



**Detecting and Assessing Pollution Events from Wildfires
Using Remote Sensing and Meteorological Data
A Data Science Approach**

Sofia Margarida Matias Rodrigues

Dissertação para a obtenção do Grau de Mestre em
**Ciências dos Dados em Agricultura, Alimentação, Floresta e
Ambiente**

Orientador(es): Ana Russo
Célia Gouveia

(Versão provisória)

2025



Acknowledgments

Abstrato

Os incêndios florestais contribuem significativamente para a poluição do ar ao libertarem matéria particulada (PM) e gases tóxicos na atmosfera, com as alterações climáticas a preverem um aumento na atividade dos incêndios florestais e na propagação do fumo, agravando os riscos para a saúde. O trabalho a ser desenvolvido será apoiado por um quadro orientado por dados para monitorizar e avaliar a poluição do ar causada por incêndios florestais, uma questão urgente de saúde e ambiental que afeta a população global. O trabalho irá analisar os resultados dos modelos atmosféricos atuais, juntamente com indicadores de deteção remota, como o Fire Radiative Power (FRP) e o Fire Radiative Energy (FRE), combinados com dados meteorológicos e aprendizagem automática, para melhorar a deteção de eventos de poluição. O objetivo é analisar os impactos transfronteiriços das emissões de incêndios florestais, avaliar tecnologias de deteção remota (e.g., MODIS, SEVIRI, Sentinel) no monitoramento de incêndios florestais e examinar métodos de ciência de dados para monitorização ambiental. Na secção de dados e metodologia, será descrita a integração de dados meteorológicos e de deteção remota, com modelos de aprendizagem automática, como Random Forest, XGBoost e Redes Neurais, usados para classificar eventos de poluição e mapear padrões espaço-temporais do fumo. A validação do modelo será realizada através da comparação dos resultados com eventos históricos extremos de incêndios florestais para verificar a sua precisão. O modelo será, então, avaliado pelo seu desempenho preditivo e fornecerá algumas perceções sobre os padrões de dispersão do fumo dos incêndios florestais, identificando fatores-chave que contribuem para os eventos de poluição. Para concluir o trabalho, serão apresentados os destaques do estudo, demonstrando como os dados de deteção remota e meteorológicos podem melhorar o monitoramento da qualidade do ar e apoiar o planeamento de políticas públicas. Serão propostos trabalhos futuros para melhorar as capacidades de monitoramento em tempo real, integrar fontes de dados adicionais e aplicar as descobertas em quadros ambientais e de saúde mais amplos. Esta pesquisa tem o potencial de informar intervenções estratégicas, reforçando ainda mais as ferramentas de tomada de decisão para gerir a poluição provocada por incêndios florestais.

Palavras-Chave: Fogos, Poluição do Ar, Dados Meteorológicos, Deteção Remota, Aprendizagem Automática

RESUMO PT 1200-1500 PALAVRAS

Abstract

Wildfires contribute significantly to air pollution by releasing particulate matter (PM) and toxic gases into the atmosphere, with climate change projected to increase wildfire activity and the spread of smoke, heightening health risks. The work that will be developed will be supported by a data-driven framework to monitor and assess air pollution from wildfires, a pressing health and environmental issue that affects the global population. The work will analyse current atmospheric models outputs coupled with remote sensing indicators, like Fire Radiative Power (FRP) and Fire Radiative Energy (FRE), combined with meteorological data and machine learning to improve pollution event detection. It aims to look upon transboundary impacts of wildfire emissions, evaluate remote sensing technologies (e.g. MODIS, SEVIRI, Sentinel) in wildfire monitoring, and examine data science methods for environmental monitoring. For the data and methodology section, it will describe the integration of meteorological and remote sensing data, with machine learning models, such as Random Forest, XGBoost, and Neural Networks, used to classify pollution events and track spatial-temporal patterns of smoke. Model validation will be performed by comparing results with historical extreme wildfire events to verify accuracy. Then the model will be evaluated by its predictive performance and have some insights into wildfire smoke dispersion patterns, identifying key factors contributing to pollution events. To conclude the work, the highlights of the study will be shown, demonstrating how remote sensing and meteorological data can improve air quality monitoring and support policy planning. Future work will be proposed to enhance real-time monitoring capabilities, integrate additional data sources, and apply findings within broader environmental and health frameworks. This research has the potential to inform strategic interventions, further strengthening decision-making tools for managing wildfire-driven pollution.

Key Words: Wildfires, Air Pollution, Meteorological Data, Remote Sensing, Machine Learning

Table of Contents

Introduction 9

Literature Review 12

Data 13

Methodology 14

Results..... 15

Discussion..... 16

Conclusion 17

References 18

Appendices 19

List of Figures

Não foi encontrada nenhuma entrada do índice de ilustrações.

List of Tables

Não foi encontrada nenhuma entrada do índice de ilustrações.

