization of african grasses (*Urochloa brizantha* and *Melinis minutiflora*) in areas without cattle livestock. Thanks to its plant strategies of fast colonization and high biomass production, these african grasses have suppressed the native grasses, herbs and shrubs, causing a decrease of biodiversity and an increase of fuel build-up (Damasceno et al., 2018; Durigan et al., 2007). Comparing fire behavior in invaded areas by african grasses vs. native grasslands, fire is able to reach higher temperatures in areas dominated by African grasses, increasing consequently fire severity and damages of vegetation (Gorgone-Barbosa et al., 2015). To exacerbate this situation, post-fire conditions can enhance African grasses seed germination, increasing their invasiveness and accelerating the Cerrado's degradation (Gorgone-Barbosa et al., 2016).

In the case of Santa Maria and Juquery, we can suggest the relationship between african grasses invasion and high fire frequencies as a trade-off. Areas with invasive grasses are more susceptible to accidental and arson fires due to its fast fuel



build-up. Consequently, the expansion and densification of invasive grasses are favored by high fire frequencies. This scenario creates a conceptual dilemma: How to ensure Cerrado grasslands conservation when fire, its evolutionary driver, becomes its biggest enemy?

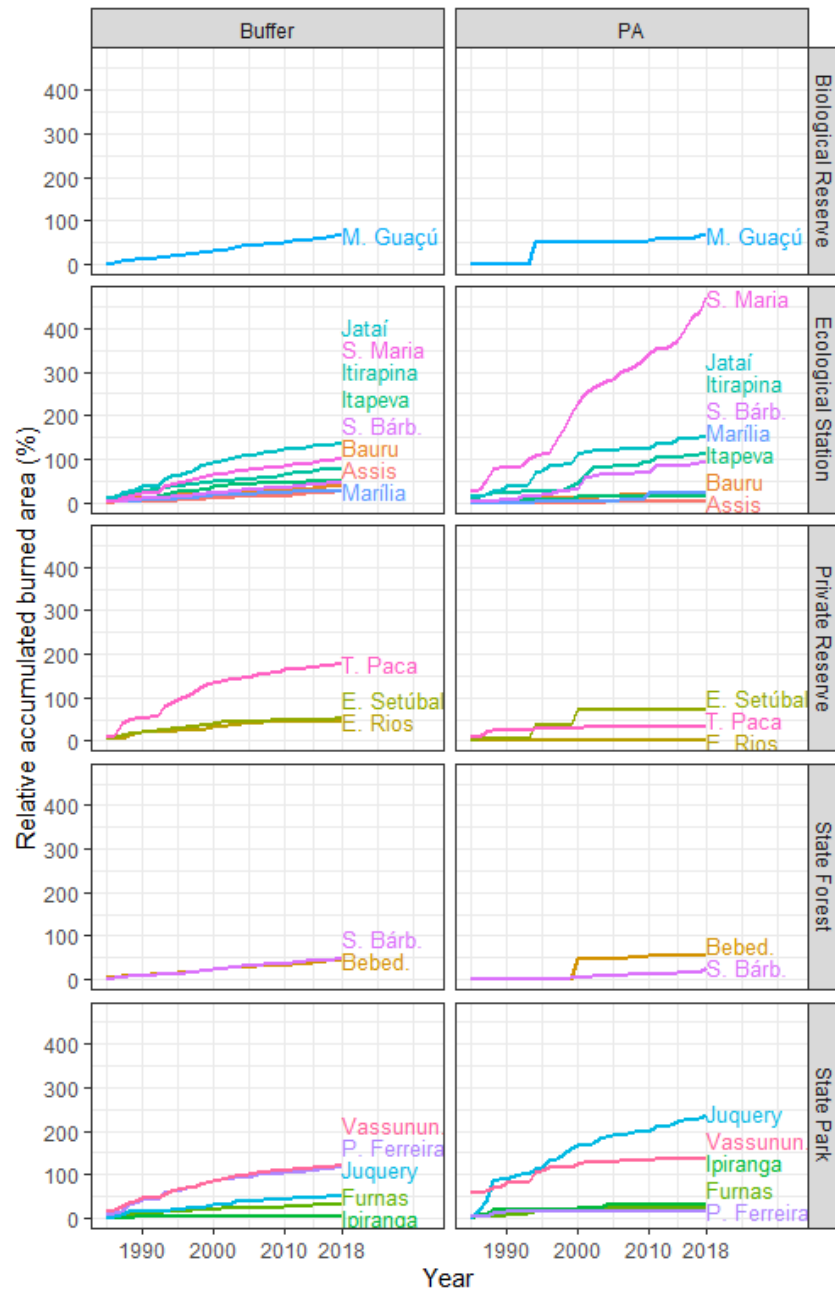
*Spatio-temporal variation of burned areas across PAs and its buffer zones*

We detected 222 km<sup>2</sup> of burned areas within the protected areas vs. 6 736 km<sup>2</sup> of burned areas in its buffer zones during the period between 1985 and 2018. Despite the buffer zones have presented an absolute sum of burned areas 30 times greater than the PAs, when we consider the relative areas ( = total burned area / total area of PA or Buffer), this situation is reversed and we observed 78% of the PAs that was burned vs. 70% of its buffer zones.

When assessing the relative accumulated burned area for each PAs, we observed that Santa Maria Ecological Station was the PA that burned the most in the past 34 years (62 km<sup>2</sup>= 473%), followed by Juquery Ecological Station (47 km<sup>2</sup>= 238%), Jataí Ecological Station (135 km<sup>2</sup>= 150%) and Itirapina Ecological Station (25 km<sup>2</sup>= 117%). The highest relative accumulated burned areas on the buffer zones was found in Toca do Paca Private Reserve (668 km<sup>2</sup> = 178%), Jataí Ecological Station (1204 km<sup>2</sup> = 136%), Vassununga State Park (875 km<sup>2</sup> = 120%) and Porto Ferreira State Park (530 km<sup>2</sup> = 116%). Our results show that despite a rigorous fire exclusion policy implemented by the PA managers, these areas tends to burn relatively more than their buffer zones (Figure 7) where fire use is allowed under legal authorization.

