Scalable & explainable ML for wildfire risk modeling in Southern Europe: A case-study in Portugal

Ophélie Meuriot[1], Jorge Soto Martin[1], Beichen Zhang[2], Francisco Camara Pereira[1], Martin Drews[1]

[1]DTU Management – Climate and Energy Policy, Division for Climate Economics and Risk Management, Technical University of Denmark
[2]Lawrence Berkeley National Laboratory, Climate and Ecosystem Sciences Division, Berkeley, United States

Motivation

Wildfires in Europe

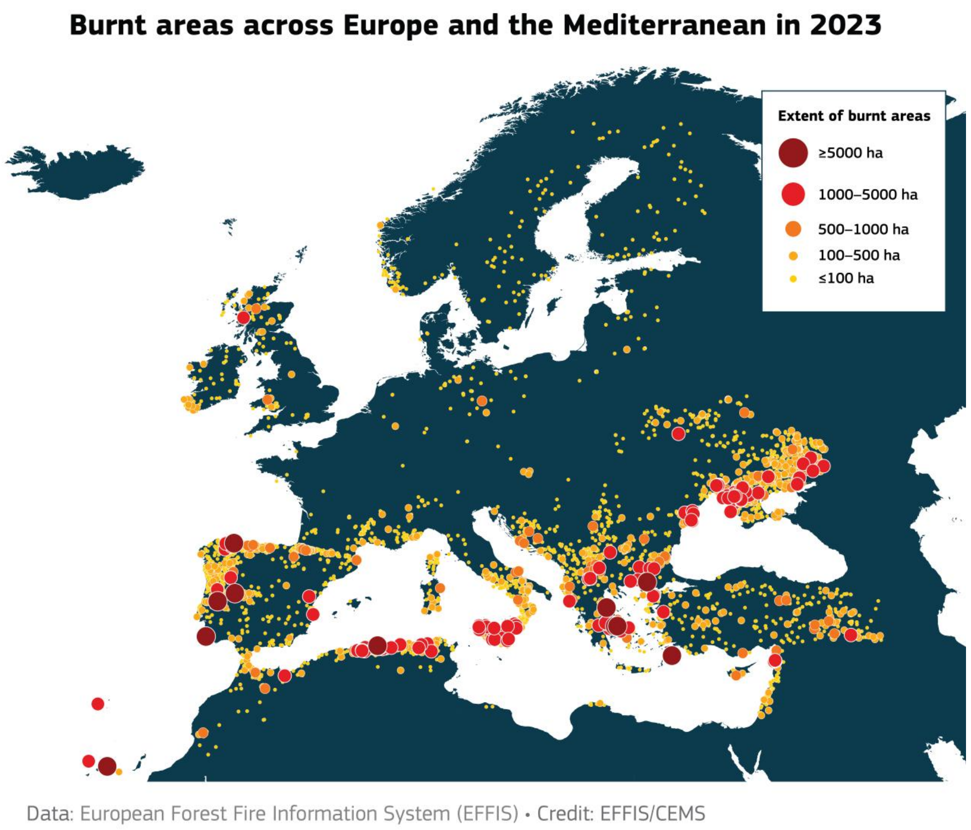
Wildfires are among Europe's most destructive natural hazards, affecting ecosystems, economies, and communities (EEA, 2024).

Fire indicators

Traditional fire indices rely mainly on weather, missing key human and landscape drivers

State of the art

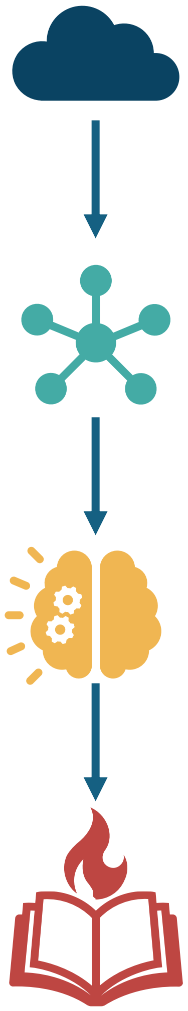
Remote sensing data enable ML-based wildfire risk assessment through data integration (Bountzouklis et al., 2023; Oliveira et al., 2012; Dorph et al., 2022).



