Leveraging Python for data modeling, visualization and remote sensing of fire hotspots in the Amazon: a case study of Santana do Araguaia, Brazil

Jackson Vinícius Mendes da Silveira

Cyberspace Institute, Federal Rural University of Amazonia, Belém, Brazil – jackson.silveira@discente.ufra.edu.br

Keywords: Amazon region, Python, data modeling, fire risk, fire hotspots.

Abstract

Conservation of the Amazon rainforest is essential because it plays an important role in global biodiversity and climate stability. In this work, python was used in data modeling, visualization and remote sensing for analysis of fire hotspots generated from the region where Santana do Araguaia is located. With the use of shapefiles and csv data and different Python libraries, like pandas, geopandas for geographical functions and libraries with graphics packages such as matplotlib, seaborn and rasterio. Methodology encompasses a series of analytical techniques to explore the relationships between environmental variables and fire risk. These techniques include generating linear regression models to study specific correlations, creating choropleth maps to visualize spatial patterns, mapping fire hotspots to identify high-risk areas and using 2d regression graphs for detailed analysis. In addition, time series analyses are conducted using static visualizations to track changes in fire risk over time, offering a comprehensive approach to understanding fire dynamics. Remote sensing techniques are used to produce normalized difference vegetation index maps and detect changes in vegetation and land cover. This research demonstrates how Python tools can reinforce fire risk analysis and support forest management blueprint, providing relevant information for conservation efforts.

1. Introduction

The Amazon is the largest continuous tropical rainforest on the planet, it serves as a home for an enormous biodiversity of plants and animals, is important for carbon sequestration and climate regulation (Rosa et al., 2013; Andela et al., 2022; Silva et al., 2005). It occupies roughly 60% of Brazil's territory and also plays a fundamental economic role for the region's inhabitants, marked by low socio-economic development, high agriculture, extractivism, and livestock activity (Silva Junior et al., 2022; Gomes et al., 2020; Rappaport et al., 2018). However, this region faces increasing threats from climate changes and human-driving activities, particularly fire occurrences (Qin et al., 2019). Forest fires in the region have serious consequences, including the loss of biodiversity, the release of excess carbon dioxide and the disruption of indigenous and local communities (Silveira et al, 2022; Andela et al., 2022; Tyukavina et al., 2017). Underlying causes of these fires are complex, usually linked to changes in land use, agricultural expansion and climate variability (Oliveira et al., 2013; Andela et al., 2022). Understanding the spatial and temporal patterns of these fires is key to developing effective fire management strategies and informing policy decisions aimed (Cheng et al., 2024; Mohd Said et al., 2017).

Santana do Araguaia, located in the southeast of Pará state has an area of 11.591,441 km² and its population is estimated to be around 32.000 thousand people with it accounts for populational growth projection by IBGE (Brazilian Institute Geography Statics). The municipality has been part of the Araguaia Integration Zone, composed by fifteen other municipalities (Oliveira et al., 2016). Although the region has had an historic occupation, it is the one with the highest levels of environmental alteration and loss of biodiversity (Becker, 2001). Historically, it has been marked by public and private initiatives for integration, colonization and the creation of road networks that have altered the socio-economic and political structures, as well as the natural landscape of the region (Penteado, 1967; Botelho et. al., 2022).

This research is based on free data provided by the National Institute for Space Research (INPE), Mapbiomas and Landsat satellite imagery. Python's extensive libraries facilitate data processing, geospatial analysis and visualization, enabling an evaluating the potential of these tools in environmental management (Lopes et al., 2021; Stancin and Jovic, 2019). Analysis methodology involves the collection and pre-processing of data sets, including satellite images and geospatial data on land cover. These datasets are utilized to form models which analysis associations among variables similar precipitation, vegetation indices and fire hotspots with their distribution. Visual and spatial analysis, for instance through maps or graphs, is employed to identify patterns of the data which enhance understanding (Zahran et al. 2020). 2018, 2021 and 2023 will be used to demonstrate the techniques and the changes that have taken place in this region.

Its conclusions and approach are also to be considered one amongst many forms of data analysis, contributing towards the broader endeavor in maintaining Amazon region through science-based governance and resource management.

2. Data

2.1 Data collection

Different types of datasets were collected to be modeled using Python libraries to analyze fire hotspots and environmental conditions in Santana do Araguaia, Brazil.