Even, different patterns of burned area were detected for PAs and its buffer zones across the time. Buffer zones have showed a continuous pattern of burned areas across the entire period, with lower variation of total burned area between years, highlighting “linear” patterns for all the buffer zones. Contrary, we found a pattern of burned area “peaks” within most of the PAs. These fire peaks range from decades in the PAs with lower fire frequencies to intervals between 4-5 years in the PAs with higher fire frequencies. These intervals without fire can be explained by the fire exclusion policies implemented within the PAs (Durigan & Ratter, 2016), with larger intervals without fire detected in PAs with forest vegetation and shorter intervals detected into PAs with predominance of grassland and invasion by african

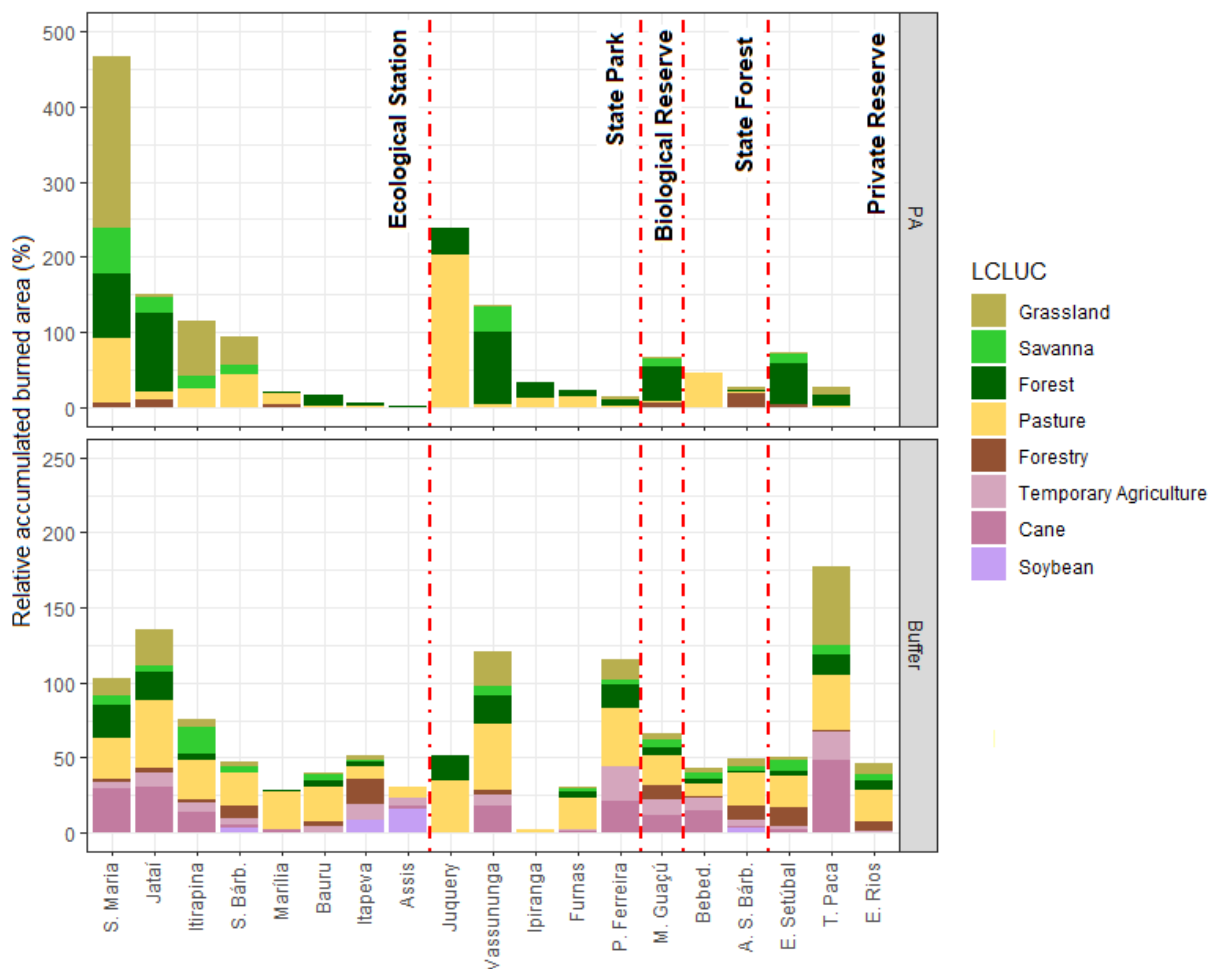
grasses. Furthermore, while we found a distribution of burned areas across every months (including wet and dry season) in the buffer zones, most of the detected burned areas within PAs were detected in mid-late dry season (from July to September) (Supplementary Figure S3).



**Figure 7.** Relative accumulated burned area (%) over time (years) for each protected areas (PA- right boxes) and its respective buffer zones (Buffer- left boxes). Labels in the right margins refers to the SNUC category of protection.