Copernicus Climate Change Service European State of the Climate | 2023 · PROGRAMME OF THE EUROPEAN UNION · Copernicus · IMPLEMENTED BY ECMWF · Emergency Management

Distribution and extent of burnt areas across Europe and the Mediterranean in 2023. Data source: European Forest Fire Information System (EFFIS). Credit: EFFIS/CEMS.

Aims



1
Develop a data-driven wildfire risk model

- Combine climate, land use, human activity, and topography to estimate **daily wildfire probability** across Southern Europe.

2
Interpret the model with explainable AI

- Assess how different predictors contribute to wildfire risk.

3
Model scalability

- Test model scalability by applying it to an unseen region and evaluate its predictive performance and explainability.

Methodology

Input

- Historical Wildfires
- Weather Data
- Land Cover
- Topography
- Human Activity

EFFIS
- Fire season (Jun – Oct)
- 50 ha - 2008 - 2023
- 50% fire – 50% no fire

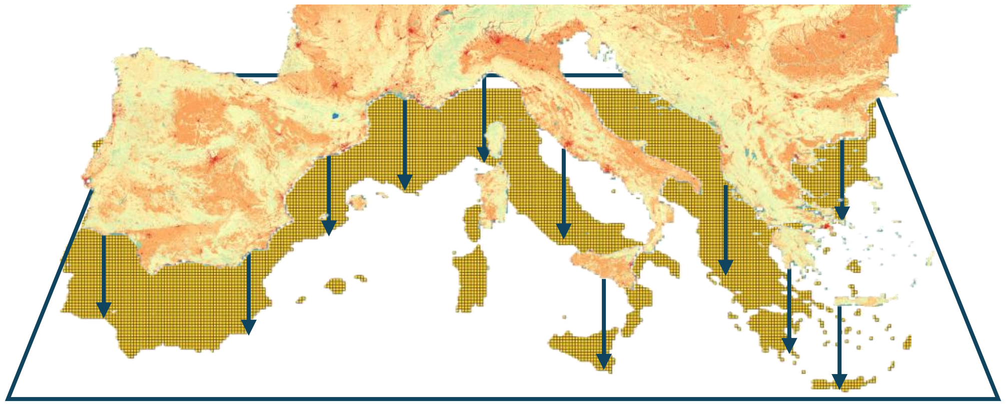
ERA5-Land
- Temperature (1d, 7d)
- Wind speed
- Relative humidity
- Precipitation (1d, 7d, 28d, 84d)

CORINE
- Forest (broadleaf & evergreen),
- Shrubland, grassland,
- cropland,
- Urban, water

EU DEM
- Elevation
- Slope
- Aspect

Global Inventory
- Road density
- Power line density
- Population density

→ Data aggregated to a common 12 x 12 km grid



Methodology

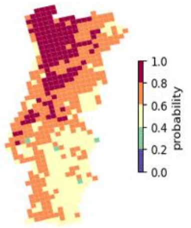
Input

- Historical Wildfires
- Weather Data
- Land Cover
- Topography
- Human Activity

Model selection & explainability

- 4 classification models
- Hyperparameter tuning on **training** data using a 5-fold CV & final model selection on **testing** data
- SHAP used for **model explainability**

Model scalability



- Best performing model is run in Portugal (**unseen region**)
- Focus on the Oct. 2017 fires during which 51 causalities were recorded
- Key predictors of the event identified using SHAP