2.1.1 Shapefiles: Were used to represent vegetation cover. A land use mask from Mapbiomas was applied to the shapefile of Santana do Araguaia to classify natural forest.

2.1.2 CSV Files: Including details such as time, location, latitude and longitude were collected in CSV format from the National Institute for Space Research (INPE).

2.1.3 Satellite Images: Were downloaded from the United States Geological Survey (USGS) website.

https://doi.org/10.5194/isprs-archives-XLVIII-3-W3-2024-27-2024

2.2 Preprocessing

2.2.1 Cleaning: Cleaning data enabled efficient selection and filtering of relevant columns and rows. This reduced the need for extensive data cleaning, as only the necessary data was extracted for analysis.

2.2.2 Reprojection: For Santana do Araguaia was reprojected using QGIS to the SIRGAS 2000 coordinate reference system. This reprojection was essential for aligning the shapefile with other spatial datasets, ensuring consistency in geographic analysis.

2.2.3 Image processing: Landsat 8 satellite images were manipulated to compose bands and generate a normalized difference vegetation index (NDVI) image was obtained. Bands composition was done with rasterio library to make the images more applicable for vegetation analysis. The NDVI was calculated using the raster calculator equation (1). This indicator is crucial for monitoring vegetation indices and greenhouse ecosystems (Ma et al., 2023).

$$NDVI = \frac{NIR - RED}{NIR + RED}, \quad (1)$$

where NDVI = normalized difference vegetation index
NIR = reflection in the near-infrared spectrum
RED = reflection in the red spectrum

2.2.4 Mapbiomas mask: In order to generate a shapefile of Santana do Araguaia that contained information on the natural forest cover present in the municipality for later use, the Mapbiomas overlay tool was applied in QGIS.

2.3 Python libraries and tools

2.3.1 Pandas: Is an easy-to-use licensed open-source data analysis library for the Python programming language (Sial et al., 2021; Ramalho, 2015). It is essential for managing and manipulating tabular data. Its powerful data structures and functions enabled efficient cleaning, filtering, and aggregation of data from the CSV files (McKinney, 2022).

2.3.2 Geopandas: It is an open-source project that facilitates the manipulation of geospatial data in Python. It extends the data types used by pandas by allowing geometric spatial operations (Lopes et al., 2021). Can support spatial operations such as merging datasets, performing spatial joins and reprojection.

2.3.3 Matplotlib: Is a graphical library for data visualization in Python, supported by other libraries such as pandas and numpy (Hunter, 2007; Lemenkova, 2020; Lopes et al., 2021). Can creating static visualizations of the data (Sial et al., 2020). Its extensive capabilities allowed for the generation of a wide range of plots, including histograms, scatter plots, and line graphs (Ramalho, 2015).

2.3.4 Rasterio: Is a widely used raster data processing tool to writing and manipulating raster files, such as resampling and compositing bands.

2.3.5 Seaborn: Was built on top of matplotlib, enhanced the ability to generate statistical visualizations (Lopes et al., 2021; Lemenkova, 2020). Has advanced plotting functions such as correlation heatmaps and regression plots for analyzing variable relationships and presenting statistical findings clearly and aesthetically.

2.3.6 Numpy: Is a numerical computation library consisting of multidimensional array objects and a collection of procedures used to process equal types of arrays, is used to build efficient computational models and numerical perspectives (Sial et al., 2020; Lemenkova, 2020).

3. Data modeling

3.1 Regression analysis

A linear regression model is a statistical method used to understand the relationship between a dependent variable and one or more independent variables (Faraway, 2014). It finds the best-fitting linear equation to predict the dependent variable based on the values of the independent variables. The basic form of a simple linear regression equation is:

$$Y = \beta 0 + \beta 1X + \epsilon, \quad (2)$$

where Y = dependent variable
X = independent variable
β0 = intercept
β1 = slope
ϵ = error term

By using Python, a linear regression model can be defined for the data for the years 2018 and 2023, we can understand how changes in rainfall and days without rain influence the risk of fire. Importing libraries `LinearRegression` from `sklearn.linear_model` is used to create and train the linear regression model. The `pandas` is employed for data manipulation and loading. Data for the years 2018 and 2023 are loaded from CSV files into DataFrames for analysis. `X_2018` and `X_2023` contain the independent variables (precipitation and rainless days) for 2018 and 2023, respectively. `y_2018` and `y_2023` are the dependent variables (fire risk) for the same years. Two linear regression models are created, one for each year. The models are trained using the fit method, which adjusts the model parameters to minimize the error between predicted and actual fire risk values.