As expected, the PAs that showed higher fire frequencies have also showed largest accumulated burned areas (Santa Maria Ecological Station and Juquery State

Park). Contrary to the peaks pattern described for the other PAs, both areas showed to be exceptions and followed the trends of its buffer zones (continuous and stable burned areas across time). These PA have a context of conflict among the PA managers and the neighbor communities. Additionally, when we assess the types of land cover that was burned on these PAs during the last three decades, we found that most of fires on these PAs were on areas covered by native grasslands and invaded areas by african grasses (treated as “pasture” in this study), while most of fires in the buffer zones occurs in areas covered by pasture and sugar-cane, following the general pattern previously related for the highly anthropized Cerrado (Figure 8).



**Figure 8.** Relative accumulated burned area (%) by land cover and land use (LCLUC) for each PAs (top) and its buffer zones (bottom) between 1985 and 2018. Note the range of values in the y-axis from the buffer zones corresponds to a half of the values from the PAs.

Interestingly, our results show that the full preservation areas have presented larger burned area and higher fire frequencies than sustainable use areas. Most of full preservation areas have burned more than their buffer zones, creating a paradox scenario: if these areas have a stronger and strict conservation status and more

restrictive laws are implemented, how they can burn relatively more than sustainable use areas and unprotected areas? This has been a historical and political issue. First, following the “myth of the untouchable nature” (Diegues, 2001), the rare remnants of open ecosystems in the São Paulo’s Cerrado were protected under the most restrictive SNUC categories in the 80’s. Second, this management strategy follows the premise of total isolation of these areas, something that is impossible considering the fact that these areas are represented by small fragments surrounded by an anthropogenic matrix. Finally, other studies have highlighted the level of protection (and restrictiveness) is inversely proportional that the perception of conservation importance of these areas by the local communities and directly proportional to the probability of conflicts between local people and managers (Fiori et al., 2006; Ferreira, 2005; Silva et al., 2009).

PAs mainly covered by forest (“cerradão” and ecotonal zones with Atlantic Rainforest) showed lower fire frequencies and accumulated burned areas than its buffer zones, independently of the SNUC protection level. However, we observed three exceptions: Jataí Ecological Station, Vassununga State Park and Mogi-Guaçu Biological Reserve. These PAs have presented peaks of burned area of native forest in the end of 90’s (equivalent to 120%, 130% and 50% of their total areas, respectively), with few remaining areas without or with a small percent of burned area during the last two decades. In all the cases, fire have affected areas of transition between savannas and forest (“cerradão”), associated to dry events, and also burning mainly disturbed forest areas with lower tree canopy.

#### *Management implications and future perspectives*

The history of land use and the transformation of land cover for agriculture on São Paulo's Cerrado created a highly fragmented landscape that can acts as barriers for fire spread (Andela & van der Werf, 2014; Archibald, 2016). In a temporal scale, this fragmentation caused the decrease of burned area, especially considering the conversion of pastures into sugar-cane croplands, just as we observed in the highly anthropized Cerrado of Sao Paulo.

In the context of highly human managed landscapes, fuel build-up trends to be lower when compared to natural areas. Residual organic matter is mechanically harvested and used as natural fertilizer in temporary croplands of sugar-cane and

soybean (Martins et al., 1999). Also, cattle consume most of biomass into pastures, having none or low fuel build-up rate available to burn during the dry season. Considering that the protected areas are small patches of native vegetation under a fire exclusion policy, these PAs are inserted into the landscape highly flammable with enough dry out vegetation to carry fire, especially those covered by grassland and african grasses (higher fuel build-up than the buffer zones) which are more susceptible to wildfires on the late dry season (Conciani et al., 2021). Fires in the late dry season are hotter, more severe and burn large areas in the Cerrado (Coutinho, 1990). Besides, most of fire occurrences that involves fire brigades and direct firefighting that have been officially registered by PA managers occurs in the late dry season (DATAGEO, 2021), pointing out that fire exclusion alone was not successful to prevent wildfires on the Cerrado and conversely created a repeated scenario of large wildfires as has been observed in other Cerrado Areas (Alvarado et al., 2018; Batista et al., 2018).

Considering the PAs dominated by grassland and savanna vegetation, we observed two different situations: i) the case of PAs dominated by african grasses that burned with a high fire frequency and; ii) PAs that had a lower fire frequency than the historical Cerrado (Dias, 2006), evidencing a process of woody encroachment and open ecosystem biodiversity loss (Oliveira & Marquis, 2002; Stevens et al., 2017). For this last case, on both Itirapina and Santa Bárbara Ecological Stations, it is highly recommended the development of an adaptive management plan that considers the use of prescribed fires spatially distributed in mosaics over the PA, especially in areas of grassland (“campo limpo”) and open savanna (“campo sujo”). This strategy has results effective for wildfires prevention by breaking the spatial continuity of fuel load (Franke et al., 2018) as well as improving the abundance and richness of forbs and graminoids in these areas (Durigan et al., 2020).

On the other hand, management strategies that aims to increase conservation effectiveness in PAs with high fire frequency and large invaded areas by african grasses needs to be carefully planned. Despite the positive effects of fire on open ecosystems that have been previously reported (Abreu et al., 2017; Durigan et al., 2020; Fidelis et al., 2019; Fidelis & Pivello, 2011), we need to consider that the densely invaded areas are environmentally degraded and should be treated as such. To treat the degraded areas by using the same conservation policies and strategies that are used to manage Cerrado’s native areas could result in an error and these

conservation intervention can be completely inefficient. Par example, fire occurrence have a high potential to favor african grasses over native grasses (Damasceno et al., 2018; Gorgone-Barbosa et al., 2015, 2016), further aggravating the environmental problem of these areas. We suggest to acknowledge these degraded areas need an active intervention to successfully reach its conservation purpose and thus managers should center the conservation efforts looking for alternatives to restore these areas.

