PRIORITIZATION OF AREAS FOR THE PREVENTION AND COMBATING OF FOREST FIRES IN THE BRAZILIAN AMAZON

Ignácio Martins Pinho [1]*,Alber Hamersson Sánchez Ipia* [2]*, Guilherme Mataveli* [3] *e Liana Oighenstein Anderson* [4]

[1] INPE, martins.pinhos@gmail.com, [2] INPE, alber.ipia@inpe.br, [3] INPE, guilherme.mataveli@inpe.br e [4] Cemaden, lianaanderson@gmail.com

RESUMO

Os incêndios florestais representam um grande desafio para a conservação da Amazônia. O fogo na Amazônia é de origem humana e causa diversos impactos negativos em escalas locais, regionais e globais. Neste trabalho visamos desenvolver um método para priorização de áreas com previsão de ocorrência de incêndios florestais. A priorização foi feita a partir do desenvolvimento de um modelo de regressão *Random Forest* para previsão de ocorrência de incêndios florestais. Para o período de análise, a área determinada como prioritária continha 79,9% do total de área de floresta queimada na Amazônia brasileira. Os resultados ressaltam o potencial do uso de dados de sensoriamento remoto para gestão territorial da Amazônia.

Palavras-chave – Amazônia, incêndios florestais, *Random Forest*, modelagem preditiva, gestão territorial.

ABSTRACT

Forest fires are a major challenge for the conservation of the Amazon. Forest fires are human related and cause negative impacts on local, regional and global scales. We aimed to develop an area priorization method with the prediction of forest fire occurrence. Prioritization was based on the development of a Random Forest regression model for predicting the occurrence of forest fires. For the period of analysis, the area determined as a priority contained 79.9% of the total area of burned forests in the Brazilian Amazon. The results highlight the potential of using remote sensing data for territorial management in the Amazon.

Key words – *Amazonia, forest fire, Random Forest, predictive modeling, territorial management.*

1. INTRODUCTION

The Amazon is the largest tropical rainforest in the world, housing the greatest terrestrial biodiversity on the planet, storing an amount of carbon equivalent to 15 years of global emissions, and playing an important role in climate regulation and ecosystem service provision [1]. However, the expansion of anthropogenic activities in the forest and climate change resulting from global warming threaten its functioning. In this context, forest fires represent a major challenge for the conservation of the biome [2]. Forest fires cause various impacts on forest dynamics, such as plant and animal mortality, reduction of biomass, and canopy cover. Carbon emissions also represent a global-level impact. Fires further cause social and economic impacts, including direct human mortality, increased respiratory diseases, water, food, and energy insecurity, and the loss of agricultural crops, infrastructure, carbon stocks, and ecosystem services.

In the Amazon, the ignition sources that trigger forest fires are mostly of human origin. Fire is used for three main purposes: (1) conversion of forests into agricultural areas; (2) maintenance of pasture areas to prevent the growth of secondary vegetation; and (3) as a tool to challenge public policies and territorial management [3]. In addition to human ignitions, there are other factors that amplify fires: climate and landscape [4]. High temperatures and lack of precipitation reduce vegetation moisture, creating a scenario conducive to fire occurrence, as the vegetation fuel becomes dry. In a global warming scenario with higher temperatures and more frequent and prolonged droughts, the Amazon rainforest becomes more flammable and, therefore, more vulnerable to fire. Furthermore, the structure of the landscape influences the occurrence of fires, as forest edges, besides being closer to anthropogenic ignition sources, have a drier and warmer microclimate, more wind, and less biomass. Finally, forest degradation caused by the fires themselves and selective logging increases the amount of dead biomass and reduces canopy cover, increasing solar incidence within the forest, making it hotter and drier.

To reduce the occurrence of forest fires in the Amazon, it is necessary for public prevention and combat policies to be scientifically grounded in the spatial and temporal patterns, processes, causes, and consequences of forest fires, which are specific to the Amazon biome. The Policy for Action on the Prevention and Control of Deforestation in the Legal Amazon (PPCDAm) was implemented in 2023, and the National Policy for Integrated Fire Management in 2024. Although the prevention and combat of forest fires are included within these policies, little has been established regarding the direction of actions and the definition of priority areas, making territorial management less efficient. The objective of this work is to develop a method that allows for the delimitation of priority areas for structuring fire response actions in the Amazon on a monthly basis. For this purpose, a model was developed to predict burned forest areas based on vectors of forest fire occurrence in the Amazon.