	2018	2023
Coefficient 1	-0.04627345	-0.00654461
Coefficient 2	0.01071018	0.01497931
Intercept	0.54746460	0.28018222
R^2	0.37641113	0.63368071

Table 1. Values found for linear regression

Regression results (Table 1) provide statistical metrics that describe the relationship between independent variables (precipitation) and the dependent variable (fire risk).

This coefficient measures the amount of change in a dependent variable for each unit change in the independent variable. For example, if the coefficient for precipitation is 0.5, then an increase of one unit in precipitation results in an increase of 0.5 units in fire risk. The higher the coefficients, the stronger the effect of the independent variable on the dependent variable.

Intercept is the value that the dependent variable would have when the independent variable is equal to zero. It sets a base measure for changes.

R^2 value describes how well the independent variables explain the variation in the dependent variable. An R^2 close to 1 means a good fit; the independent variables describe most of the variation in fire risk.

For both years, the coefficients show the impact of each predictor on fire risk. By comparing these coefficients and R^2 values across years, we can assess if and how the relationships between precipitation, rainless days, and fire risk have changed over time.

3.2 2D regression plot

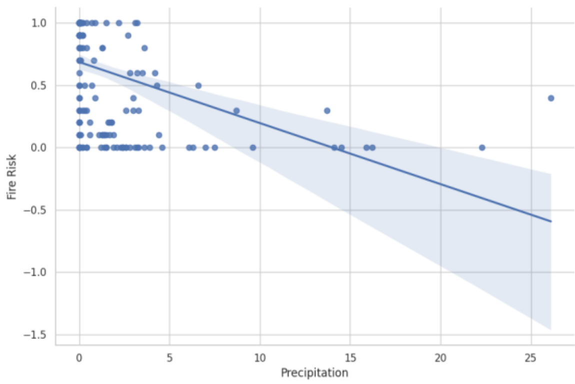

Figure 1. 2018 regression: Precipitation vs. fire risk.

(Fig. 1) was used to visualize the relationships between two key variables in a two-dimensional space. For instance, plotting `Precipitation` against `Fire Risk` helps us understand how changes in precipitation might impact the fire risk. These plots reveal the trend or relationship between the variables. For example, if the plot shows a positive slope, it indicates that higher precipitation is associated with increased fire risk. Conversely, a negative slope would suggest that higher precipitation is linked to reduced fire risk. Slope, or steepness, and direction of the regression line in the plot are informative vis-à-vis the strength and nature of the relationships between variables. The steeper the slope, the stronger the relationship; the flatter the slope, the weaker it is.

4. Visualization

4.1 Choropleth maps

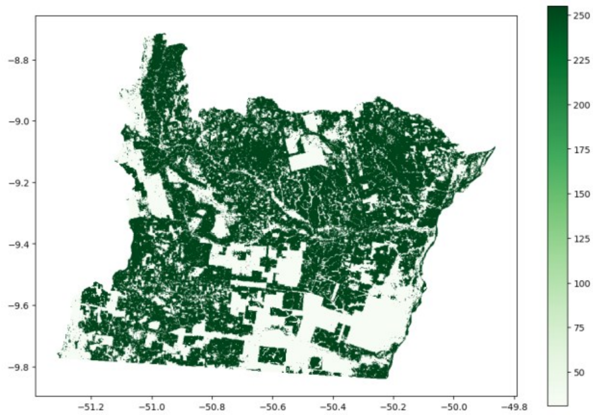

Figure 2. Natural forest cover in Santana do Araguaia in 2021.

(Fig 2) represents the forest cover data of the Santana do Araguaia study area with data from Mapbiomas Collection. The map helps in visualizing the extent and distribution of different forest covers.

The shapefile containing the forest cover data `sta_flo_nat.shp` is loaded using the geopandas library. The information contained in this shapefile includes geographic and attribute data related to forest cover in the study area. The shapefile data is loaded and combined with its existing attributes. The `df.columns` are updated to the expected format, containing the forest cover data along with geometry information. Using `Matplotlib.pyplot`, a figure and axis is created. The plot method of geopandas is applied to the map variable. This visualizes the forest cover data `column` parameter: `flo_nat` uses a column made up of the different levels of forest cover. `cmap = 'Greens'` creates a green gradient for these varying levels of forest coverage.