Most of these degraded areas are currently managed by following the same non active-intervention policy applied within full preservation PAs. This could be a misconception since most of these densely invaded areas are located in the border between the PA and the neighbor human communities. These degraded areas inside the PAs accumulate high amounts of fuel during the annual cycles, and the absence of management by the conservation system is often interpreted by local communities as abandoned by the government being seen as unproductive lands. This perception of abandon and unproductiveness of these degraded areas have created a historical scenario of conflict between managers and local population, with constant human invasions and unauthorized uses of these areas.

Since these social conflicts already exists, mainly based in the differences of perception between local communities and the conservation goals of PA managers, every management plan biased only by institutional interest is subjected to failure. New strategies are thus need to consider social and cultural aspects, and local communities should be included in the management councils to participate on the discussions about the management decisions. Until this happens, fire will continue to be used as a political pressure tool (Kull, 2002), favoring the invasion of african grasses and compromising the effective conservation of these PAs.

### *Conclusion*

São Paulo's Cerrado fire regime is mainly drive by anthropogenic factors and most of burned areas in the last three decades are concentrated in pastures and sugar-cane croplands. Variations in burned area across time have follow the variations in land use and land cover, highlighting the anthropogenic drivers of fire regime.

Protected areas covered by forest formations burned rarely across time, except for degraded areas. Conversely, protected areas dominated by grassland

formation burned more than its buffer zones. Surprisingly, all the protected areas with higher fire frequencies and larger burned areas were categorized into the most restrictive levels of conservation (SNUC), questioning the effectiveness of the national fire exclusion policy.

Since all PAs on savanna and grassland formation are periodically burning, management policies needs change to allow that managers apply prescribed fires to generate a landscape mosaic of burned and unburned areas, and thus reduce large and uncontrollable fire occurrences, increasing conservation effectiveness. Managers also should recognize that areas with dominance of african grasses do not fulfil its conservation purpose and needs to be restored. Strategies of management needs to be rethink by the decision makers and PA managers to include the local communities as active actors of conservation as a way to decrease social tensions, change the perception of conservation about these protected areas and mitigate conflicts.

### *Acknowledgments*

We thank to Instituto Florestal and Fundação Florestal for providing historical and geospatial data about the protected areas and for logistical assistance. D.E.C. is currently receiving financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) financial code 001 and received technical scholarship support from Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grant#441968/2018-0).

### *References*

- Abreu, R. C. R., Hoffmann, W. A., Vasconcelos, H. L., Pilon, N. A., Rossatto, D. R., & Durigan, G. (2017). The biodiversity cost of carbon sequestration in tropical savanna. *Science Advances*, 3(8), e1701284. <https://doi.org/10.1126/sciadv.1701284>
- Alencar, A., Z. Shimbo, J., Lenti, F., Balzani Marques, C., Zimbres, B., Rosa, M., Arruda, V., Castro, I., Fernandes Márcico Ribeiro, J. P., Varela, V., Alencar, I., Piontekowski, V., Ribeiro, V., M. C. Bustamante, M., Eyji Sano, E., & Barroso, M. (2020). Mapping Three Decades of Changes in the Brazilian Savanna Native Vegetation Using Landsat Data Processed in the Google Earth Engine Platform. *Remote Sensing*, 12(6), 924. <https://doi.org/10.3390/rs12060924>
- Alvarado, S. T., Andela, N., Silva, T. S. F., & Archibald, S. (2020). Thresholds of fire response to moisture and fuel load differ between tropical savannas and grasslands across continents. *Global Ecology and Biogeography*, 29(2), 331–344. <https://doi.org/10.1111/geb.13034>
- Alvarado, S. T., Fornazari, T., Cóstola, A., Morellato, L. P. C., & Silva, T. S. F. (2017). Drivers of fire occurrence in a mountainous Brazilian cerrado savanna: Tracking long-term fire regimes using

- remote sensing. *Ecological Indicators*, 78, 270–281.  
<https://doi.org/10.1016/J.ECOLIND.2017.02.037>
- Alvarado, S. T., Silva, T. S. F., & Archibald, S. (2018). Management impacts on fire occurrence: A comparison of fire regimes of African and South American tropical savannas in different protected areas. *Journal of Environmental Management*, 218, 79–87.  
<https://doi.org/10.1016/J.JENVMAN.2018.04.004>
- Alves, D. B., & Pérez-Cabello, F. (2017). Multiple remote sensing data sources to assess spatio-temporal patterns of fire incidence over Campos Amazônicos Savanna Vegetation Enclave (Brazilian Amazon). *Science of the Total Environment*, 601–602, 142–158.  
<https://doi.org/10.1016/j.scitotenv.2017.05.194>
- Andela, N., & van der Werf, G. R. (2014). Recent trends in African fires driven by cropland expansion and El Niño to La Niña transition. *Nature Climate Change*, 4(9), 791–795.  
<https://doi.org/10.1038/nclimate2313>
- Archibald, S., Nickless, A., Govender, N., Scholes, R. J., & Lehsten, V. (2010). Climate and the inter-annual variability of fire in southern Africa: a meta-analysis using long-term field data and satellite-derived burnt area data. *Global Ecology and Biogeography*, 19(6), 794–809.  
<https://doi.org/10.1111/j.1466-8238.2010.00568.x>
- Archibald, Sally. (2016). Managing the human component of fire regimes: lessons from Africa. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1696), 20150346. <https://doi.org/10.1098/rstb.2015.0346>
- Arévalo, J., & Naranjo-Cigala, A. (2018). Wildfire Impact and the “Fire Paradox” in a Natural and Endemic Pine Forest Stand and Shrubland. *Fire*, 1(3), 44. <https://doi.org/10.3390/fire1030044>
- Augusto Verola Mataveli, G., Elisa Siqueira Silva, M., Pereira, G., da Silva Cardozo, F., Shinji Kawakubo, F., Bertani, G., Cezar Costa, J., de Cássia Ramos, R., & Valéria da Silva, V. (2018). Satellite observations for describing fire patterns and climate-related fire drivers in the Brazilian savannas. *Hazards Earth Syst. Sci*, 18, 125–144. <https://doi.org/10.5194/nhess-18-125-2018>
- Batista, E. K. L., Russell-Smith, J., França, H., & Figueira, J. E. C. (2018). An evaluation of contemporary savanna fire regimes in the Canastra National Park, Brazil: Outcomes of fire suppression policies. *Journal of Environmental Management*, 205, 40–49.  
<https://doi.org/10.1016/J.JENVMAN.2017.09.053>
- Bradstock, R. A. (2010). A biogeographic model of fire regimes in Australia: current and future implications. *Global Ecology and Biogeography*, 19(2), 145–158. <https://doi.org/10.1111/j.1466-8238.2009.00512.x>
- Bray, S. C., Ferreira, E. R., & Ruas, D. G. G. (2000). *As políticas da agroindústria canavieira eo proálcool no Brasil*. Unesp-Marília Publicações.
- Carlos, A., Júnior, F., & Hespanhol, A. N. (2006). *OS EFEITOS DAS POLÍTICAS VOLTADAS AO SETOR SUCROALCOOLEIRO NO ESTADO DE SÃO PAULO 1* (Vol. 1, Issue 6).
- Conciani, D., Santos, L., Silva, T. S. F., Durigan, G., & Alvarado, S. T. (2021). Human-Climate Interactions Shape Fire Regimes in the Cerrado of São Paulo state, Brazil. *Journal for Nature Conservation*.