List of Abbreviations

IPMA – Instituto Português do Mar e da Atmosfera

CAMS – Copernicus Atmosphere Monitoring Service

PM – Particulate Matter

MODIS – Moderate Resolution Imaging Spectroradiometer

WRF-Chem – Weather Research and Forecasting Model with Chemistry

WRF – Weather Research and Forecasting

SEVIRI – Spinning Enhanced Visible and Infrared Imager

MSG – Meteosat Second Generation

WHO – World Health Organization

WMO – World Meteorological Organization

FRP – Fire Radiative Power

FRE – Fire Radiative Energy

CDS – Climate Data Store

ADS – Atmosphere Data Store

ECMWF – European Centre for Medium-Range Weather Forecasts

Introduction

The World Health Organization (WHO) defines air pollution as a contamination by any chemical, physical or biological agent that changes the natural characteristics of the atmosphere, either indoors or outdoors. There is a large amount of air pollution sources, the most common categories being transportation vehicles (that use motors), industrial facilities and fires. Key air pollutants that pose significant public health risks include particulate matter, carbon monoxide, ozone, nitrogen dioxide, and sulfur dioxide. Data and studies provided by WHO show that 99% of the global population breathes contaminated air, based on the established guideline limits for each pollutant ([World Health Organization, s.d.](#)).

Wildfires are a derivative of a fire that is uncontrolled and can take large proportions because they are usually situated in rural areas with difficult access. Historically, they could be a beneficial tool to certain natural landscapes by clearing the underbrush and allowing seed release for some species. However, climate change and anthropogenic influence contribute to extreme episodes and cause larger burned areas, which causes a major imbalance in the ecosystems and carbon storage ([National Geographic, 2023](#)). The proportions that a fire can take depend on the weather conditions, such as wind, high temperatures, lack of rainfall for a long period of time, topography, road access and available means ([National Geographic, 2025](#)).

The WHO states that wildfires are increasing its frequency, severity and duration, bringing concerns to the health of those affected by its exposure. The smoke is a mixture of hazardous air pollutants, like PM2.5, NO2 and Ozone (O3). Even though they are aggravated by meteorological extremes, like droughts, heat waves and high winds, wildfires can also be a source that impact the climate when they release large amounts of carbon dioxide (CO2) and other greenhouse gases into the atmosphere ([World Health Organization, s.d.](#)). The PM2.5 is produced both directly from combustion and through atmospheric chemical reactions, making it the main concern for public health, exceeding air quality standards – sometimes reaching over 500 µg/m3, when the standard values, established by the WHO, should be 0-10 µg/m3 for a very good air quality ([De Sario, Katsouyanni, & Michelozzi, 2013](#)) ([QualAr, 2019](#)).

Meteorological extremes or extreme weather events are characterized by the World Meteorological Organization (WMO) as rare events “at a particular place and time of year, with unusual characteristics in terms of magnitude, location, timing, or extent” ([World Meteorological Organization, s.d.](#)). Climate change is one of the main reasons for the increase in the impacts of extremes events that take place, and wildfires are directly linked to that, in a sense that often occur alongside or a result of other climate extremes (e.g.

Commented [SM1]: Adicionar à lista de abreviaturas?

heatwaves, droughts or high winds) (AghaKouchak, et al., 2020). The Mediterranean region is particularly susceptible to weather and climate variability, being more prone to larger occurrences of fires, while being directly related to extreme events (Bento, et al., 2023). That being said, there is a concern that wildfires will become more aggravated over the years and with that, the air quality will degrade even more. So it is important to develop studies relating these two topics, air pollution and wildfires, to come up with strategies that promotes interventions and initiatives.

Atmospheric dispersion models e estatísticos (ML e AI)

data available - <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview> & <https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-forecasts?tab=overview> or <https://ads.atmosphere.copernicus.eu/datasets/cams-europe-air-quality-reanalyses?tab=overview>

The study will have data from two/three different sources that will be described in more detail further. Copernicus is the European Union's Earth observation program that provides free and open data about our planet's environment, by using a network of satellites, ground-based sensors and airborne measurements. It delivers real-time information on climate change, land use, oceans, atmospheric conditions and emergency responses to natural disasters like wildfires or floods. Their data can be used to provide relevant information to help service providers, public authorities, international organizations or even for academic and research purposes (Copernicus, s.d.). From Copernicus there are two sources that are going to be used to retrieve data – Climate Data Store (CDS) and Atmosphere Data Store (ADS). From CDS, the dataset that is going to be used is ERA5 hourly data on single levels from 1940 to present. ERA5 is defined as the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for the global climate and weather, replacing the ERA-Interim reanalysis. Reanalysis is a method used to create long-term, consistent datasets of past atmospheric conditions by combining weather model simulations with real-world observations, in this way it is reconstructing historical climate and weather patterns instead of predicting the future. A key process behind reanalysis is data assimilation, where new observational data is continuously merged with previous model predictions to improve accuracy (Copernicus, s.d.). The other source, ADS, provides the dataset called CAMS European air quality reanalyses