By analyzing (Fig 2), researchers and conservationists can pinpoint areas of high forest cover that may require protection or areas with reduced cover that could be targeted for restoration efforts. This information is critical for making informed decisions about forest management and conservation strategies.

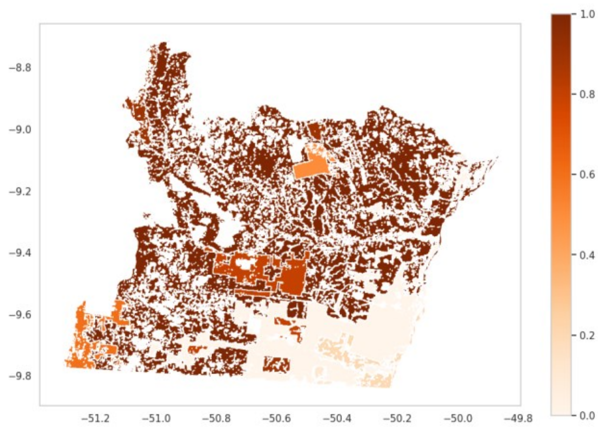

Figure 3. Fire hotspot data superimposed on natural forest.

https://doi.org/10.5194/isprs-archives-XLVIII-3-W3-2024-27-2024

In addition, the forest cover map can be used in conjunction with fire hotspot data to provide a comprehensive view of forest cover and fire risk in Santana do Araguaia (Fig 3).

Forest cover data is loaded from a shapefile (`sta_flo_nat.shp`) using geopandas. This shapefile contains geographic information about different levels of forest cover. Fire hotspot data is loaded from a CSV file (`sta_focos_2021.csv`) using pandas. This file includes longitude, latitude, and fire risk information for various fire incidents. The fire hotspot data is converted into a `GeoDataFrame` using geopandas, with points created from the longitude and latitude columns. This step allows for the spatial representation of fire hotspots. The coordinate reference system (CRS) of the fire hotspot data is set to EPSG:4326 (WGS 84) and then transformed to EPSG:4674 (SIRGAS 2000), aligning it with the CRS of the forest cover shapefile. A spatial join (`gpd.sjoin`) is performed between the forest cover `GeoDataFrame` and the fire hotspots `GeoDataFrame`. This join matches fire hotspots with corresponding forest cover areas based on their spatial intersection. Duplicate entries are removed to ensure each geographic location is represented only once. The `plot` method is used to display this combined data, with fire hotspots represented in shades of orange based on their risk level.

(Fig 3) shows both forest cover and fire hotspots, allowing for the assessment of how fire incidents relate to forest density. Areas with high fire risk are overlaid on the forest cover map, providing a visual representation of the interaction between fire risk and forested areas.

4.2 Spatial analysis

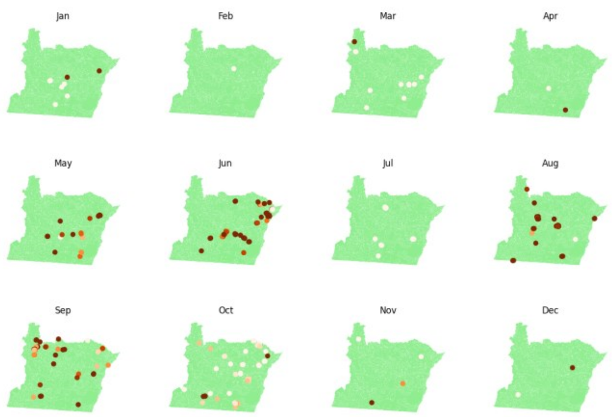

Figure 4. Monthly fire hotspots maps 2018.

(Fig 4) presents a series of small maps that illustrate the fire hotspots for each month of 2018. These maps allow for a visual analysis of how fire risk fluctuates over time and across different regions within the study area.

2018 data is loaded from a CSV file. The `DataHora` column is converted to datetime format to extract the month (`Mes`) for grouping. Fire hotspots data is grouped by month to prepare for visualization. The data is filtered for each specific month. Scatter plots are created for each month. The `color intensity` of points represents fire hotspots, and maps are displayed in a grid format. Maps are arranged side by side in a grid layout to facilitate easy comparison. The plots are shown with the months labeled as abbreviations (e.g., Jan, Feb, Mar) and the intensity of fire visualized.

(Fig 4) provide a visual representation of the fire risk distribution throughout 2018.

