IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10642963

MULTIMODAL DATASET FOR WILDFIRE RISK PREDICTION IN CYPRUS

Maria Prodromou [1,2], Stella Girtsou[3], George Leventis[1,2], Dimitris Koumoulidis[1,2], Marios Tzouvaras[1,2], Christodoulos Mettas[1,2], Alexis Apostolakis [3], Mariza Kaskara[3], Haris Kontoes[3], Diofantos Hadjimitsis[1,2]

1 ERATOSTHENES Centre of Excellence, Limassol, 3012, Cyprus
2 Department of Civil Engineering and Geomatics, Cyprus University of Technology, Limassol, 3036, Cyprus
3 National Observatory of Athens, Operational Unit BEYOND Centre for Earth Observation Research and Satellite Remote Sensing IAASARS/NOA, GR-152 36 Athens, Greece

ABSTRACT

Wildfires detection is a major issue for authorities. There are various causes of fire events with the most common being human influence. A fire risk prediction model through the analysis of geo-environmental and climate data is important for early warning and fire management. In this work, a dataset from multiple modalities, including road density, travelers, forest-agriculture interface, burned areas from historical fire events, metrological data, land cover, vegetation indices from data cube, is generated. Artificial intelligence and machine learning models can use this multimodal dataset to improve forest fire management.

Index Terms— data cube, wildfires, Cyprus, fire management

1. INTRODUCTION

Forest significantly impact ecosystems worldwide and are frequent geophysical event, with an estimated 3.5-5 million km^2 [1], [2] burned area annually based on satellite data. In the Mediterranean, increases in the frequency and affected areas have been noted, leading in property loss, deaths and significant environmental degradation [3], [4].

Cyprus is located in the Eastern Mediterranean, an area where forest fires frequently occur, especially during the summer period. Factors like drought, hot summers and flammable vegetation contribute to the increased risk of forest fires [5], [6]. Effective fire management strategies, including early warning systems and the use of remote sensing, are crucial for mitigating these risks.

Effective wildfire risk prediction and management depend on the up-to-date, spatial explicit representation of the environment, mainly focusing on the biomass and characteristics of live and dead vegetation, which is the primary factor influencing fire behavior and risk [7]. Furthermore, an effective wildfire risk prediction can improve the decision-making, evaluation, and risk management, which is a prerequisite for the achievement of the EU forest strategy for the 2030 goal to improve the quantity and quality of EU multi-functional forests [7]–[9].

Taking that into consideration, forest fires can be analyzed using a data cube approach, which integrates various variables related to wildfire drivers and historical records of ignitions and burned areas. This allows for the assessment of machine learning models in tasks such as short-term wildfire danger forecasting and final burned area estimation. The use of data cubes in wildfire modeling provides opportunities for the implementation of additional tracks to mitigate the increasing threat of wildfires in the Mediterranean region [10], [11]. Additionally, the integration of spatial parameters with the probability of fire occurrence using classical frequent itemset algorithms and GIS can help in spatial forecasting of forest fires and generating forest fire risk zone maps with good prediction accuracy [12]. Satellite observations, specifically using Moderate Resolution Imaging Spectroradiometer (MODIS) data, can be used to develop unsupervised spatio-temporal data mining methods for mapping forest fires, overcoming limitations in both data and methods employed by prior efforts [12]. Other approaches include leveraging a nearest neighbor imputation model that integrates LiDAR and Landsat data sources into a space-time data cube, enabling efficient analysis of vast forest structure data in the form of time series analyses [13].

The performance of machine learning models commonly used in wildfire modeling tasks can be influenced by the quality and representativeness of the data used. To that end, approaches like [14], [15] combine data cubes, machine learning, and geospatial ontology-based data access (OBDA) technologies. This integration allows for effective harmonization of diverse data sources, enhancing the accuracy and efficiency of fire risk computations. The use of geospatial OBDA technologies is a key feature, facilitating semantic querying and improving data interpretation, thereby offering a more advanced fire risk management tool.

Highlighting the importance of forest fire monitoring and management for effective climate mitigation, ecosystem conservation, and the reduction of biodiversity loss, the

current work aims to contribute to fields like efficient mapping, analysis and management of forest fires, especially within the Eastern Mediterranean and Middle East (EMMENA) region. This is achieved through the creation of data cubes that contain satellite data alongside several vegetation and meteorological data. In-detail information can be found at the respective "Data Collection" section.

As shown above, available -though limited studies in the field- literature emphasize the utilization of data cube frameworks for managing complex spatial-temporal datasets, thus improving the capability to map and monitor forest attributes and their changes over time.

