

# A tool for prioritizing deforestation hotspots in the Brazilian Amazon

Alber Sanchez<sup>1</sup>, Guilherme Mataveli<sup>1</sup>, Gabriel de Oliveira<sup>2</sup>,  
Michel E. D. Chaves<sup>1,5</sup>, Ricardo Dalagnol<sup>3</sup>, Fabien H. Wagner<sup>3</sup>,  
Celso H. L. Silva Junior<sup>4</sup>, Luiz E. O. C. Aragao<sup>1</sup>

<sup>1</sup>National Institute for Space Research, Brazil; <sup>2</sup>University of South Alabama, USA; <sup>3</sup>University of California, USA; <sup>4</sup>State University of Maranhão; <sup>5</sup>São Paulo State University in Tupã, Brazil.

alber.ipia@inpe.br



## Introduction

Despite deforestation reduction promises by different governments, no administration has achieved zero illegal deforestation [Pereira et al. 2019]. Deforestation policies and their enforcement are subject to government changes, challenging long-term environmental planning. As a result, deforestation in the Brazilian Amazon has had a positive trend since 2012 (See Figure 1).

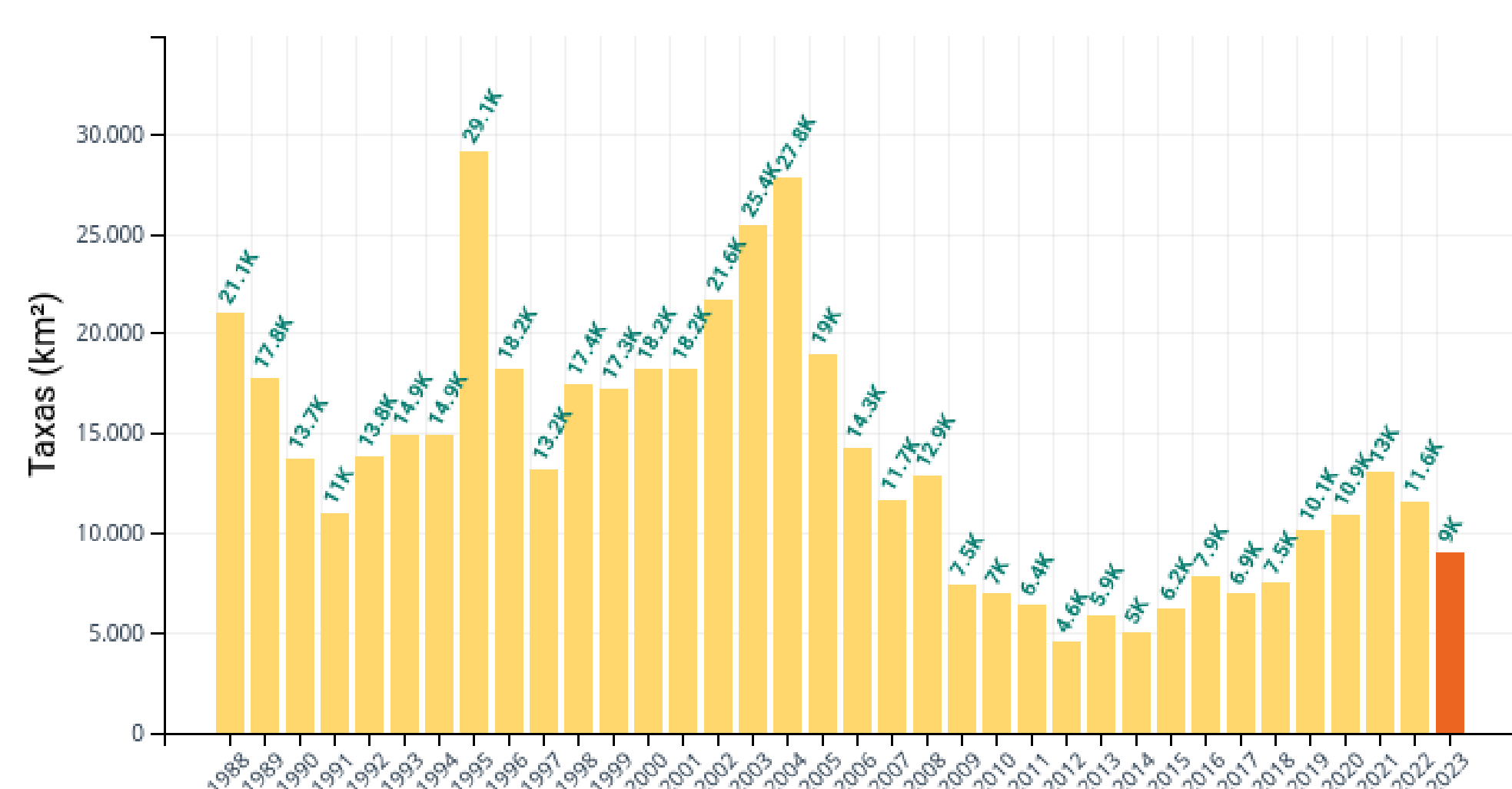


Figure 1: Deforestation in the Brazilian Amazon (km<sup>2</sup>). Source INPE.

We believe that policies regarding deforestation should be subject to public scrutiny and based on scientific principles. For this reason, we proposed an index for prioritizing areas in the Brazilian Amazon for law enforcement actions (check “*Science-based planning can support law enforcement actions to curb deforestation in the Brazilian Amazon*” [Mataveli et al. 2022]). Our index is based on variables aggregated into a regular grid of 25x25 km, 2019 to 2023 using the Random Forest algorithm [Mataveli et al. 2023].

To ensure the openness and transparency of our proposal, we prepared a data and software bundle using the R language (hereby called package) that allows the public and other research teams to reproduce our methods and findings. This package is what we show here.

## Scientific reproducibility with R

According to the Association for Computing Machinery (ACM) badging system, there are three types of reproducibility [Association for Computing Machinery 2022]:

- Repeatability. Research teams can reliably repeat their computations.
- Reproducibility. Independent research teams can obtain the same results using the authors’ software.
- Replicability. Independent research teams obtain the same results using their own artifacts.

R is a programming (scripting) language for statistical computing and graphics [R Core Team 2022]. R is extensible through packages, which enable it to load and run code, data, demos, examples, documentation, tests, and consistency checks [Wickham 2015], making it well suited for distributing scientific results [Pebesma et al. 2012]. Since the inception of our index and paper, R packages have allowed us to achieve both repeatability and reproducibility. The former was because the code and data required to run our experiment were packaged as a single unit, and the latter was because by making our package available online (see Links below), we allowed reproducibility. Because replicability depends on independent research groups, their own data, and their own software, it cannot be achieved.