4.3 Time series analysis

The data that contained datetime information is parsed to extract month and year when needed as these can be very useful in understanding how different achievements vary seasonally. Finally, it is visualized with seaborn, where a line plot will be drawn for fire risks over every month of different years, represented through different lines. Due to the lack of information for 2023, the corresponding line stops in July.

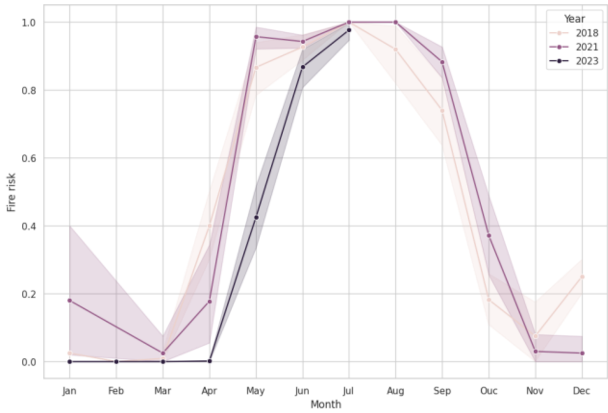

Figure 5. Variation of fire risk over the months of 2018, 2021 and 2023

With pandas library, the code load data from CSV files for the years 2018, 2021 and 2023. These files contain fire hotspot data, including information about fire risk (`RiscoFogo`), the date and time of occurrence (`DataHora`), and other relevant variables. The `DataHora` column is converted to a datetime format to enable the extraction of the month and year, which are stored in new columns `Mes` and `Ano`, respectively. This step is essential for the time series analysis, as it aligns the data along a common temporal axis (months of the year).

Code concatenates data from all three years into a single `DataFrame` called `data_total`. This combined dataset allows for a unified analysis of fire risk trends across different years. The data is then filtered to remove any rows where the fire risk (`RiscoFogo`) is negative, ensuring the analysis focuses only on valid data points.

Using the seaborn library, a line plot is created to visualize the fire risk across different months, with separate lines for each year. The `hue` parameter differentiates the years, allowing for easy comparison. The plot is customized with labels for the x-axis (Months) and y-axis (Fire risk), and the x-axis ticks are labeled with month abbreviations (e.g., Jan, Feb). A legend is added to indicate which line corresponds to which year. Finally, the (Fig 5) is displayed, showing how fire risk varies across months and between years.

5. Remote sensing

5.1 NDVI calculation

A 2018 image from the Landsat 8 satellite was used, which did not cover the entire municipality, but can be used in our example, and can be replaced by an image from any other year. It was used bands 4 (red) and 5 (near-infrared).

Code imports the necessary libraries such as `rasterio` for manipulating raster data, `numpy` for numerical operations, and `matplotlib.pyplot` for visualization. It opens the TIFF files containing the red and near-infrared (NIR) bands of Landsat 8 using `rasterio.open()`. The red and NIR bands are read into arrays. The (1) ensures accurate values, reflecting the difference between NIR and red band intensities relative to their sum.

Metadata from the red band is used to maintain spatial resolution and other properties. The array is saved as a new TIFF file, ensuring correct data type `(float32)` and single-band format.

NDVI data is visualized with `matplotlib.pyplot`, using `imshow()` with a “inferno” color map. This colormap provides a vivid representation of values, where areas with higher vegetation health appear in brighter colors, while lower vegetation or barren areas are shown in darker colors. The color bar accompanying the plot offers a scale for interpreting values.

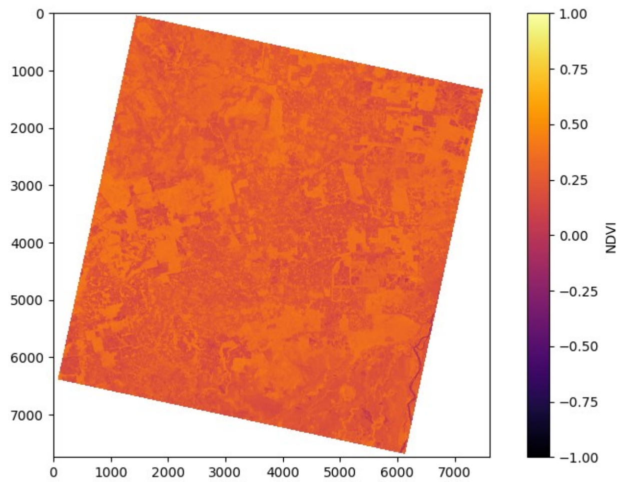

Figure 6. NDVI generated from bands 4 and 5.

