Implementation of a Random Forest classifier to examine wildfire predictive modeling in Greece using diachronically collected fire occurrence and fire mapping data

Authors: Alex Apostolakis, Stella Girtsou, Charalampos Kontoes, Ioannis Papoutsis, Michalis Tsoutsos
e-mails: {alex.apostolakis, sgirtsou, kontoes, ipapoutsis, mtsoutsos}@noa.gr

Beyond center of excelence, IAASARS, National Observatory of Athens, Greece

Presenter: Alex Apostolakis

Recent Forest fire disasters

- 2019 Bush fires in New South Wales of Australia burned about 1.65 million hectares
- 2019 Brazilian Space Agency has reported an 83% increase in fire occurrences compared to the same period of the previous year.
- 2018 Attica wildfires spread up to a speed of 124 km/h resulting to more than a hundred casualties.

Climate change impact

The problem of the wildfires becomes considerably important if we account for the climate change scenarios which suggest substantial warming and increase of heat waves, drought and dry spell events across the entire Mediterranean in the future years.

Model categories for fire risk prediction

- **Theoretical (or physics-based) :** Theoretical models are entirely based on equations that describe the physics of the related to the fire ignition physical phenomena like fluid mechanics, combustion and heat transfer

- **Data-driven models :** Data-driven models (also known as empirical models before the data science developments) are purely based on the correlations between data extracted from historical fire records and their related parameters. Machine learning models belong to this category.

Why Machine Learning models

- Machine Learning algorithms are designed to automatically formulate the complex mathematical relations between the input parameters. In Physical-based models the mathematics of those relations should be known in advance.

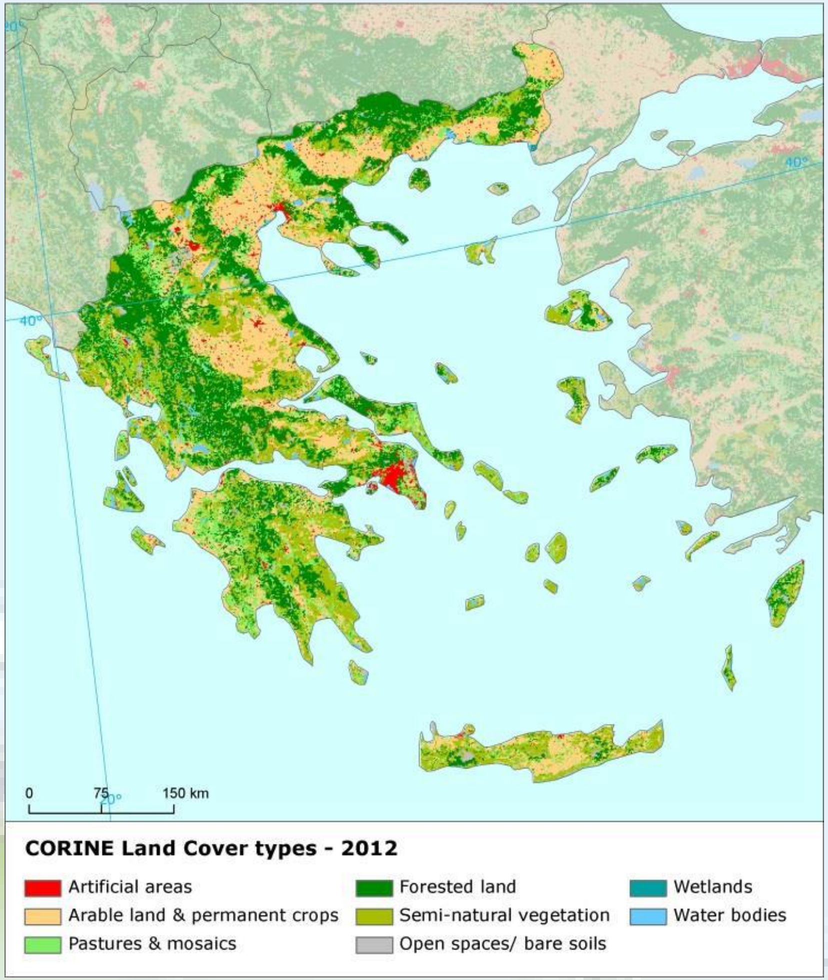
- With Machine Learning it is relatively easy to add or suppress input parameters, and thus select combinations of parameter sets that works best for the model prediction.