## Results

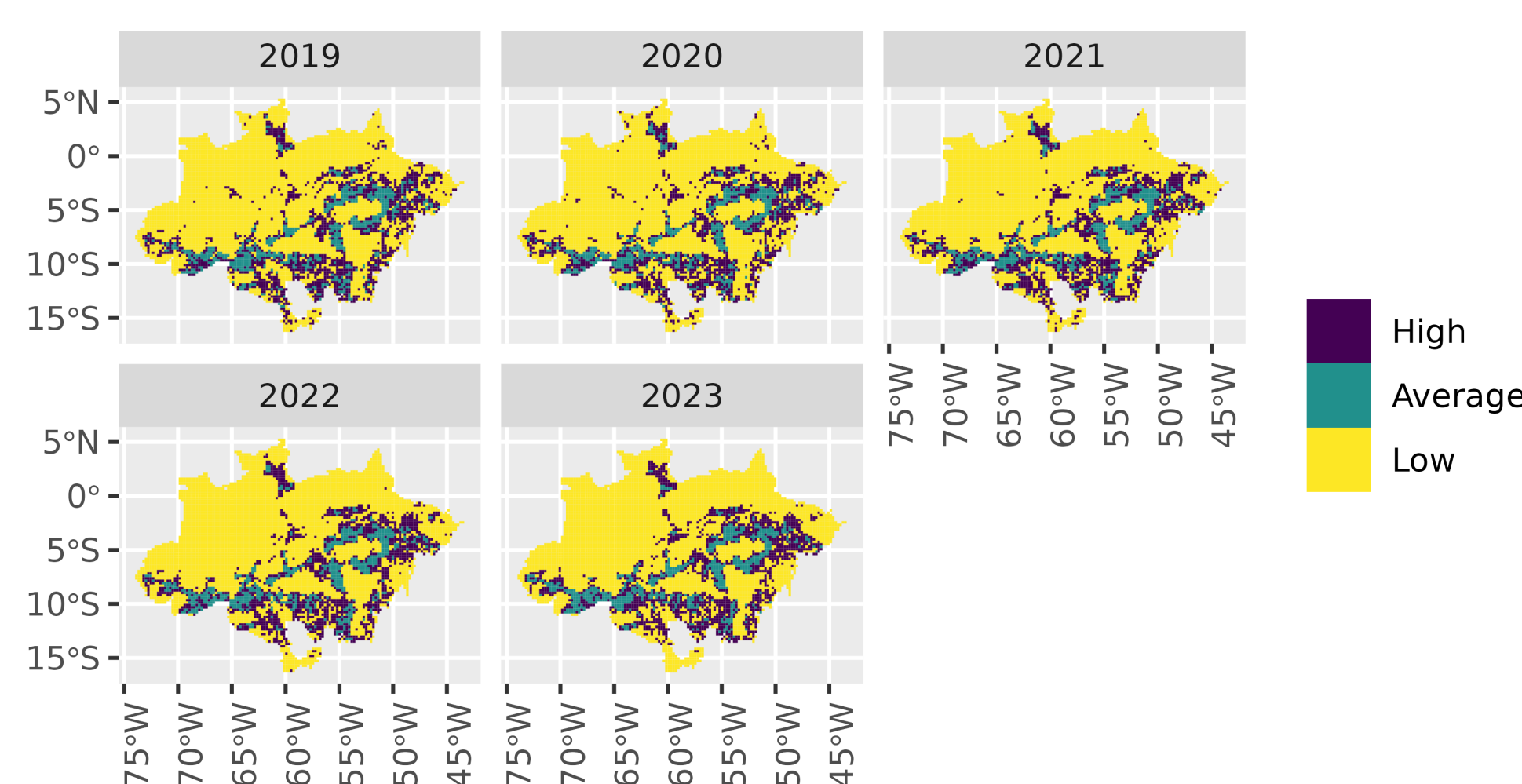


Figure 2: Prioritization index as stored in our package.

## Package description

Our package bundles both the code and the data required to prioritize deforestation areas for 2022 and 2023, allowing users to reproduce the results presented in our paper [Mataveli et al. 2022] and its update [Mataveli et al. 2023] (see Figure 2).

Our package is available online (see Links below).

The code in this package fits the model presented in our paper (*fit\_model()*) and estimates its accuracy (*estimate\_accuracy()*), which is achieved by adjusting 100 models’ data and then cross-validating them. An additional function (*results\_to\_shp()*) applies thresholds to the results to turn them into categories (e.g. low, average, and high) and exports them to a vector file compatible with GIS software. These functions take only one parameter, the output directory (*out\_dir*).

The results are the cross-validation models along with their metrics and error estimates, and the final model and its estimations. These results are stored in the R data format, CSV files, and GeoPackage files.