5.2 Comparison and change detection

After creating (Fig 6), a section was cut out of the area that is part of the case study, which is the westernmost part of the municipality. The code was used to generate a TIFF file for 2021 as well, so that a comparison process could be carried out in this interval to verify the change that had occurred.

Using `rasterio.open()` for 2018 and 2021, it extracts the NDVI data along with the image's spatial transformation and coordinate reference system (CRS). To ensure that both images are comparable, it checks if the dimensions of the images match. If not, it reprojects the 2021 image to the CRS and resolution of the 2018 image using `calculate_default_transform` and `reproject`. This step aligns the spatial attribute of the images for accurate comparison. By replacing background values (zeros) in both images with `NaN` to avoid distortions in the analysis. Difference between 2021 and 2018 is computed to feature changes in vegetation health over the period. Positive values indicate an increase in vegetation, while negative values suggest a decrease.

Three subplots are created (Fig 7) to display the images for both years, along with the variation between the two. Images are plotted using a green colormap, with background values set to white for better visibility of vegetation. The variation map is visualized with the "oranges" colormap to emphasize the extent and magnitude of changes in vegetation health.

Figure 7. Vegetation indices clipped for the relevant area.

6. Results and discussion

6.1 Statistical analysis

Statistical analysis, including regression models, provides valuable information on the relationships between the main variables affecting fire risk, such as precipitation, rainless days and fire hotspots indices. For example, a negative correlation between rainfall and fire risk suggests that lower rainfall is associated with higher fire danger, a common trend in fire-prone regions.

Regression models further quantify these relationships, providing coefficients that indicate how changes in variables such as rainfall and the number of dry days affect fire risk. By applying these models to data from different years, such as 2018 and 2023, it is possible to compare how these links may evolve over time.

6.2 Fire hotspot distribution

Examining the distribution of fire hotspots provides valuable information on the spatial dynamics of fire incidents (Mohd Said et al., 2017). The maps depicting the distribution of fire outbreaks in Santana do Araguaia reveal the geographical patterns of fire occurrences. Areas of high intensity, where fire incidents are more frequent, are clearly highlighted. This spatial representation not only identifies regions with significant fire activity, but also helps to understand how different types of land cover are affected. For example, areas with dense vegetation may show different fire patterns compared to other regions. By overlaying fire data with land cover types, the analysis facilitates targeted interventions and resource allocation, which are crucial for mitigating fire risks and improving forest management strategies.

6.3 Time series

Time series analysis of fire risk in 2018, 2021 and 2023 provides a temporal perspective of fire activity. Monthly graphs illustrate how fire risk fluctuates throughout the year, showing periods of increased or decreased fire incidents. This analysis helps identify seasonal trends and patterns, such as peak fire seasons or months with exceptionally high or low fire risk. Choropleth maps provide a detailed overview of the intensity of fire hotspots on a yearly basis, offering a clearer understanding of how fire risk varies over time. This temporal insight is essential for developing seasonal fire management plans and predicting future fire risks based on historical data. Uniform with Viana et al. (2024) understanding these trends can inform preventative measures and improve response strategies during periods of high risk.

6.4 Remote sensing

According to Holmgren and Thuresson (1998) remote sensing data, especially from satellite images, plays a key role in assessing vegetation health and detecting changes in land cover. Normalized difference vegetation index products provide a measure of vegetation health by comparing the reflectance of red and near-infrared light. (Fig 7) reveal areas with healthy, stressed or degraded vegetation, offering information on how fire incidents and other factors affect plant health over time. The change detection maps also illustrate how land cover has evolved, displaying areas of deforestation, forest regrowth or other important alterations.

6.5 Discussion

Rosa et al. (2013) indicated that by integrating statistical analysis, spatial maps, time series analysis and remote sensing data provides a comprehensive framework for understanding fire risk and its environmental impacts. Statistical correlations and regression models offer a quantitative grasp of how key variables influence fire risk, enabling predictive modeling and the development of targeted mitigation strategies. Spatial maps of fire hotspots reveal critical areas of high fire intensity, allowing for targeted conservation efforts and better fire management. Time series analyses offer a dynamic view of how fire risk fluctuates over time, enabling the development of seasonal and long-term strategies to manage and mitigate fire risks. Remote sensing data enriches this depth of understanding by providing detailed information on vegetation health and land cover changes, presenting the ecological consequences of fire activity.