Model selection

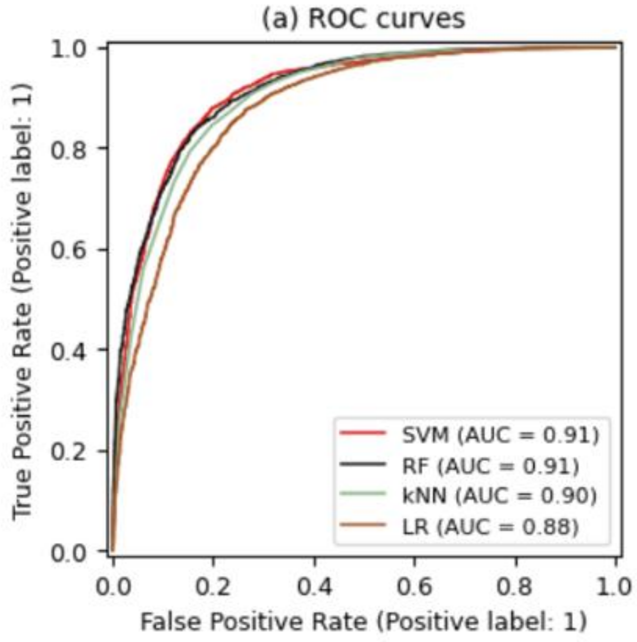
Models trained

- Support Vector Machine (SVM), Random Forest (RF), K-nearest neighbors (kNN) and Logistic Regression (LR)

Model performance (test data)

- SVM and RF both with F1 = 0.91, AUC = 0.91

➢ RF model selected for its scalability and computational cost .



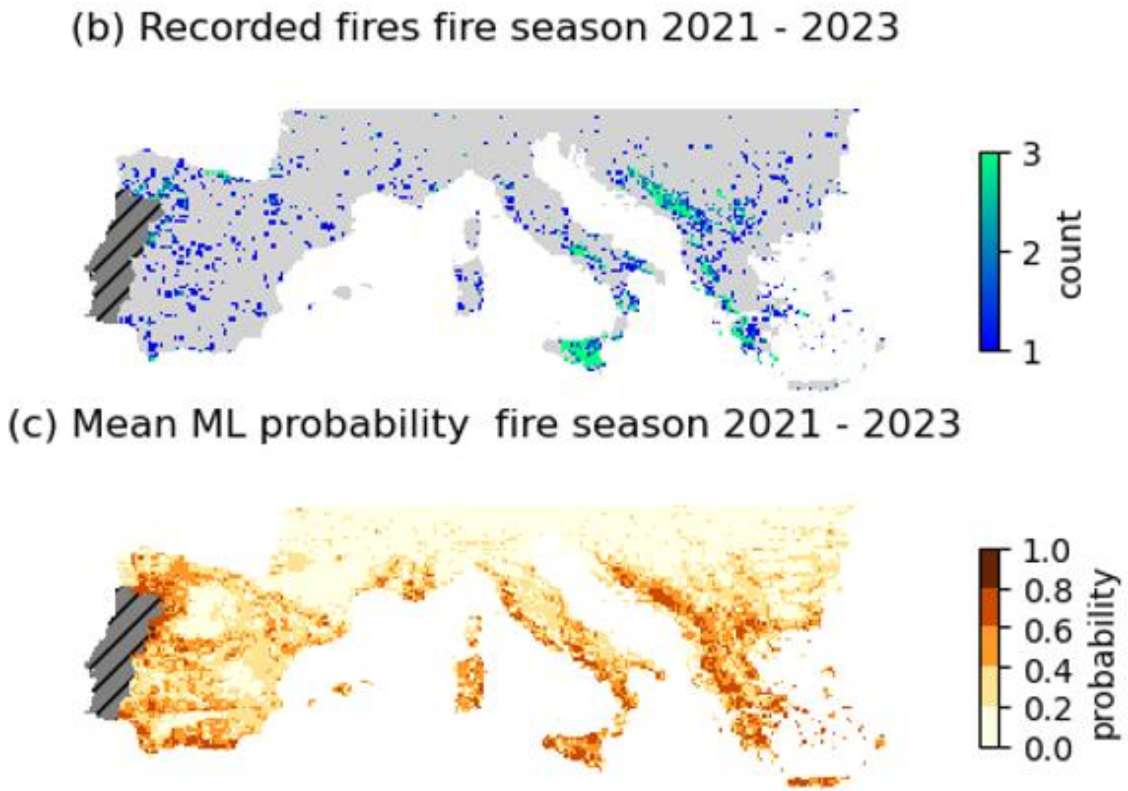
Model	F1 score	Log loss
SVM	0.91	0.38
RF	0.91	0.38
kNN	0.90	0.72
LR	0.88	0.44

Model performance in Europe – test years

RF model evaluation over 2021 - 2023

- Model run daily with the classification probability used to assess wildfire risk.
- Good agreement between historical record of fires and mean wildfire probability.