We ran our code using R 4.3.1 and GNU/Linux Ubuntu 20.04.6 (Kernel 5.15.90.1) LTS on top of Windows 10 Subsystem for Linux 1.2.5.0. We used 16 of the 32 available cores in an Intel Xeon E5-2640 v3 2.593 GHz processor with 32 GB of memory. Given these software and hardware, *fit\_model* took 13 minutes to run (user 9535.76, system 129.55, elapsed 752.90), *results\_to\_shp* took 14 hours (user 640670.17, system 10554.66, elapsed 50190.13), and *results\_to\_shp* took a second (user 0.91, system 0.02, elapsed 0.972).

Regarding data, our package includes a grid of 25 km<sup>2</sup> resolution, the pre-computed index (see Figure 2), and the variables required by our model. This includes areas of deforestation, indigenous or protected lands; distances to waterways, highways, deforestation; and the number of fire spots (see Figure 3).

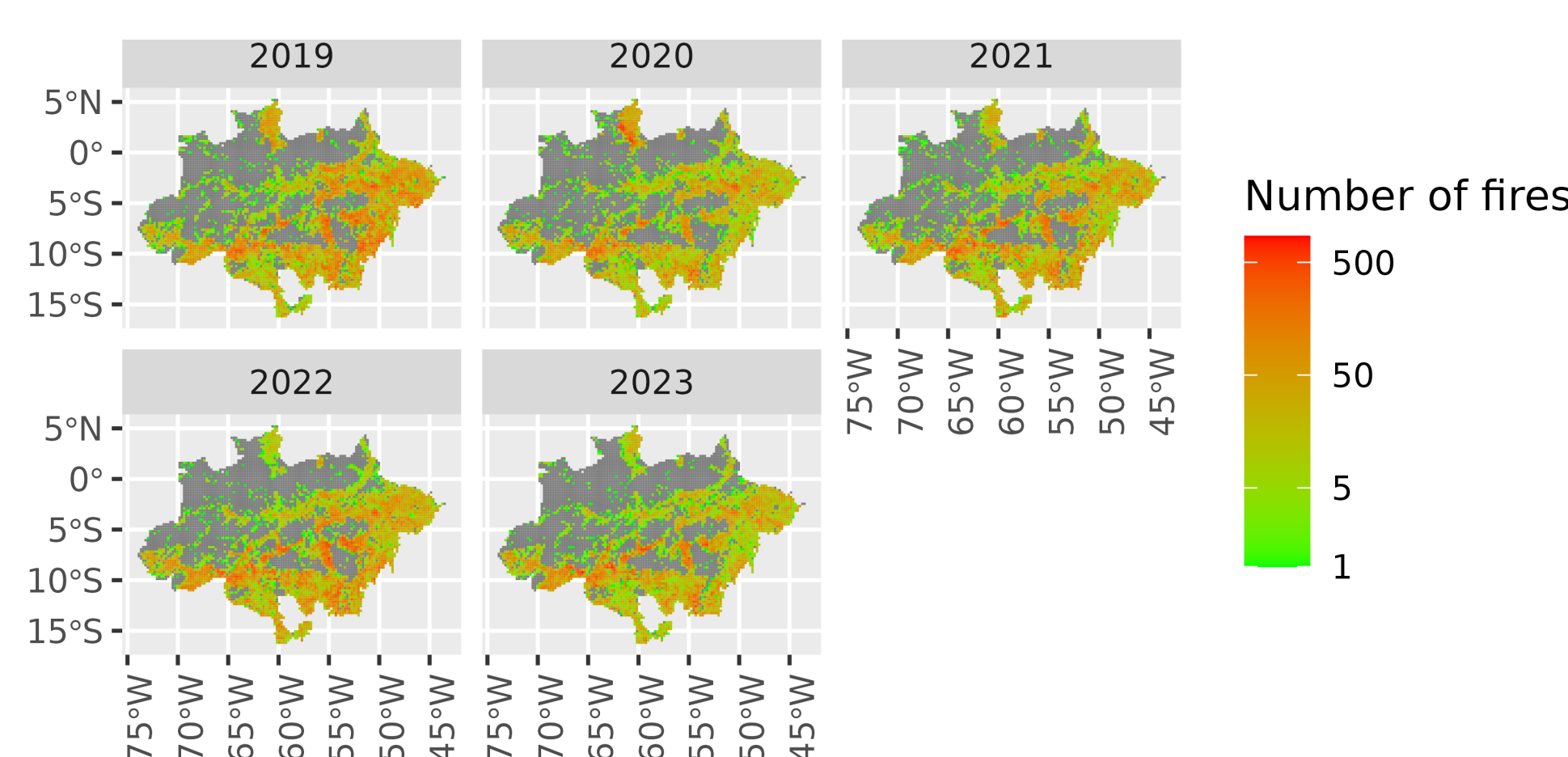


Figure 3: Number of fires. One of the variables for computing out prioritization index.

## Final remarks

We have presented the R package *prioritizeddeforestation-hotspots* (see Links below), which enables users to reproduce the results presented in the paper “*Science-based Planning Can Support Law Enforcement Actions to Curb Deforestation in the Brazilian Amazon*” [Mataveli et al. 2022, Mataveli et al. 2023]. This tool comprises not only the software but also the data used during the writing and analysis stages of the aforementioned paper. In this way, we provide other research teams with the opportunity to review our conclusions and the potential to start extending our research to cover new hypotheses.