- Coutinho, L. M. (1990). *Fire in the Ecology of the Brazilian Cerrado* (pp. 82–105). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-75395-4\\_6](https://doi.org/10.1007/978-3-642-75395-4_6)
- Coutinho, Leopoldo Magno. (1990). O cerrado e a ecologia do fogo. *Ciência Hoje*, 12(68), 23–30.
- CPTEC/INPE. (2020). *El Niño e La Niña - CPTEC/INPE*. <http://enos.cptec.inpe.br/>
- Damasceno, G., Souza, L., Pivello, V. R., Gorgone-Barbosa, E., Giroldo, P. Z., & Fidelis, A. (2018). Impact of invasive grasses on Cerrado under natural regeneration. *Biological Invasions*, 20(12), 3621–3629. <https://doi.org/10.1007/s10530-018-1800-6>
- DATAGEO. (2021). *Infraestrutura de Dados Espaciais Ambientais do Estado de São Paulo – IDEA-SP - DataGeo*. <https://datageo.ambiente.sp.gov.br/>
- De Fiori, A., Prof, O., Eduardo, J., & Santos, D. (2006). UNIVERSIDADE FEDERAL DE SÃO CARLOS CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E RECURSOS NATURAIS A PERCEPÇÃO AMBIENTAL COMO INSTRUMENTO DE APOIO DE PROGRAMAS DE EDUCAÇÃO AMBIENTAL DA ESTAÇÃO ECOLÓGICA DE JATAÍ (LUIZ ANTÔNIO, SP).
- Dias, B. F. (2006). *Degradação ambiental: Os impactos do fogo sobre a diversidade do cerrado*. In I. Garay and B. Becker (Eds.), *Dimensões Humanas da Biodiversidade: O Desafio de Novas Relações Homem-Natureza no Século XXI*. Editora Vozes.
- Diegues, A. C. S. (2001). *O mito moderno da natureza intocada* (Vol. 4). Hucitec São Paulo.
- Durigan, G., Pilon, N. A. L., Abreu, R. C. R., Hoffmann, W. A., Martins, M., Fiorillo, B. F., Antunes, A. Z., Carmignotto, A. P., Maravalhas, J. B., Vieira, J., & Vasconcelos, H. L. (2020). No Net Loss of Species Diversity After Prescribed Fires in the Brazilian Savanna. *Frontiers in Forests and Global Change*, 3, 13. <https://doi.org/10.3389/ffgc.2020.00013>
- Durigan, G., & Ratter, J. A. (2016). The need for a consistent fire policy for Cerrado conservation. *Journal of Applied Ecology*, 53(1), 11–15. <https://doi.org/10.1111/1365-2664.12559>
- Durigan, G., Siqueira, M. F. de, & Franco, G. A. D. C. (2007). Threats to the Cerrado remnants of the state of São Paulo, Brazil. *Scientia Agricola*, 64(4), 355–363. <https://doi.org/10.1590/S0103-90162007000400006>
- Ferreira, C. P. (2005). *Percepção Ambiental na Estação Ecológica de Jureia - Itatins* [Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo]. <https://doi.org/10.11606/D.90.2005.tde-01122008-154923>
- Fidelis, A., & Pivello, V. R. (2011). Biodiversidade Brasileira. In *Biodiversidade Brasileira* (Vol. 0, Issue 2). <http://www.icmbio.gov.br/revistaeletronica/index.php/BioBR/article/view/102>
- Fidelis, A., Rosalem, P., Zanzarini, V., Camargos, L. S., & Martins, A. R. (2019). From ashes to flowers: a savanna sedge initiates flowers 24 h after fire. *Ecology*, 100(5), e02648. <https://doi.org/10.1002/ecy.2648>
- Fiori, A. N., & Fioravanti, C. (2001). Os caminhos para salvar o cerrado paulista. *FAPESP Pesquisa, São Paulo*, 63, 38–43.