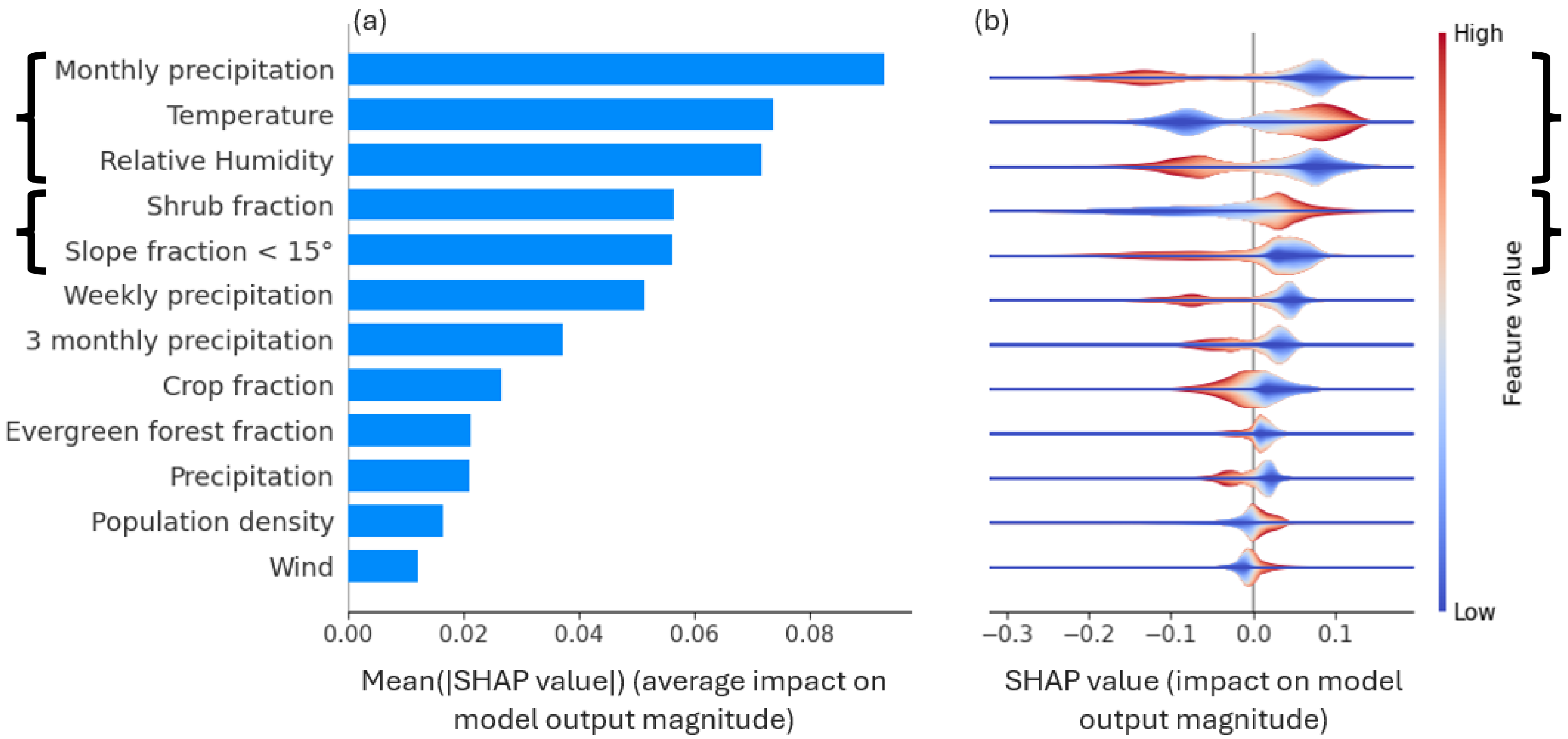
➢ RF model can successfully represent high wildfire risk areas



Model explainability: top 12 predictors

SHAP values from the testing data (2021 – 2023)

- SHapley Additive exPlanations (SHAP) method can be used to assess the contribution of each feature to an individual prediction (Lundberg and Lee, 2017).