The aim of this study is to generate a dataset for fire risk prediction from different modalities. Section 2 describes the modalities of the dataset. Section 3 gives a statistical analysis. Section 4 gives the discussion and conclusions.

2. DATA COLLECTION AND PROCESSING

2.1. Data Collection

According to data provided by the Forest Department for the period 2000-2022, as shown in Figure 1, 37.2% of fire events in Cyprus were deliberate, followed by agricultural activities that correspond to 19.7%. Obviously, the majority of fire events can be linked to human activities while a smaller percentage are caused by natural causes (e.g. lighting).

Moreover, abandoned agricultural lands significantly increase the risk of fires particularly during the summer period where the flammable vegetation components increase. Additionally, grazing and specifically overgrazing constitutes land degradation. The proximity to the roads and urban areas further increases vulnerability. In addition, the lack of preventive measures in private forests is also an important threat.

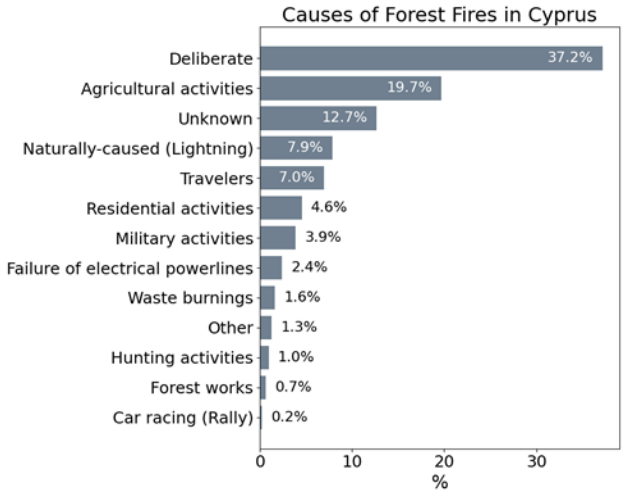

Figure 1. Causes of forest fires in Cyprus for the period 2000-2022

To effectively model, map and predict forest fires, it is essential to compile a comprehensive set of independent explanatory variables known as fire causative factors. These factors are selected based on their potential correlation with the unique characteristics of the area investigated, the historical fire events, and the availability of relevant data [16]. The variables most frequently employed to simulate and forecast the incidence of fires are obtained from the analysis of topography, precipitation, and vegetation conditions, due to their significant correlations with fuel conditions [17].

Throughout this study, the following eight variables were collected that encompass a diverse range of geo-environmental and climate factors:

1. **Road density:** To comprehensively model human activities and predict deliberate wildfires, a road density layer was meticulously crafted using OpenStreetMap (OSM). This layer is based on the premise that a more intricate road network enhances accessibility to forested areas, consequently elevating the likelihood of wildfire occurrences.
2. **Travelers**: Activities like picnic, natural trails and camping sites significantly influence the occurrence of fires, so three layers were generated using the data that provided by the Republic of Cyprus Open Data Portal.
3. **Forest-agriculture interface:** Recognizing the heightened wildfire ignition risk in regions characterized by intricate forest-agriculture mixtures [18], specialized layer was developed detailing the forest-agriculture interface in Cyprus. This layer serves as a valuable resource in refining the accuracy of our fire risk prediction models.
4. **Burned Areas:** The burned areas were collected by EFFIS - European Forest Fire Information System, which supports the services in charge of the protection of forests against fires in the EU and neighbor countries and enriched with data from Forest Department in Cyprus.
5. **Digital Elevation Model (DEM):** The DEM for the purposes of this study was retrieved from Copernicus' EU-DEM. This product is a hybrid one, based on SRTM and ASTER GDEM data fused by a weighted averaging approach with a spatial resolution at 25m (https://land.copernicus.eu/imagery-in-situ/eu-dem). The DEM was used to obtain the elevation, slope, and aspect. These topographic features were selected because they are widely recognized as key variables that are closely linked to fire-related phenomena. They provide a significant influence on the spatiotemporal patterns of various other variables, including precipitation, temperature, humidity, evapotranspiration, soil moisture, and wind speed. Furthermore, they play a crucial role in determining the occurrence and behavior of fires, such as their speed and spread [19]
6. **Land cover.** The Copernicus' Corine Land Cover (https://land.copernicus.eu/dashboards/clc-clcc-2000--2018) (CLC) datasets from the years 2012 and 2018, with

a spatial resolution of 100m, were obtained and processed for this study. Land cover facilitates the development of fire probability models because it serves as an indirect indicator of the landscape's susceptibility to fire [20]. For example, the Xerothermic environment in the Mediterranean region, particularly during the summer season, is marked by high air temperatures and low relative humidity. These conditions create conducive environments for incidents with high flammability.