- Franke, J., Barradas, A. C. S., Borges, M. A., Menezes Costa, M., Dias, P. A., Hoffmann, A. A., Orozco Filho, J. C., Melchiori, A. E., & Siegert, F. (2018). Fuel load mapping in the Brazilian Cerrado in support of integrated fire management. *Remote Sensing of Environment*, 217, 221–232. <https://doi.org/10.1016/J.RSE.2018.08.018>
- G1. (2016). *G1 - Helicóptero flagra fugitivos em mata incendiada perto de presídio em SP - notícias em São Paulo*. <http://g1.globo.com/sao-paulo/noticia/2016/10/helicoptero-flagra-fugitivos-em-mata-incendiada-perto-de-presidio-em-sp.html>
- Gorgone-Barbosa, E., Pivello, V. R., Baeza, M. J., Fidelis, A., Gorgone-Barbosa, E., Pivello, V. R., Baeza, M. J., & Fidelis, A. (2016). Disturbance as a factor in breaking dormancy and enhancing invasiveness of African grasses in a Neotropical Savanna. *Acta Botanica Brasilica*, 30(1), 131–137. <https://doi.org/10.1590/0102-33062015abb0317>
- Gorgone-Barbosa, E., Pivello, V. R., Bautista, S., Zupo, T., Rissi, M. N., & Fidelis, A. (2015). How can an invasive grass affect fire behavior in a tropical savanna? A community and individual plant level approach. *Biological Invasions*, 17(1), 423–431. <https://doi.org/10.1007/s10530-014-0740-z>
- Guidon, N. (1992). As ocupações pré-históricas do Brasil (excetuando a Amazônia). *História Dos Índios No Brasil*, 2, 37–52.
- IBGE. (2014). *Cidades do Brasil*. <https://cidades.ibge.gov.br/>
- Kronka, F. J. N., Nalon, M. A., Matsukuma, C. K., Kanashiro, M. M., Ywane, M. S. S., Lima, L., Guillaumon, J. R., Barradas, A. M. F., Pavão, M., & Manetti, L. A. (2005). Monitoramento da vegetação natural e do reflorestamento no Estado de São Paulo. *Simpósio Brasileiro de Sensoriamento Remoto*, 12, 16–21.
- Kull, C. A. (2002). Madagascar aflame: Landscape burning as peasant protest, resistance, or a resource management tool? *Political Geography*, 21(7), 927–953. [https://doi.org/10.1016/S0962-6298\(02\)00054-9](https://doi.org/10.1016/S0962-6298(02)00054-9)
- Lentile, L. B., Holden, Z. A., Smith, A. M. S., Falkowski, M. J., Hudak, A. T., Morgan, P., Lewis, S. A., Gessler, P. E., & Benson, N. C. (2006). Remote sensing techniques to assess active fire characteristics and post-fire effects. *International Journal of Wildland Fire*, 15(3), 319. <https://doi.org/10.1071/WF05097>
- Martins, D., Velini, E. D., Martins, C. C., De Souza, L. S., & Jovem Pesquisador, B. (1999). EMERGÊNCIA EM CAMPO DE DICOTILEDÔNEAS INFESTANTES EM SOLO COBERTO COM PALHA DE CANA-DE-AÇÚCAR 1. In *Planta Daninha* (Issue 1).
- Mistry, J., Berardi, A., Andrade, V., Krahô, T., Krahô, P., & Leonardos, O. (2005). Indigenous fire management in the cerrado of Brazil: The case of the Krahô of Tocantins. *Human Ecology*, 33(3), 365–386. <https://doi.org/10.1007/s10745-005-4143-8>
- Murphy, B. P., Andersen, A. N., & Parr, C. L. (2016). The underestimated biodiversity of tropical grassy biomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1703), 20150319. <https://doi.org/10.1098/rstb.2015.0319>
- Oliveira, P. S., & Marquis, R. J. (2002). *The cerrados of Brazil : ecology and natural history of a neotropical savanna*. Columbia University Press. [https://books.google.com.br/books?hl=pt-BR&lr=&id=TXvY\\_kAFPUOC&oi=fnd&pg=PA159&dq=HOFFMANN,+William+A.%3B+MOREIRA,+A](https://books.google.com.br/books?hl=pt-BR&lr=&id=TXvY_kAFPUOC&oi=fnd&pg=PA159&dq=HOFFMANN,+William+A.%3B+MOREIRA,+A)

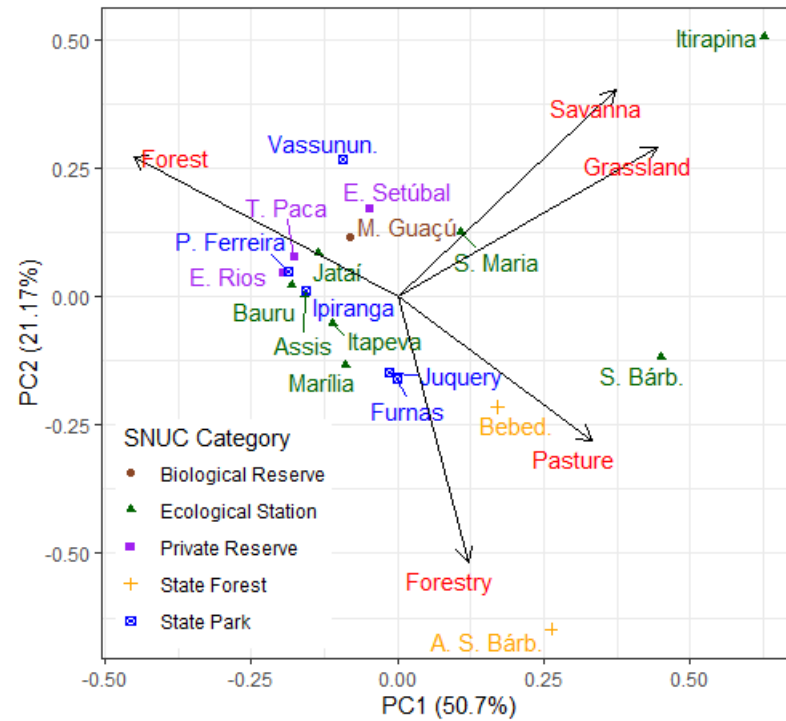
driana+G.+The+role+of+fire+in+population+dynamics+of+woody+plants.+The+Cerrados+of+Brazil.+Ecology+and+Natural+History+of+a+Neotropical+Savanna,+p.+159-177,+2002.&ots=\_HuLK8HL7q&sig=RjM-OdKdC1RiZYbG8xYOC47gWo4#v=onepage&q&f=false