2. METHODS

The development of this study was divided into two stages. The first stage involved the collection and processing of remote sensing data products to estimate variables representing the vectors of forest fire occurrence in the Amazon. For the study, the area of interest was divided into a grid where each cell measures 0.5° in width, and the analysis period was set between November 1 and 30, 2023.

This timeframe was chosen to align the weather forecast data with the other datasets. The second stage included the development of a Random Forest regression model to predict the burned forest area during the analysis period and subsequently delineate priority areas. The prioritization model has two key features: (1) the prediction is sub-seasonal (one month period), allowing sufficient time to direct response actions; (2) the area classified as a priority encompasses a significant portion of forest fires but is not so extensive as to hinder response actions.

2.1. Study Area

The study area encompasses the entire Brazilian Amazon biome, covering over 4 million km², which represents 60% of the entire Amazon rainforest. The biome has an equatorial climate and is predominantly composed of forest formations. However, due to its vast spatial dimensions, there is considerable heterogeneity in climate, structure, and forest composition.

2.2. Data e Variables Used

Data from various remote sensing products were used to estimate variables representing forest fire occurrence vectors in the Amazon. Values were estimated for each grid cell. To define forested versus non-forested areas, the 2023 forest cover mask from PRODES was utilized. The variables used in the model are described below:

- **ba**: Burned forest area per cell during the analysis period (November 1 to 30, 2023). **DETER**.
- **ba_pm**: Burned forest area per cell in the 30 days prior to the analysis period. **DETER**.
- **ba_py**: Burned forest area per cell during the period between November 1 and 30, 2022. **DETER**.
- **da_pm**: Deforestation alerts area per cell in the 30 days prior to the analysis period. **DETER**.
- **da_py**: Deforestation alerts area per cell during the period between November 1 and 30, 2022. **DETER**.
- **af_pm**: Number of active fire detections in non-forested areas per cell in the 30 days prior to the analysis period. **FIRMS/VIIRS**.
- **af_py**: Number of active fire detections in non-forested areas per cell during the period between November 1 and 30, 2022. **FIRMS/VIIRS**.
- **ed, np, mpa, fc, hydro**: Forest edge density, number of forest fragments, average area of forest fragments, percentage of forest cover, and percentage of water body cover per cell. PRODES. **PRODES**.
- **pc**: Percentage of pasture cover per cell. **MapBiomas.**
- **pa, il, pp**: Percentage of area occupied by conservation units, indigenous territories, and registered private properties in the Cadastro Ambiental Rural (CAR) for each cell. **Atlas Agropecuário - Imaflora**.
- **biomass**: Average biomass of forested areas within each cell. **CCI - Biomass**.
- **road_d**: Road density for each cell. **OpenStreetMap.**
- **prec_f, temp_f**: Average daily precipitation forecast and average temperature forecast per cell during the analysis period. **BAM-1.2 - CPTEC/INPE**.
- **ahd**: Accumulated water deficit up to the analysis period [5]. **CHIRPS**.
- **prec_a, temp_a**: Occurrence or not of precipitation and surface temperature anomaly in the previous month for each cell [6]. **CHIRPS e MOD21C3**.

2.3. Modeling and Delimitation of Priority Areas

With the described data, regression models using the Random Forest algorithm were created to predict the area of burned forest per cell during the analysis period. The non-parametric machine learning algorithm Random Forest uses a set of decision trees to make predictions based on a dataset. This method was used because the algorithm handles different types of data well, is easy to implement and optimize, and performs well against overfitting. The implementation was done in R using the Ranger package.

The Pearson correlation coefficient was calculated between each independent variable and the dependent variable (ba). The variables were then ranked according to the absolute value of their respective correlation coefficients. A series of models was created, starting with the highest-ranked variable, and adding one more variable at each step until a model using all variables was reached. During this process, the model's hyperparameters were optimized through grid search. At each step, the model's accuracy (R^2) was calculated. Additionally, each model was used to make predictions of burned forest area. From these predictions, each cell in the grid was classified into priority classes for fire prevention and suppression. The classification was done using percentiles. The high, medium, and low priority classes were defined as greater than 90, between 90 and 70, and less than 70 percentiles, respectively. The following calculation was then performed to evaluate the classification quality of each model:

% of burned forest area per class = Burned forest area per class / Total burned forest area

A priority classification was also performed using DETER data on burned forest area. Since this classification used actual data, it represents the best classification to be achieved. The classifications made from the models' predictions were compared with this ideal classification to assess how efficient the predictions were.