7. **Meteorology**: All meteorological factors analyzed, including Temperature, Dewpoint, Wind speed, Wind direction, and Precipitation, were retrieved from the ERA5-Land reanalysis datasets, which are accessible through the Copernicus EO program. These factors significantly contribute to the occurrence and intensity of forest fires, and their variability directly impacts the frequency and severity of wildfires. These variables were downscaled to a finer resolution of 500m using the nearest neighbor interpolation.

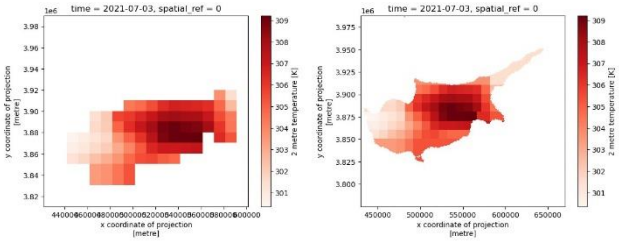

Figure 2. Resample of the spatial resolution for the Meteorological dataset at 500m.

8. **Vegetation indices:** The Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) were obtained from NASA products of the Moderate Resolution Imaging Spectroradiometer (MODIS), MYD13A1 and MOD13A1. Each pixel value of those products is an average of the daily products collected within 16 days, with 500m spatial resolution (https://modis.gsfc.nasa.gov/data/dataprod/mod13.php). The importance of vegetation condition in modeling fire occurrence is well-established, as it directly influences the probability of burning and subsequent fire characteristics, such as size and severity [21]. The NDVI and the EVI are extensively utilized as they effectively capture the characteristics and variations in surface biomass, providing valuable insights into the composition, distribution, and dynamics of vegetation communities [22].

2.2. Data Processing

The collected data were preprocessed as shown in Figure 3 in order to clean and normalize to ensure consistency across the dataset (spatially and temporally). Additionally, the data annotation was conducted in order to label the data with relevant metadata to make the retrieval and analysis of the data more efficient. After that, the annotated data were populated into the data cube, structured to support easy access and their manipulation.

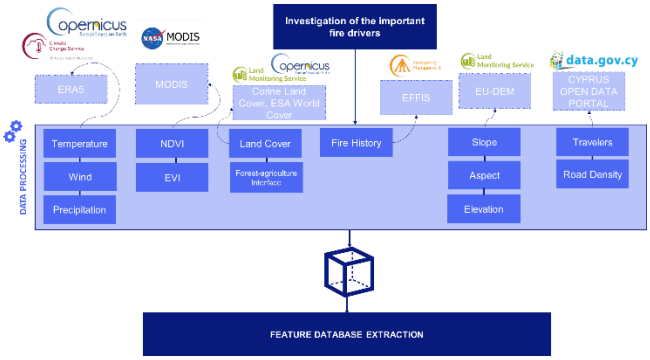

Figure 3 Workflow diagram for the proposed methodology

3. EXPLORING TRENDS AND RELATIONSHIPS USING THE DEVELOPED MULTIMODAL DATASET

In this section, examples of the visualizations for various factors are presented. Data cubes offer a significant advantage through their capability to provide direct access to harmonized data, allowing for straightforward queries in space and time. To build successful wildfire risk prediction models, a comprehensive study of the data is essential to recognize the specificities of the dataset and address any potential discrepancies to ensure the robustness and reliability of the models.

The following Figure 4 is a clear demonstration of datacube's capability to provide immediate data access, showcasing an intuitive, color-coded delineation between fire-affected regions (yellow) and unaffected areas (purple). These plots, ready to be utilized as training labels, highlight a significant imbalance between the two classes. This disparity is a critical factor to consider when developing machine learning algorithms, as it necessitates strategies to ensure balanced representation and prevent model bias.

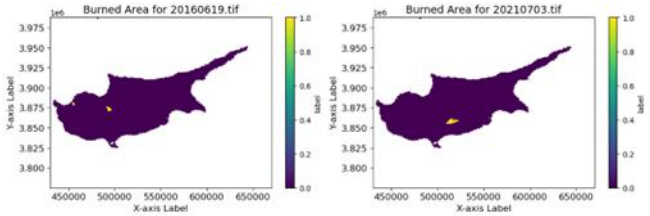

Figure 4. Visualization of burned areas through the data cube.