- Povo, G. do. (2006). *Sem-terras destruíram 40 anos de pesquisa, diz diretor da Estação Experimental São Simão*. <https://www.gazetadopovo.com.br/vida-publica/sem-terras-destruiram-40-anos-de-pesquisa-diz-diretor-da-estacao-experimental-sao-simao-ab5xkdq02tlzbqmv8vyiwww9a/>
- R7. (2017). *Pelo menos 27 presos fogem de penitenciária em Franco da Rocha - São Paulo - Estadão*. <https://sao-paulo.estadao.com.br/noticias/geral,pelo-menos-27-presos-fogem-de-penitenciaria-em-franco-da-rocha,70001938202>
- Ribeiro, J. F., & Walter, B. M. T. (2017). *Fitofisionomias do bioma cerrado*. <https://www.alice.cnptia.embrapa.br/handle/doc/554094>
- Silva, T. S. da, Cândido, G. A., & Freire, E. M. X. (2009). Conceitos, percepções e estratégias para conservação de uma estação ecológica da Caatinga nordestina por populações do seu entorno. *Sociedade & Natureza*, 21(2), 23–37. <https://doi.org/10.1590/s1982-45132009000200003>
- Simon, M. F., Grether, R., Queiroz, L. P. de, Skema, C., Pennington, R. T., & Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences*, pnas.0903410106. <https://doi.org/10.1073/PNAS.0903410106>
- Soares-Filho, B., Rajão, R., Macedo, M., Carneiro, A., Costa, W., Coe, M., Rodrigues, H., & Alencar, A. (2014). Cracking Brazil's forest code. *Science*, 344(6182), 363–364.
- Souza, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F. T., Hasenack, H., Matsumoto, M., G. Ferreira, L., Souza-Filho, P. W. M., de Oliveira, S. W., Rocha, W. F., Fonseca, A. V., Marques, C. B., Diniz, C. G., Costa, D., Monteiro, D., Rosa, E. R., Vélez-Martin, E., ... Azevedo, T. (2020). Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing*, 12(17), 2735. <https://doi.org/10.3390/rs12172735>
- Stevens, N., Lehmann, C. E. R., Murphy, B. P., & Durigan, G. (2017). Savanna woody encroachment is widespread across three continents. *Global Change Biology*, 23(1), 235–244. <https://doi.org/10.1111/gcb.13409>

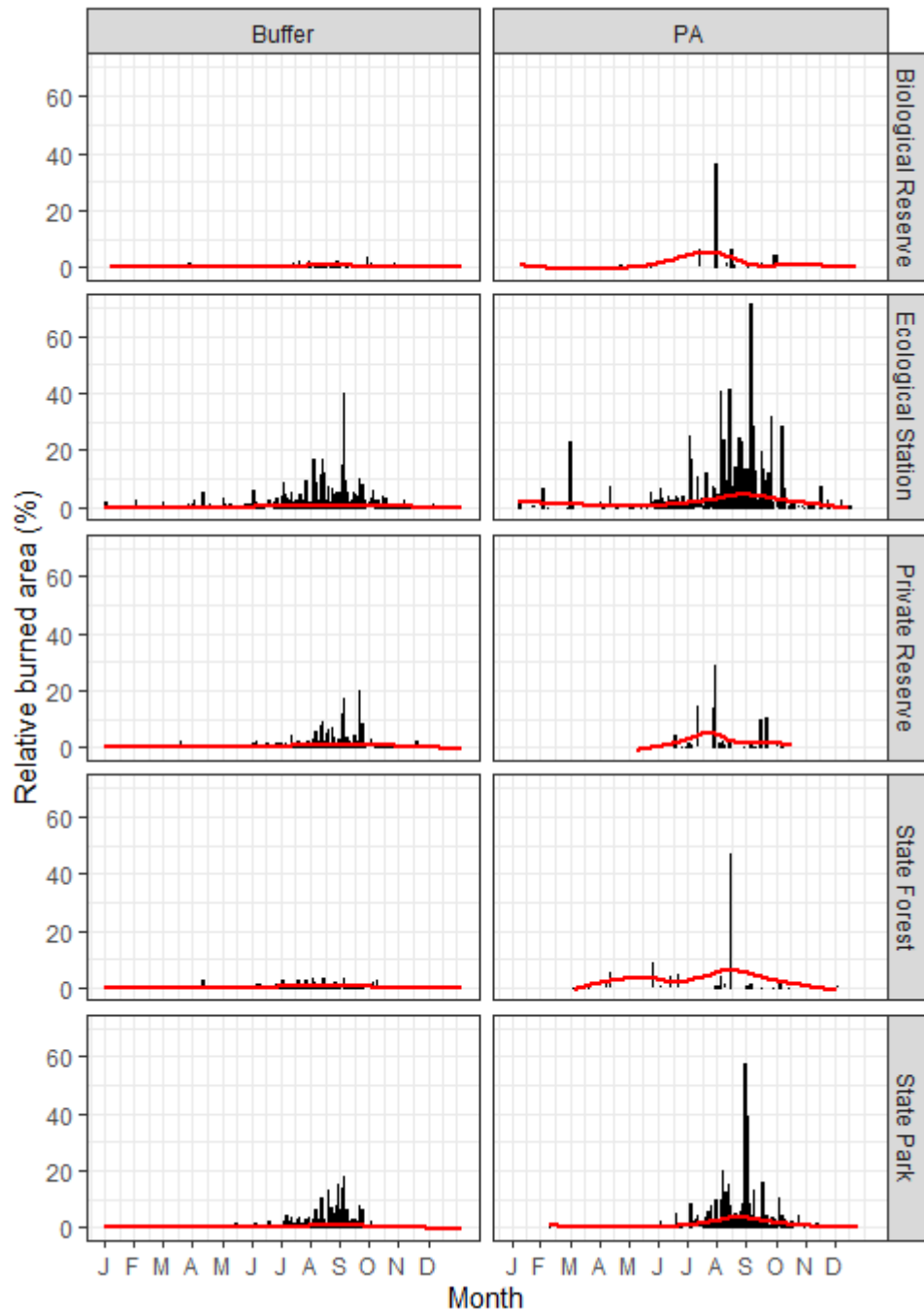
*Supplementary***Supplementary Table S1.** MapBiomass collection 5 pixel value legend.