High model output

Low: Precipitation, Relative humidity, Slope fraction < 15°

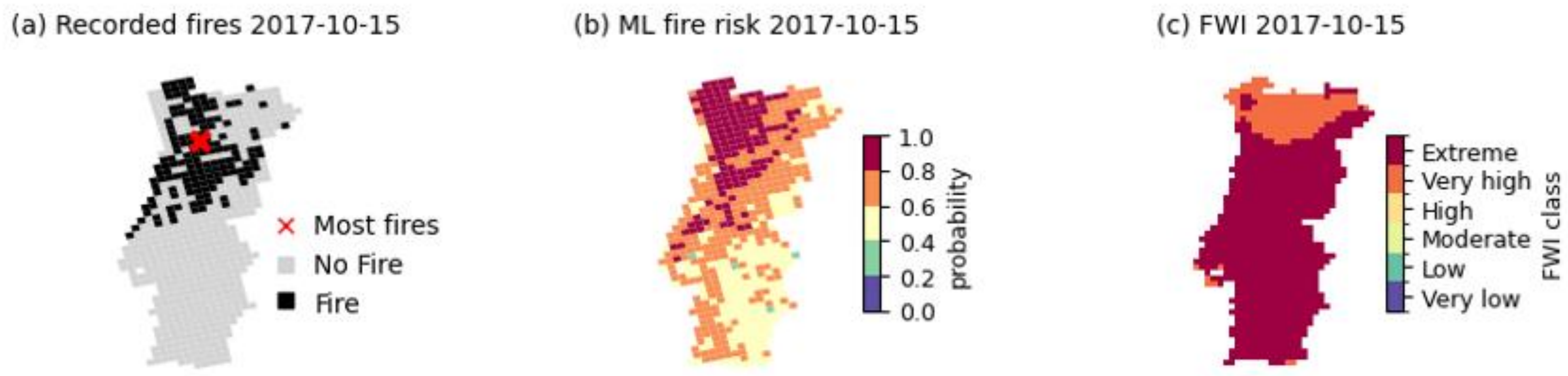
High: Temperature, Shrub fraction

→ **Feature importance** → **Impact on the model's prediction**

Model scalability: A case study in Portugal

October 2017 fires: spatial assessment

- The RF model is run daily in Portugal (unseen region) and we first focus here on the day where the most fires were recorded: **15th of October 2017**

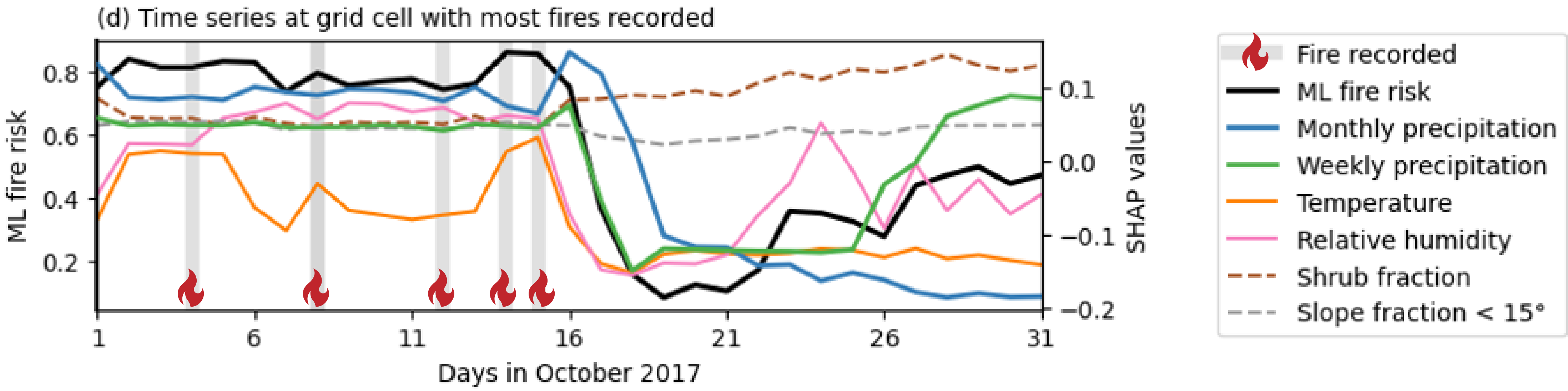


- ➢ Close alignment between the RF model **high probability** (0.8-1.0) and the **recorded fires** on the same day
- ➢ The **Fire Weather Index** is, however**, not able to capture** the areas where wildfires were recorded