## References

- [Association for Computing Machinery 2022] Association for Computing Machinery (2022). Artifact review and badging. <https://www.acm.org/publications/policies/artifact-review-and-badging-current>. Accessed: 2023-09-18.
- [Mataveli et al. 2022] Mataveli, G., de Oliveira, G., Chaves, M. E. D., Dalagnol, R., Wagner, F. H., Ipia, A. H. S., Silva-Junior, C. H. L., and Aragão, L. E. O. C. (2022). Science-based planning can support law enforcement actions to curb deforestation in the Brazilian Amazon. *Conservation Letters*.
- [Mataveli et al. 2023] Mataveli, G. A. V., Oliveira, G. d., Chaves, M. E. D., Silva, R. D. d., Wagner, F. H., Sanchez Ipia, A. H., Silva-Junior, C. H. L., Dutra, D. J., and Aragão, L. E. O. e. C. d. (2023). *Determinação de áreas prioritárias para o combate ao desmatamento na Amazônia em 2023*. Instituto Nacional de Pesquisas Espaciais, São José dos Campos.
- [Pebesma et al. 2012] Pebesma, E., Nüst, D., and Bivand, R. (2012). The R software environment in reproducible geoscientific research. *Eos. Commentarii Societatis philologae Polonorum*, 93(16):163.
- [Pereira et al. 2019] Pereira, E. J. d. A. L., Silveira Ferreira, P. J., De Santana Ribeiro, L. C., Sabadini Carvalho, T., and De Barros Pereira, H. B. (2019). Policy in Brazil (2016–2019) threaten conservation of the Amazon rainforest. *Environmental Science & Policy*, 100:8–12.
- [R Core Team 2022] R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [Wickham 2015] Wickham, H. (2015). *R Packages*. O’Reilly Media, Sebastopol, CA, first edition edition.

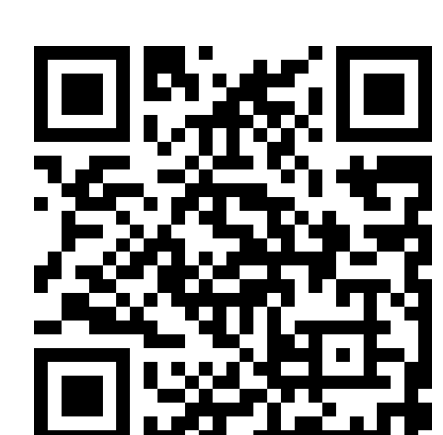
## Links



Email.



Code.



Paper.

- Email: alber.ipia@inpe.br
- Code: <https://github.com/albhasan/prioritizeddeforestationhotspots>
- Paper: <https://doi.org/10.1111/conl.12908>