<b>MapBiomass LCLUC class</b>	<b>Pixel value</b>
<b>Forest</b>	
Forest formation	3
Savanna formation	4
Mangrove	5
Forest plantation	9
<b>Natural non-forest formations</b>	
Wetland	11
Grassland formation	12
Salt flat	32
Rocky outcrop	29
Other non-forest formations	13
<b>Farming</b>	
Pasture	15
Agriculture	18
Temporary crop	19
Soybean	39
Sugar cane	20
Perennial crop	36
Mosaic of agriculture and pasture	21
<b>Non vegetated area</b>	
Beach and dune	23
Urban infrastructure	24
Mining	30
Other non-vegetated areas	25
<b>Water</b>	
River, lake and Ocean	33
Aquaculture	31
Non observed	27

**Supplementary Figure S2.** Principal Component Analysis (PCA) considering the proportion of predominantly LCLUC by protected area for the year of 1985 (72% of explanation). Black arrows represent the loadings of each variable; Red labels shows the name of each variable (class of land cover); Colored labels point each one of the PAs where colors represent the SNUC category.



**Supplementary Figure S3.** Distribution of relative burned area (%) over months considering accumulated values for the interval between 1985 and 2018. Right boxes show the results for the buffer zones while right boxes show the results for the protected areas (PA). Black bars represent the observed value of accumulated burned area for each julian day (1-365). Red curve represents the trends estimated by multiple non-parametric LOESS regressions (locally estimated scatterplot smoothing).



## CONSIDERAÇÕES FINAIS

Em termos gerais, resumimos os principais resultados e contribuições deste trabalho da seguinte forma:

1. As bibliotecas em código aberto em linguagem R, o algoritmo Random Forest e etapas de pós-processamento foram suficientes para gerar uma metodologia automatizada e reproduzível para mapeamento de áreas queimadas no Cerrado. O código fonte do projeto pode ser acessado através do link [https://github.com/musx/FireGIS\\_SP](https://github.com/musx/FireGIS_SP)
2. O produto de áreas queimadas gerado conseguiu atingir qualidade satisfatória para sua posterior aplicação em análises ambientais e ecológicas (79% de acurácia, 16% de omissão e 9% de comissão).
3. Apesar de satisfatório, o produto de áreas queimadas mostrou-se limitado para o correto mapeamento de cicatrizes de queimadas em áreas úmidas, havendo alta comissão nestas áreas.
4. O desenvolvimento da plataforma “FireGIS SP” para análise interativa dos resultados irá contribuir na disseminação e divulgação dos resultados para comunidade científica, gestores e demais usuários. A plataforma pode ser acessada através do link [https://bit.ly/FireGIS\\_SP](https://bit.ly/FireGIS_SP)
5. A combinação entre dados de área queimada e dados de uso e cobertura do solo mostrou-se extremamente eficiente na compreensão dos regimes de queimas do Cerrado paulista.
6. O regime de queimas do Cerrado paulista é essencialmente antrópico, respondendo diretamente pelos padrões de uso e cobertura do solo nas últimas três décadas. Isto é, transições no uso da terra de pastagem para cana-de-açúcar, incentivadas principalmente por políticas públicas de crédito e subsídios ao setor sucroalcooleiro, explicam as maiores variações regionais no regime de queimas.

7. As zonas de amortecimento das Unidades de Conservação repetem o padrão geral de queimas identificado para o Cerrado paulista.
8. Considerando as Unidades de Conservação (UCs), aquelas cobertas principalmente por florestas e cerradão pouco queimaram ou nunca queimaram ao longo da série (1985-2018). Por outro lado, UCs campestres sofreram queimas de grandes proporções em intervalos de 7-8 anos e UCs com dominância de gramíneas africanas apresentaram alta frequência de fogo, potencialmente favorecendo espécies invasoras em detrimento das espécies nativas.
9. Uma vez que as UCs campestres queimam periodicamente por motivo acidental ou criminoso, faz-se necessário o delineamento de novas estratégias de manejo que viabilizem a prescrição de queimas controladas pelos gestores destas UCs como forma de promover a fragmentação do material combustível e a manutenção da biodiversidade.
10. O fogo foi muito frequente em UCs conflituosas e dominadas por gramíneas africanas. Nesse contexto, faz-se necessário a integração da comunidade local em um modelo de gestão participativa como forma de alterar a percepção de abandono, fomentar a educação ambiental e buscar alternativas viáveis e realistas de restauração para as áreas degradadas.