Following, a visual examination of trends and relationships between temperature, precipitation and burned areas from 2000 until 2022 is presented in Figure 5 and a representation of burned areas. The time series analysis plot shows distinct trends and relationships among temperature, precipitation, and burned areas from 2000 to 2022. The data indicates significant variability in all three parameters. The temperature (red line) shows a fluctuating pattern without a

clear increasing or decreasing trend over the years. Additionally, the precipitation (blue line) also varies with notable peaks around 2005, 2008 and 2017. The burned areas present a decrease over the period, particularly after 2010. Furthermore, the peak in burned areas shows an inverse relationship between precipitation and the extent of burned areas. In contrast the temperature peaks do not consistently align with the highest burned areas, indicating that while temperature plays a role, precipitation is a more dominant factor in influencing the burned areas. Also, this trend can correlate with the status of Cyprus where the 37,2% of fire events were deliberate which are not directly influenced by natural climatic factors which could explain this lack. Taking into account this finding we incorporate human activities in our multimodal dataset.

Moreover 2000, 2007, 2016 and 2021 were the most disastrous in total burned area. Although in 2004 more than 400 fire events were recorded, the total burned area of that year remained at low levels, which suggests that prevailing meteorological and environmental conditions may not have favored rapid fire spread and/or that the suppression efforts managed to put out the fires timely. The latter is also linked with the wildfire's location and the accessibility to the fire trucks.

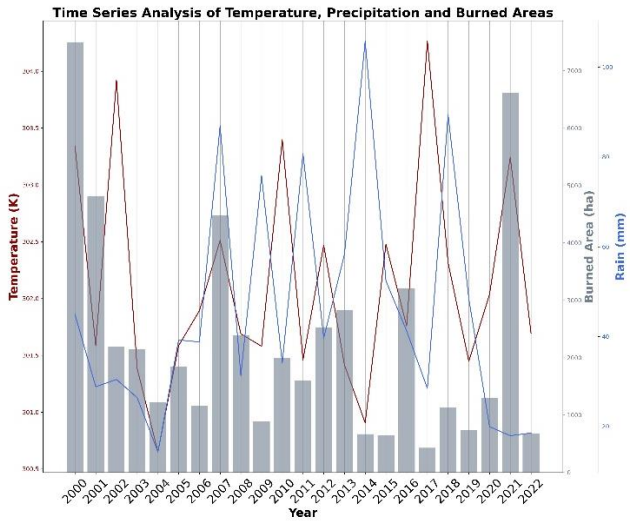

Figure 5. Time series analysis of temperature, precipitation and burned areas in Cyprus for the period 2000-2022.

4. DISCUSSION AND CONCLUSIONS

The spatiotemporal analysis of wildfire occurrence and identification of wildfire risk areas are of significant importance, especially for researchers, managers, and decision makers in the field of hazard mitigation and climate change. Developing an effective fire risk prediction model using environmental data is essential for early warning and fire management. In this study, a comprehensive multimodal dataset was created, integrating data from various sources such as road density, traveler activity, forest-agriculture interfaces, historical burned areas, meteorological data, land cover, and vegetation indices. Multidimensional arrays, in other words datacubes, are a common practice to combine and harmonize different data sources for more comprehensive analysis. This study aims to enhance the ERATOSTHENES Data Cube in order to integrate satellite and vector data from several sources, which will be used in the future to calculate the Fire Risk Prediction Model for Cyprus using machine learning algorithms. The created dataset supports the development of artificial intelligence and machine learning models to improve forest fire management. Apart from that, this dataset can also be utilized for the development of a digital twin for modeling wildfires in the future. Overall, this study demonstrates the potential of using data cube/ multimodal datasets to improve wildfire risk prediction and management. By integrating diverse data sources and utilizing advanced machine learning techniques, authorities can develop more effective strategies for mitigating the impact of wildfires, particularly in areas like Eastern Mediterranean where fire risks are widespread.

5. ACKNOWLEDGEMENTS

The authors acknowledge the 'EXCELSIOR': ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment H2020 Widespread Teaming project (www.excelsior2020.eu). The 'EXCELSIOR' project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857510, from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development and the Cyprus University of Technology.

6. REFERENCES

[1] L. Giglio, L. Boschetti, D. P. Roy, M. L. Humber, and C. O. Justice, "The Collection 6 MODIS burned area mapping algorithm and product," Remote Sens Environ, vol. 217, pp. 72–85, Nov. 2018, doi: 10.1016/j.rse.2018.08.005.

