Deforestation Correlation with Air Pollution in Amazon

Danial Mahmoudi

January 2025

1 Introduction

The Amazon rainforest is crucial for global environmental stability, but its ongoing deforestation has far-reaching consequences, including potential impacts on air pollution in urban centers like São Paulo which faces an additional layer of complexity from the influence of Amazon deforestation. This report seeks to answer the question: **How does deforestation in the Amazon correlate with pollution outcomes in Brazil, specifically in São Paulo?** By investigating this relationship, the aim is to better understand the connections between deforestation, atmospheric changes, and urban pollution.

2 Used Data

2.1 Global Forest Watch - Amazon Deforestation (SAD Alerts)

- **Source:** Global Forest Watch
- **Data Description:** This dataset contains geo-referenced deforestation and degradation alerts in the Brazilian Amazon from 2008 to 2018. Key attributes include the date of the alert, the affected area, and spatial coordinates.
- **Reason for Choosing:** The dataset is highly relevant for assessing the environmental impact of deforestation on pollution outcomes in São Paulo, particularly given the focus on deforestation's role in altering atmospheric conditions.
- **License Obligations and Compliance:** The dataset is provided under the Open Data Commons Attribution License (ODC-By). Proper attribution has been ensured in all derived materials and outputs to meet license requirements.

2.2 Kaggle - Air Pollution in São Paulo

- **Source:** Kaggle
- **Data Description:** This dataset includes hourly measurements of air pollutants such as PM10, PM2.5, NO2, O3, CO, SO2, Benzene, and Toluene in São Paulo, spanning from 2013 onward.

- **Reason for Choosing:** The dataset directly supports the analysis of the impact of Amazon deforestation on pollution levels in São Paulo.
- **License Obligations and Compliance:** The dataset is used under the Kaggle license, and attribution to the original source is explicitly provided in this report.

2.3 Final Data Structure

- **Deforestation Data:** A monthly aggregated dataset containing the total deforested area (in square kilometers) for each month from May 2013 to December 2018.
- **Pollution Data:** A monthly aggregated dataset containing average pollutant levels (PM10, PM2.5, NO2, O3, CO, SO2, Benzene, Toluene) for the same period.

3 Analysis

3.1 Pollutants Most Correlated with Deforestation

To analyze the relationship between deforestation and air pollution, a Spearman's rank correlation analysis was conducted. This method was chosen due to the following:

- **Non-Normality:** Shapiro-Wilk tests confirmed that key variables (e.g., deforestation area, PM10, and others) were not normally distributed, as illustrated in the Q-Q plot (Figure 1).
- **Non-Linearity:** Regression models revealed low explanatory power ($R^2 = 0.037$–0.233), indicating non-linear relationships.
- **Robustness:** Spearman's correlation is robust to outliers and captures monotonic relationships, unlike Pearson's correlation.

Based on the correlation heatmap (Figure 2), the pollutants strongly correlated with deforestation are:

- **PM2.5:** A strong positive correlation ($r = 0.49$) was identified. This aligns with findings by Andreae & Merlet (2001), which attribute fine particulate emissions to slash-and-burn practices during deforestation.
- **PM10:** A moderate positive correlation ($r = 0.41$) was observed. This supports conclusions by Andreae & Merlet (2001) who demonstrated that biomass burning releases significant coarse particulate matter (PM10) as a result of deforestation and land clearing practices.
- **NO2:** A moderate positive correlation ($r = 0.45$) was noted, which supports conclusions by Yokelson et al. (2008), who demonstrated that biomass burning releases significant nitrogen oxides (NOx) as a result of deforestation and land clearing practices.
- **O3:** Interestingly, ozone exhibited a negative correlation ($r = -0.29$). Studies like Sillman et al. (1999) suggest Ozone is primarily formed in urban and industrial areas rather than directly from biomass burning.

Benzene and Toluene showed the strongest correlation ($r = 0.50$ and $r = 0.51$, respectively), but they were excluded from further analysis for two main reasons:

- These pollutants are heavily influenced by industrial and urban activities, making them less reliable as indicators of deforestation-related pollution.
- Significant portions of the Benzene and Toluene data were missing and had to be zero-filled, introducing potential biases and limiting the robustness of the analysis.

3.2 Key Findings

- **PM2.5** and **PM10** are the most reliable indicators of deforestation, given their strong correlation and support from prior studies.
- **NO2** also shows a moderate correlation, reinforcing its relevance in deforestation-related pollution.
- **Ozone's negative correlation** highlights the robustness of the analysis by capturing both positive and negative relationships.

3.3 Deforestation Trends vs Pollutants

In addition to the correlation analysis, further exploration was done by plotting the trends of deforestation alongside the pollutant levels.

- **Deforestation vs PM2.5:** A trend analysis was conducted by plotting the deforestation area against PM2.5 levels over time (Figure 3). Due to differing scales, the deforestation area was plotted on the left y-axis, while PM2.5 levels were plotted on the right y-axis for better clarity, revealing a clear upward trend in both deforestation and PM2.5, further supporting the correlation between the two.
- **Deforestation vs O3:** A similar trend analysis was conducted between deforestation and O3 (Figure 4). A negative trend was observed, where increased deforestation is associated with a decrease in ozone levels. This supports the negative correlation observed earlier, as ozone is primarily formed in urban and industrial areas rather than directly from biomass burning.