3. RESULTS

In November 2023, the area of burned forest in the Brazilian Amazon biome totaled 2123.68 km². During this period, forest fires occurred in 197 of the 471 cells that make up the study area. There was a concentration of fires in the central and eastern regions of the Brazilian Amazon, although fires also occurred in other regions. Figure 1 shows the spatial distribution of the burned forest area during this period.

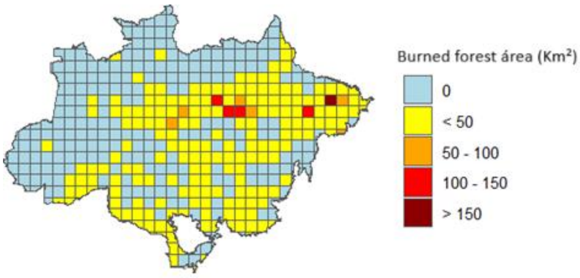

Figure 1: Burned forest area (km²) for each grid cell between 1 and 30 of November of 2023, acording to DETER data.

Figure 2 presents the ranking of variables based on the absolute value of the Pearson correlation coefficient. Regression models were created following the order of this ranking.

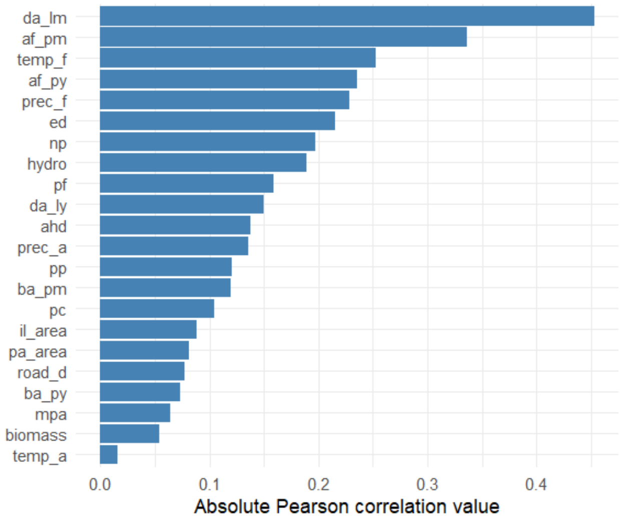

Figure 2: Ranking of variables based on the absolute value of the correlation coefficients between the variables and the burned forest area. The variables prev_prec, CF, DHA, anom_prec, TI, UC, AMF and BM showed a negative coefficient value.

The performance of the models, both in terms of the R^2 value and the percentage of burned forest area in cells classified as high priority, is shown in Figure 3. The maximum R^2 value (0.209) was achieved using only the first three variables from the ranking (da_pm, af_pm and temp_f). However, this model had the third worst performance regarding the percentage of burned area in cells classified as high priority (73.6%). For this metric, the model with six variables (da_pm, af_pm, temp_f, af_py, prec_f and ed) showed the third best performance (79.9%). Adding more variables did not significantly increase this percentage; only models with 19 and 22 variables showed a slightly higher percentage (80.1% and 80.2%, respectively).

The model with only six variables was considered the best for two reasons: (1) the difference between the percentages is very small; (2) with a significantly smaller number of variables, the model is much simpler to explain, making it more parsimonious.

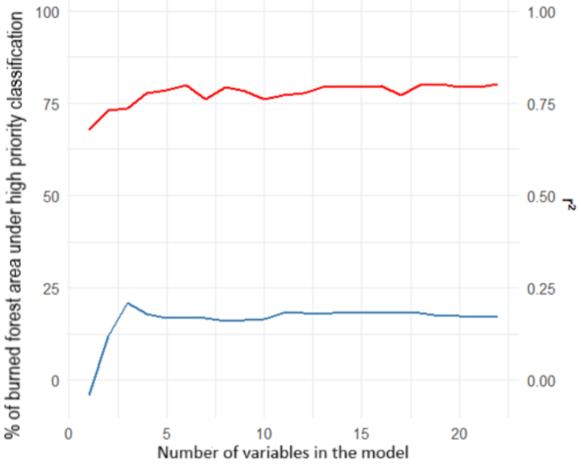

Figure 3: Model performance according to the number of variables. In red: the percentage of burned forest area that occurred in cells classified with the highest priority relative to the total burned forest area. In blue: the R^2 value for each model.

The priority area classification based on the selected model is shown in Figure 4, along with the reference classification based on the actual burned forest area data. The model correctly classified 32 high-priority cells and made mistakes in 15. Both classifications concentrated high-priority areas in the central and eastern regions of the Brazilian Amazon, where there was greater occurrence of forest fires. The model's classification did not prioritize areas in the southern part of the biome, where there was also a significant amount of fires. Table 1 also presents a comparison between the classification made by the model and the classification made using burned forest data. It can be observed that the burned forest area within the high-priority class determined by the model represents 89.6% of the burned forest area within the high-priority class determined by the reference classification.