[2] J. Lizundia-Loiola, G. Otón, R. Ramo, and E. Chuvieco, "A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data," Remote Sens Environ, vol. 236, p. 111493, Jan. 2020, doi: 10.1016/j.rse.2019.111493.

[3] J. G. Pausas and V. R. Vallejo, "The role of fire in European Mediterranean ecosystems," in Remote Sensing of Large Wildfires, Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 3–16. doi: 10.1007/978-3-642-60164-4_2.

[4] I. Chrysafis, C. Damianidis, V. Giannakopoulos, I. Mitsopoulos, I. M. Dokas, and G. Mallinis, "Vegetation Fuel Mapping at Regional Scale Using Sentinel-1, Sentinel-2, and DEM Derivatives—The Case of the Region of East Macedonia and Thrace, Greece," Remote Sens (Basel), vol. 15, no. 4, Feb. 2023, doi: 10.3390/rs15041015.

[5] E. Aragoneses and E. Chuvieco, "Generation and mapping of fuel types for fire risk assessment," Fire, vol. 4, no. 3, Sep. 2021, doi: 10.3390/fire4030059.

[6] M. L. Pettinari and E. Chuvieco, "Fire Danger Observed from Space," Surv Geophys, vol. 41, no. 6, pp. 1437–1459, Nov. 2020, doi: 10.1007/s10712-020-09610-8.

[7] S. Girtsou, A. Apostolakis, G. Giannopoulos, and C. Kontoes, "A Machine Learning Methodology for Next Day Wildfire Prediction," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE, Jul. 2021, pp. 8487–8490. doi: 10.1109/IGARSS47720.2021.9554301.

[8] S. Kondylatos, I. Prapas, G. Camps-Valls, and I. Papoutsis, "Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the Mediterranean," no. DL, pp. 1–22, 2023, [Online]. Available: http://arxiv.org/abs/2306.05144

[9] X. C. Chen et al., "A new data mining framework for forest fire mapping," in 2012 Conference on Intelligent Data Understanding, IEEE, Oct. 2012, pp. 104–111. doi: 10.1109/CIDU.2012.6382190.

[10] G. Matasci et al., "A space-time data cube: Multi-temporal forest structure maps from landsat and lidar," in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, Jul. 2017, pp. 2581–2584. doi: 10.1109/IGARSS.2017.8127523.

[11] D. Bilidas et al., "Fire Risk Management using Data Cubes, Machine Learning and OBDA systems," in Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems, New York, NY, USA: ACM, Nov. 2023, pp. 1–4. doi: 10.1145/3589132.3625615.

[12] A. Apostolakis, S. Girtsou, G. Giannopoulos, N. S. Bartsotas, and C. Kontoes, "Estimating Next Day's Forest Fire Risk via a Complete Machine Learning Methodology," Remote Sens (Basel), vol. 14, no. 5, p. 1222, Mar. 2022, doi: 10.3390/rs14051222.

[13] B. T. Pham et al., "Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction," Symmetry (Basel), vol. 12, no. 6, p. 1022, Jun. 2020, doi: 10.3390/sym12061022.

[14] J. T. Abatzoglou and C. A. Kolden, "Relationships between climate and macroscale area burned in the western United States," Int J Wildland Fire, vol. 22, no. 7, p. 1003, 2013, doi: 10.1071/WF13019.

[15] D. A. Schmidt, A. H. Taylor, and C. N. Skinner, "The influence of fuels treatment and landscape arrangement on simulated fire behavior, Southern Cascade range, California," For Ecol Manage, vol. 255, no. 8–9, pp. 3170–3184, May 2008, doi: 10.1016/j.foreco.2008.01.023.

[16] Y. Vetrita and M. A. Cochrane, "Fire Frequency and Related Land-Use and Land-Cover Changes in Indonesia's Peatlands," Remote Sens (Basel), vol. 12, no. 1, p. 5, Dec. 2019, doi: 10.3390/rs12010005.

[17] M. R. Levi and B. T. Bestelmeyer, "Digital soil mapping for fire prediction and management in rangelands," Fire Ecology, vol. 14, no. 2, pp. 1–12, 2018, doi: 10.1186/s42408-018-0018-4.

[18] S. Testa, K. Soudani, L. Boschetti, and E. Borgogno Mondino, "MODIS-derived EVI, NDVI and WDRVI time series to estimate phenological metrics in French deciduous forests," International Journal of Applied Earth Observation and Geoinformation, vol. 64, pp. 132–144, Feb. 2018, doi: 10.1016/j.jag.2017.08.006.