Study Area

- Greece's territory (131.957 square kilometers) located in the southeast of the Mediterranean climatic zone, with mild and rainy winters, warm and dry summers and extended periods of sunshine.

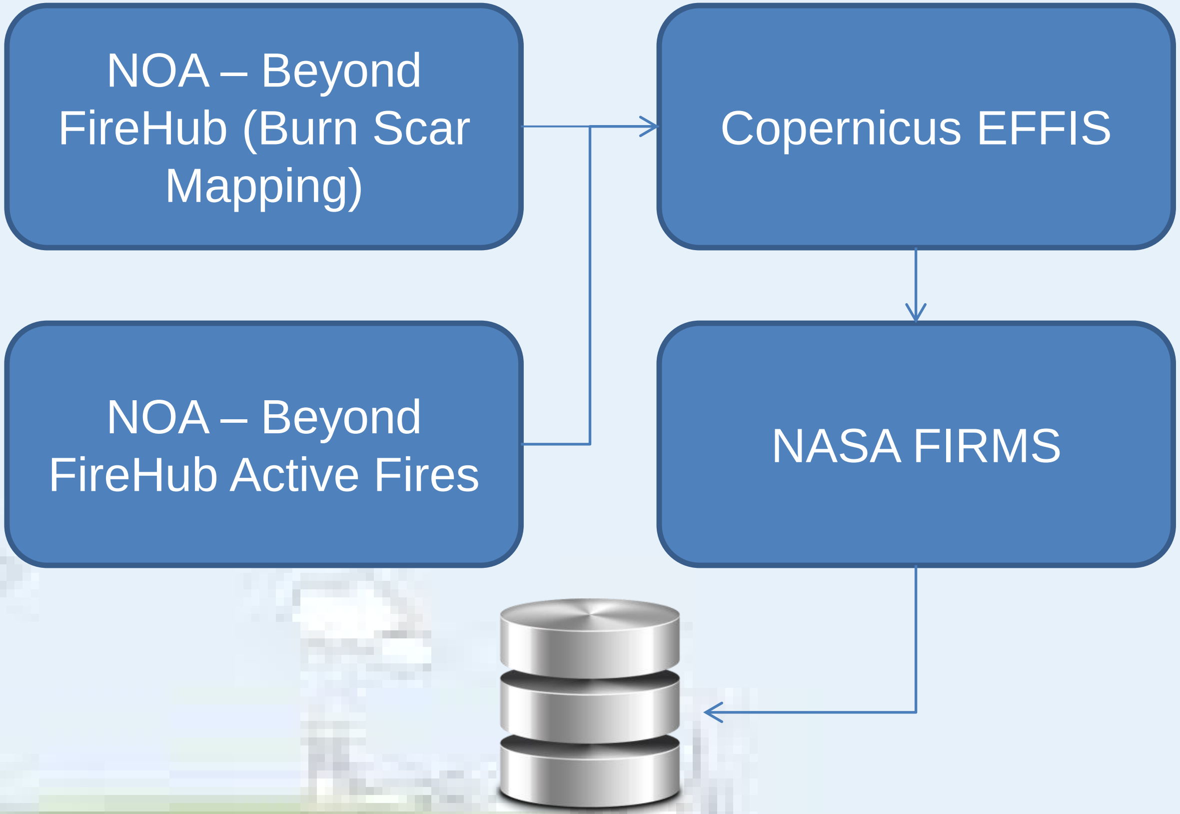
- 58.8% of the total surface, represents low altitude areas (0-500m) which are prone to fire ignition The topography and the dominant north winds in combination with the vegetation types in the central and southern parts of Greece are between the prime drivers for fire ignition during the summer period.

- Vegetation cover makes Greece particularly prone to fire hazard and fire risk as coniferous and mixed forests, sclerophyllous vegetation, natural grasslands, transitional woodlands, semi natural and pasture areas correspond to approximately 72% of the total surface of the country.