3.4 Figures

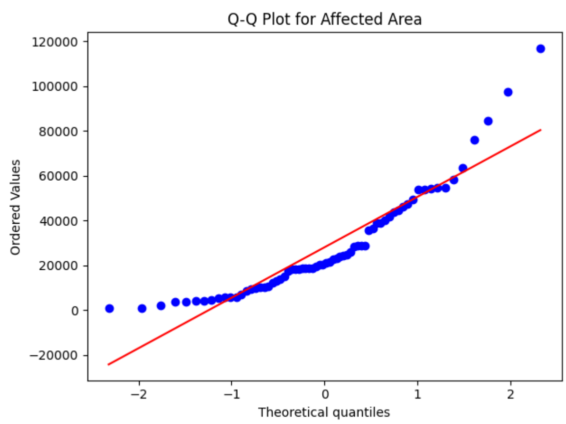

Figure 1: Q-Q plot for affected area.

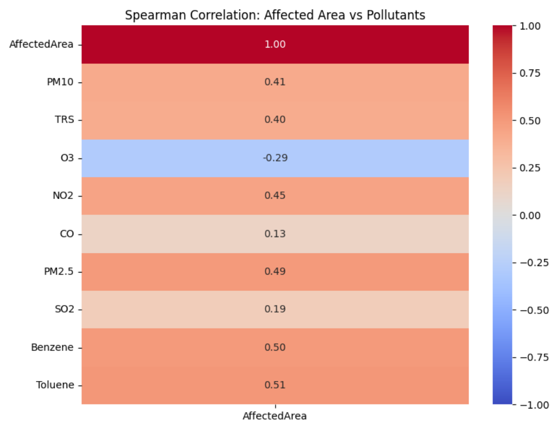

Figure 2: Spearman's rank correlation.

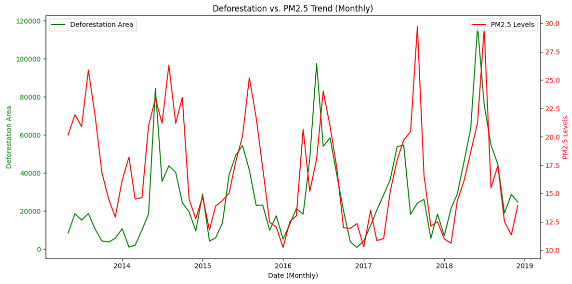

Figure 3: Deforestation trend vs PM2.5.

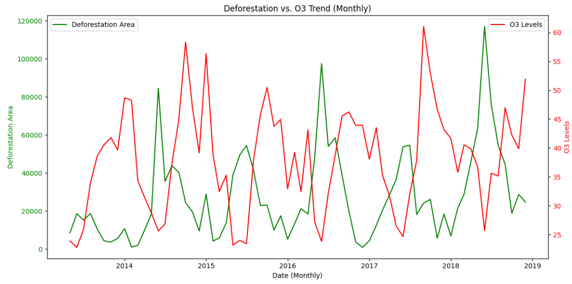

Figure 4: Deforestation trend vs O3.

4 Conclusions and Limitations

In this analysis, I aimed to explore the correlation between deforestation and various pollutants. Using Spearman's rank correlation, I found a significant positive correlation between deforestation and pollutants such as PM2.5, PM10, and NO2. These results provide strong evidence that deforestation is associated with increased pollution levels. However, my analysis did not intend to directly study the causes of this correlation, such as biomass burning (a primary method of deforestation), but rather sought to confirm that the observed correlations were genuine and not the result of random factors. My goal was to show that a plausible causation exists, supported by existing literature.

While the correlation is evident, several limitations must be acknowledged. My study primarily relies on statistical correlations, which can suggest a relationship but do not prove direct causality. Additional factors, such as local industrial activities or seasonal variations, could influence the observed trends. Furthermore, the reliance on available data introduces potential biases, especially due to missing data for some pollutants. Despite these limitations, I believe the analysis successfully demonstrates a clear relationship between deforestation and pollution, offering a foundation for further investigation into the underlying causes.

References

- Andreae, M. O., & Merlet, P. (2001). Emission of trace gases and aerosols from biomass burning. *Global Biogeochemical Cycles, 15*(4), 955-966. `https://doi.org/10.1029/2000GB001382`
- Yokelson, R. J., Christian, T. J., Karl, T. G., & Guenther, A. (2008). The tropical forest and fire emissions experiment: Laboratory fire measurements and synthesis of campaign data. *Atmospheric Chemistry and Physics, 8*, 3509–3527. `https://doi.org/10.5194/acp-8-3509-2008`
- Sillman, S. (1999). The relation between ozone, NOx, and hydrocarbons in urban and polluted rural environments. *Atmospheric Environment, 33*(12), 1821-1845. `https://doi.org/10.1016/S1352-2310(98)00345-8`
- Global Forest Watch. (2019). *Deforestation alerts dataset.* Retrieved from https://data.globalforestwatch.org/datasets/gfw::sad-alerts/about
- Kaggle. (2020). *Air Pollution at São Paulo, Brazil since 2013.* Retrieved from https://www.kaggle.com/datasets/danlessa/air-pollution-at-so-paulo-brazil-since-2013