4. DISCUSSION

From the comparison between the models, it becomes clear that the R^2 value provides little information for evaluating the model's quality. Since the goal is not to make an exact

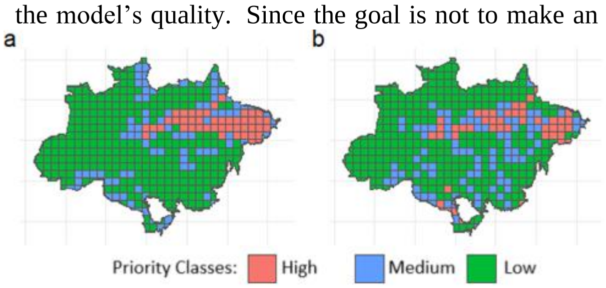

Figure 4: a: Priority classification created from the predictions of the best model; b: Reference priority classification created from the actual burned forest area data from DETER.

Classification based on model predictions			Classification based burned forest area data from DETER		
Priority class	Burned forest area (Km²)	Percentage of Burned forest area	Priority class	Burned forest area (Km²)	Percentage of Burned forest area
High	1696,23	79,87	High	1893,27	89,15
Medium	319,85	15,06	Medium	211,82	9,97
Low	107,60	5,07	Low	18,59	0,88

Table 1: Comparison between the classifications made by the model and based on DETER burned forest area data.

prediction of the burned forest area, but rather to create a priority classification from these predictions, assessing the classifications themselves is a better approach to determining the model's quality. Therefore, the model composed of 6 variables was selected as the best. These variables correspond to the vectors of forest fire occurrences in the Amazon. The ignition sources are represented by variables that quantify deforestation and the use of fire in non-forested areas. Weather conditions are represented by the weather forecast variables. Lastly, forest structure is represented by the edge density variable. The simplicity, conceptual consistency, and good performance of the classification derived from this model were the reasons for its selection as the most suitable model.

The classification made from the selected model showed satisfactory results. With 79.9% of the burned forest area occurring within the high-priority class, the goal of delimiting small priority areas that concentrate a significant portion of fire occurrences was achieved. Thus, this method of prioritizing areas for fire suppression and prevention can help direct effective actions to reduce and mitigate the impacts of fire in the Brazilian Amazon.

Although it shows good potential for delimiting priority areas for fire suppression and prevention, the method still has limitations and needs to be optimized. Initially, there are limitations regarding the data itself. Remote sensing products tend to underestimate fire occurrence in dense forests because tree canopies make it difficult to detect fire in the undergrowth. Uncertainties and limitations also exist in all other data used. Therefore, constructing and comparing models that use other data as the dependent variable should be done. Lastly, the model's classification needs to be tested using datasets from other times of the year. For this model to become operational, its classification must maintain a consistent performance over time. This could be the greatest challenge, as the relative importance of each variable should not remain constant over time, since the vectors of forest fire occurrences in the Amazon vary over time and space.

5. CONCLUSIONS

The threat of fire, becoming increasingly frequent and intense, requires Brazil to take more effective actions to reduce the occurrence of forest fires in the Amazon rainforest. This is necessary for the maintenance of the biome, its ecological interactions, ecosystem services, and for Brazil to meet its carbon emission reduction goals. Given the vast dimensions of the Amazon, the use of remote sensing products and machine learning presents great potential to aid in the territorial management planning of the Amazon. In this case, the model created can help direct actions for fire prevention and suppression.

6. REFERENCES

[1] IPCC. Climate change 2021: The physical science basis (eds masson-delmotte, v. et al.). *https://www.ipcc.ch/report/ar6/wg1/FullReport, Cambridge Univ. Press*, 2021.

[2] David M. Lapola et. al. The drivers and impacts of amazon forest degradation. *Science*, 379, 2023.

[3] Jos Barlow et. al. Clarifying amazonia's burning crisis. *Global Change Biology*, 26, 2020.

[4] Celso H.L. Silva-Junior et. al. Forest fragmentation and fires in the eastern brazilian amazon–maranhão state, brazil. *Fire*, 5, 2022.

[5] Luiz Eduardo O.C. Aragão et. al. Spatial patterns and fire response of recent amazonian droughts. *Geophysical Research Letters*, 34, 2007.

[6] Liana O. Anderson et. al. Remote sensing detection of droughts in amazonian forest canopies. *New Phytologist*, 187, 2010.