MODIS

What is frp and fre

Fire Radiative Power (FRP), in megawatts (MW), is one of the variables used to characterize wildfires, it is obtained “from the radiance at the 4- μ m band of satellite sensors and represents the instantaneous radiative energy that is released from actively burning fires”. FRP can be used in many ways to provide information about biomass burning, land cover dynamics and hydrological cycles. But for this study, it gives insights about the rate of emissions in relation to the rate of biomass combustion. This allows to estimate trace gas and aerosol emissions or smoke production. The variable can be obtained from multiple polar-orbiting and geostationary satellites and can be obtained on one of the main sources already mentioned, the MODIS (Li, Zhang, Kondragunta, & Csiszar, 2018). Based on its proportionality to the amount of burned biomass, higher values of FRP indicate more severe fires and in consequence larger levels of smoke production, leading to higher emissions of particulate matter and other pollutants (Duraó, Alonso, Russo, & Gouveia, 2024). Fire Radiative Power (FRE) is estimated via temporal integration from the measures of FRP (during the lifetime of a fire). By representing the total amount energy release during a fire, it can also provide the total amount of consumed biomass (Instituto de Meteorologia, 2009). Or <https://user.eumetsat.int/catalogue/EO:EUM:DAT:MSG:FRP-SEVIRI>

Model selection

Estudos feitos

Objectives gerais e particulares

Research questions to answer

Literature Review

Air Pollution:

- Discuss the impacts of particulate matter and toxic gases emitted from wildfires, with a focus on the transboundary nature of pollution and its exacerbation due to climate change.

Remote Sensing in Wildfire Monitoring:

- Review remote sensing tools, including MODIS, SEVIRI, and Sentinel products, used in monitoring wildfire activity and emissions. Discuss the utility of FRE and FRP as measures of biomass combustion rate and their effectiveness in tracking pollution.

Data Science in Environmental Monitoring:

- Explore data science methods in atmospheric and environmental sciences, including machine learning models used in pollution detection and forecasting.

Data

EXPLICAR ERA5 E CAMS DATA

Data Collection:

- Meteorological Data: Gather atmospheric data, such as temperature, humidity, wind speed, and direction, from sources like CAMS and monitoring stations.
- Remote Sensing Data: Use FRE and FRP data from MODIS, SEVIRI, and Sentinel-3 for monitoring fire radiative power and energy release.

Data Processing:

- Integrate meteorological and remote sensing data with pollution monitoring station data to create a comprehensive dataset.
- Preprocess data by cleaning and transforming it to address missing values, noise, and scaling for model input.

Methodology

Model Development:

- Supervised Machine Learning Models: Implement and compare models (e.g., Random Forests, XGBoost, and Neural Networks) to classify pollution events and identify anomalies related to wildfire smoke.
- Spatial and Temporal Analysis: Use geospatial tools to map the spatial reach of pollution plumes, identifying the temporal dynamics of smoke dispersal patterns.

Evaluation and Validation:

- Validate the model using metrics such as accuracy, F1-score, and area under the curve (AUC) for classification tasks.
- Cross-reference predicted pollution events with historical extreme events (e.g., Portugal's 2017 megafires) to validate spatial and temporal accuracy. Analyse the impact of compound events (e.g. droughts and heatwaves) to the magnitude of fire-driven pollution events.

Results

Model Performance:

- Present the results of model performance, discussing the predictive accuracy in detecting pollution events and identifying contributing factors from FRE, FRP, and meteorological data.

Discussion

Spatial and Temporal Impact Analysis:

- Discuss the spatial and temporal patterns of wildfire smoke distribution. Evaluate how FRP and FRE data correlate with pollutant concentrations across affected regions.

Limitations and Future Improvements:

- Discuss potential limitations of the model, such as sensitivity to specific atmospheric conditions or data quality, and suggest directions for future improvement.

Conclusion

Summary of Findings:

- Summarize the effectiveness of remote sensing and meteorological data in detecting and mapping wildfire-induced pollution events.

Implications:

- Discuss the potential of this approach to support decision-making interventions and policy planning.

Future Work:

- Suggest advancements in real-time monitoring systems, the inclusion of additional data sources, and potential integration with other environmental and health monitoring frameworks.

References

- Durao, R., Alonso, C., Russo, A., & Gouveia, C. (14-19 de April de 2024). *The role of fire radiative power to estimate fire-related smoke pollution*. Vienna, Austria: EGU General Assembly 2024. doi:<https://doi.org/10.5194/egusphere-egu24-13237>
- Instituto de Meteorologia. (23 de March de 2009). *Fire Radiative Power Pixel - MSG*. Obtido em 24 de February de 2025, de [geonetcastamericas.noaa.gov](https://www.geonetcastamericas.noaa.gov): https://www.geonetcastamericas.noaa.gov/products/navigator/details/EO_EUM_DAT_MSG_FRP-SEVIRI.html
- Li, F., Zhang, X., Kondragunta, S., & Csiszar, I. (2 de May de 2018). Comparison of Fire Radiative Power Estimates From VIIRS and MODIS Observations. *Journal of Geophysical Research: Atmospheres*, 4545-4563. doi:<https://doi.org/10.1029/2017JD027823>
- National Geographic. (19 de October de 2023). *Wildfires: How They Form, and Why They're so Dangerous*. Obtido em 2 de February de 2025, de education.nationalgeographic.org: <https://education.nationalgeographic.org/resource/wildfires-how-they-form-and-why-theyre-so-dangerous/>