Findings emphasize the importance of employing an assortment of analytical tools to address complex environmental challenges, enhance data-driven and improve fire management practices. The integration of these methods underscores the need for continuous innovation and adaptation in environmental monitoring to address evolving challenges and support sustainable land management practices.

7. Conclusion

This study headlining the importance of employing a range of analytical methods to tackle the intricate challenges of fire risk management in the Amazon region. The use of statistical techniques, spatial mapping, temporal analysis and remote sensing has provided a nuanced understanding of fire dynamics, revealing the main drivers and patterns of fire activity.

Through statistical analysis, significant links were identified between variables such as rainfall and fire risk, offering predictive acumen. Spatial maps identified areas of significant risk effectively, aiding targeted conservation efforts. Temporal analyses revealed seasonal variations and trends, crucial for timely interventions. The remote sensing component, especially NDVI and change detection, offered a clear view of vegetation health and land cover changes, framing the broader environmental impacts of fire events.

Ultimately, this work demonstrates the potential of analytical tools to inform fire management strategies and starring the need for continued innovation and collaboration in environmental research. The preservation of the Amazon will depend on our ability to adapt and refine these methodologies, ensuring that they evolve to meet the changing conditions and challenges facing this critical ecosystem.

References

Rosa I.M.D., Purves D., Souza C., Jr, Ewers R.M., 2013: Predictive Modelling of Contagious Deforestation in the Brazilian Amazon. *PLoS ONE* 8(10): e77231. doi.org/10.1371/journal.pone.0077231.

Andela N., Morton D.C., Schroeder W., Chen Y., Brando P.M., Randerson J.T., 2022: Tracking and classifying Amazon fire events in near real time. *Science Advances*. 2022 Jul 29;8(30):eabd2713. doi.org/10.1126/sciadv.abd2713.

Silva, J.M.C., Rylands, A.B., Fonseca, G.A.B., 2005: The Fate of the Amazonian Areas of Endemism. *Conservation Biology*, 19, 689−694. doi.org/10.1111/j.1523-1739.2005.00705.x.

Silva Junior, C.A.d., Lima, M., Teodoro, P.E., Oliveira-Júnior, J.F.d., Rossi, F.S., Funatsu, B.M., Butturi,W., Lourençoni, T., Kraeski, A., Pelissari, T.D., et al., 2022: Fires Drive Long-Term Environmental Degradation in the Amazon Basin. *Remote Sensing* 14, no. 2: 338. doi.org/10.3390/rs14020338.

Gomes, F.B.R., Canella, C.M., Branco, O.E.A., Castro, M.C.C.S., Castro S.R., 2020: Deforestation in Legal Amazon: A Panel Data Analysis of Potential Interferers. *Journal of Management and Sustainability*, vol. 10(2), pages 1-97, December. doi.org/10.5539/jms.v10n2p97.

Rappaport, D.I., Morton, D.C., Longo, M., Keller, M., Dubayah, R., dos-Santos, M.N., 2018: Quantifying long-term changes in carbon stocks and forest structure from Amazon forest degradation. *Environmental Research Letters*, 13(6), 065013. doi.org/10.1088/1748-9326/aac331.

Qin, Y., Xiao, X., Dong, J., Zhang, Y., Wu, X., Shimabukuro, Y., Arai, E., Biradar, C., Wang, J., Zou, Z., Liu, F., Shi, Z., Doughty, R., Moore, B., 2019: Improved estimates of forest cover and loss in the Brazilian Amazon in 2000–2017. *Nature Sustainability* 2, 764–772. doi.org/10.1038/s41893-019-0336-9.

Silveira, M.V.F., Silva-Junior, C.H.L., Anderson, L.O., Aragão, L.E.O., 2022: Amazon fires in the 21st century: The year of 2020 in evidence. *Global Ecology and Biogeography*, 31, 2026–2040. doi.org/10.1111/geb.13577.

Tyukavina, A., Hansen, M.C., Potapov, P.V., Stehman, S.V., Smith-Rodriguez, K., Okpa, C., Aguilar, R., 2017: Types and rates of forest disturbance in Brazilian legal Amazon, 2000–2013. *Science Advances*, 3(4), e1601047. doi.org/10.1126/sciadv.1601047.

Oliveira, L.J.C., Costa, M.H., Soares-Filho, B.S., Coe, M.T, 2013: Large-Scale Expansion of Agriculture in Amazonia May Be a No-Win Scenario. *Environmental Research Letters.* 2013, 8, 24021. doi.org/10.1088/1748-9326/8/2/024021.