Model scalability: A case study in Portugal

October 2017 fires: temporal assessment & explainability

- We now focus on the location where the most fires were recorded and assess the temporal variability of the model output as well as the SHAP values for the top predictors

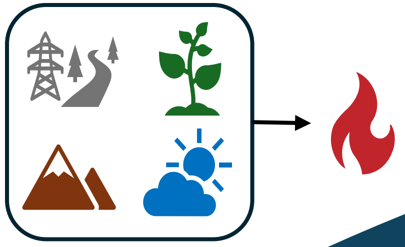


➢ ML fire risk **high when fires were recorded** and drops in the second half of the month.

➢ Drop in fire risk associated with drop in the SHAP values of the **weather variables**.

➢ **Shrub and slope fraction** contribute to high fire risk

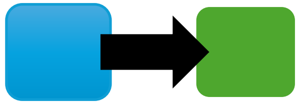
Conclusion & next steps



Suitability of data driven models to represent wildfire risk in Southern Europe

SHAP

Explainability of ML model used to identify influential predictors for wildfire risk



Scalability of model demonstrated by running model in unseen region

Pathways to Impact

- Light-weight and adaptable ML model
- Identification of key wildfires drivers
- Support for policy & land management decisions

Next steps

- Extend analysis to other regions
- Future projections of wildfire risk

Any questions ?
ophme@dtu.dk

References

C. Arderne, C. Zorn, C. Nicolas, and E. E. Koks. Predictive mapping of the global power system using open data. Scientific Data, 7(1):19, 2020. ISSN 2052-4463. doi: 10.1038/s41597-019-0347-4. URL https://doi.org/10.1038/s41597-019-0347-4.

C. Bountzouklis, D. M. Fox, and E. Di Bernardino. Predicting wildfire ignition causes in southern France using explainable artificial intelligence (xai) methods. Environmental Research Letters, 18,4 2023. ISSN 17489326. doi: 10.1088/1748-9326/acc8ee.

A. Dorph, E. Marshall, K. A. Parkins, and T. D. Penman. Modelling ignition probability for human-and lightning-caused wildfires in Victoria, Australia. Natural Hazards and Earth System Sciences,22:3487–3499, 10 2022. ISSN 16849981. doi: 10.5194/nhess-22-3487-2022.

European Environment Agency. CORINE Land Cover 2018 (raster 100 m), Europe, 6-yearly. Copernicus Programme, European Environment Agency, 2020. doi: 10.2909/960998c1-1870-4e82-8051-6485205ebbac.

Forest Fire Emergency Copernicus. European Forest Fire Information System (EFFIS). https://forest-fire.emergency.copernicus.eu/, 2025. Accessed: 2025-08-06

GISCO. EU-DEM: Digital Elevation Model. Copernicus Programme, European Environment Agency, 2016. URL https://ec.europa.eu/eurostat/web/gisco/geodata/digital-elevation-model/eu-dem.

Copernicus Programme, European Environment Agency, 2016. URL https://ec.europa.eu/eurostat/web/gisco/geodata/digital-elevation-model/eu-dem.

S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30, 2017.

J. R. Meijer, M. A. J. Huijbregts, K. C. G. J. Schotten, and A. M. Schipper. Global patterns of current and future road infrastructure. Environmental Research Letters, 13(6):064006, 2018. doi: 10.1088/1748-9326/aabd42. URL https://doi.org/10.1088/1748-9326/aabd42.

J. Muñoz Sabater. ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019. doi: 10.24381/cds.e2161bac

S. Oliveira, F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. Pereira. Modeling spatial patterns of fire occurrence in mediterranean Europe using multiple regression and random forest. Forest Ecology and Management, 275:117–129, 7 2012. ISSN 03781127. doi: 10.1016/j.foreco.2012.03.003.

A. J. Tatem. Worldpop, open data for spatial demography, 2017. ISSN 2052-4463. URL https: //doi.org/10.1038/sdata.2017.4.