Cheng Z., Wu S., Wei D., Pan H., Fu X., Lu X., Yang L., 2024: Current Status of Research on Wildland Fire Impacts on Soil Environment and Soil Organisms and Hotspots Visualization Analysis. *Fire*. 2024; 7(5):163. doi.org/10.3390/fire7050163.

Mohd Said, S., Zahran, E., Shams, S., 2017: Forest Fire Risk Assessment Using Hotspot Analysis in GIS. *The Open Civil Engineering Journal*. 11. doi.org/ 10.2174/1874149501711010786.

Oliveira, R., Venturieri, A., Sampaio, S., Lima, A., Rocha, E., 2016: Dinâmica de uso e cobertura da terra das regiões de integração do Araguaia e Tapajós/PA, para os Anos de 2008 e 2010. *Revista Brasileira de Cartografia*, No 68/7: 1411-1424, Brazil. doi.org/10.14393/rbcv68n7-44369.

Becker, B.K., 2001. Revisão das políticas de ocupação da Amazônia: é possível identificar modelos para projetar cenários. *Parcerias Estratégicas,* v. 6 n. 12, 2001, 135-159.

Penteado, A.R., 1967: *Problemas de colonização e de uso da terra na região bragantina do estado do Pará,* Vol. I & II. Federal University of Pará.

Botelho, J. Jr., Costa, S.C.P., Ribeiro, J.G., Souza, C.M. Jr., 2022: Mapping Roads in the Brazilian Amazon with Artificial Intelligence and Sentinel-2. *Remote Sensing* 14, no. 15: 3625. doi.org/10.3390/rs14153625.

Lopes, G., Delbem, A., Sousa, J., 2021. Introdução à Análise de Dados Geoespaciais com Python. *Minicursos do XIV Encontro Unificado de Computação do Piauí (ENUCOMPI) e XI Simpósio de Sistemas de Informação (SINFO)* (pp.82-106), Brazil. doi.org/10.5753/sbc.7669.6.5.

Stancin I., Jovic A., 2019. An overview and comparison of free Python libraries for data mining and big data analysis. *2019 42nd Int. Conv. Inf. Commun.Technol. Electron. Microelectron. (MIPRO) Opatija, Croatia, 2019 - Proc., pp. 977–982*. doi:10.23919/MIPRO.2019.8757088.

Zahran, E.M., Shams, S., Said, S.N., 2020: Validation of Forest Fire Hotspot Analysis in GIS Using Forest Fire Contributory Factors. *Systematic Reviews in Pharmacy* 2020;11(12):249-255. doi.org/10.31838/srp.2020.12.40.

Sial, A.H., Yahya, S., Rashdi, S., Khan, D.A., 2021: Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python. *International Journal of Advanced Trends in Computer Science and Engineering*. doi.org/10.30534/ijatcse/2021/391012021.

Ramalho, L., 2015: *Fluent Python: Clear, Concise, and Effective Programming*. O'Reilly Media.

McKinney W., 2022: *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter*. O'Reilly Media.

Ma, R., Xia, C. L., Zhang, J. Q., Shen, X. J., 2023: Spatiotemporal variation of vegetation NDVI and its response to climate change in the temperate grassland regions of China. *Chinese Journal of Ecology*. 42, 395–405. doi.org/10.13292/j.1000-4890.202302.015.

Hunter J.D., 2007: Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95. doi:10.1109/MCSE.2007.55.

Lemenkova, P., 2020. Python Libraries Matplotlib, Seaborn and Pandas for Visualization Geo-spatial Datasets Generated by QGIS. *Analele stiintifice ale Universitatii "Alexandru Ioan Cuza" din Iasi - seria Geografie*, vol. 64(1), pp. 13-32, 2020. doi.org/10.15551/scigeo.v64i1.386.

Faraway, J., 2014: *Linear Models with R*. Chapman & Hall/CRC Texts in Statistical Science.

Viana, C.M., Oliveira, S., Rocha, J., 2024. Introductory Chapter: Time Series Analysis. *IntechOpen*. doi.org/10.5772/intechopen.1004609.

Holmgren, P., Thuresson, T., 1997: Satellite remote sensing for forestry planning — A review. *Scandinavian Journal of Forest Research*, 13, 90-110. doi.org/10.1080/02827589809382966.