Forest fire inventory

- ✓ An exhaustive forest fire inventory of fire occurrences and burn scar maps was compiled by exploiting diachronic data generated by the FireHub system of BEYOND (Active Fires – AF and Burn Scar Mapping - BSM), the NASA FIRMS and the European Forest Fire Information System (EFFIS/JRC).
- ✓ Every grid cell intersecting with a BSM polygon from FireHub and the corresponding AF detections had been labeled as fire cell. In a following step any remaining AF evidences were checked spatially and temporally against the EFFIS and FIRMS datasets.
- ✓ This process has returned a set of about 12500 “fire” cells
- ✓ An equivalent dataset of “non-fire” cells was generated through a simple random selection spatially expanding over the entire Greece.

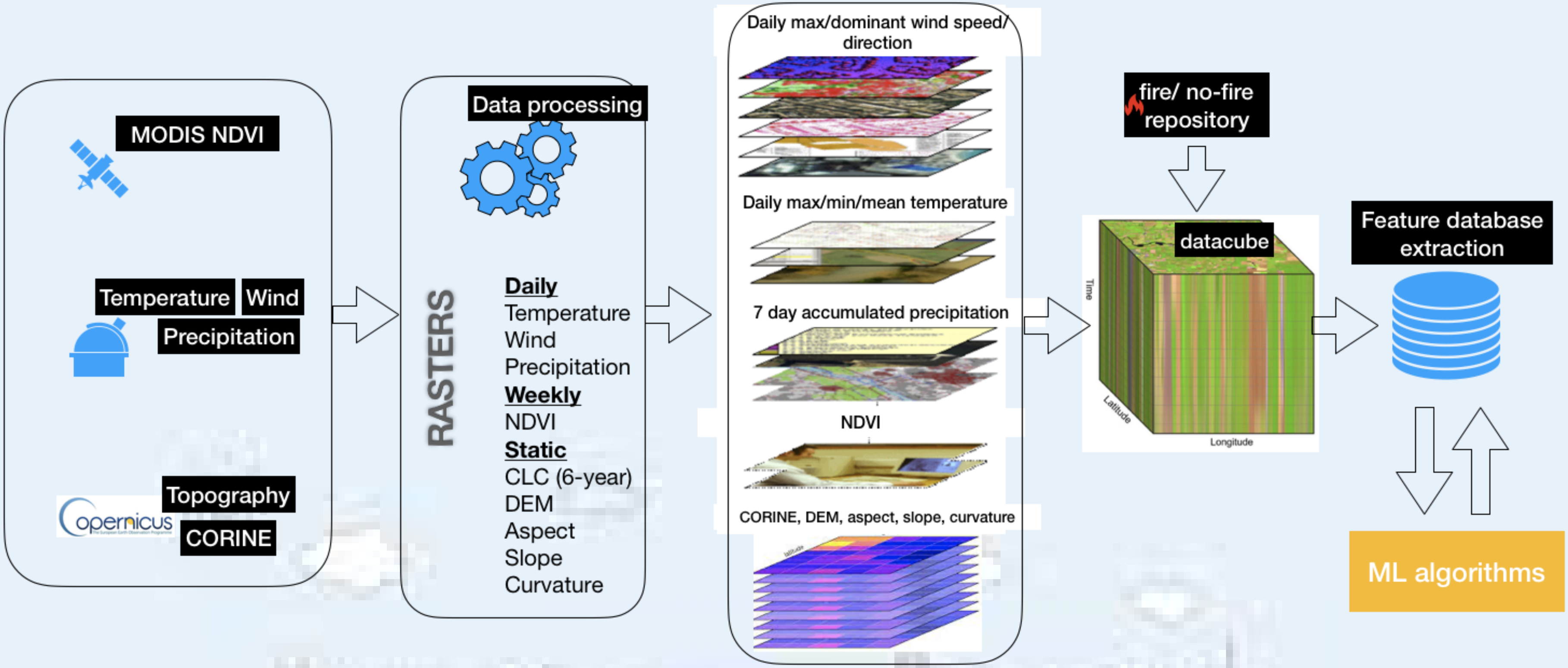


Fire Inventory 2010-2018 at 500m grid resolution for ML training

Features used in model training

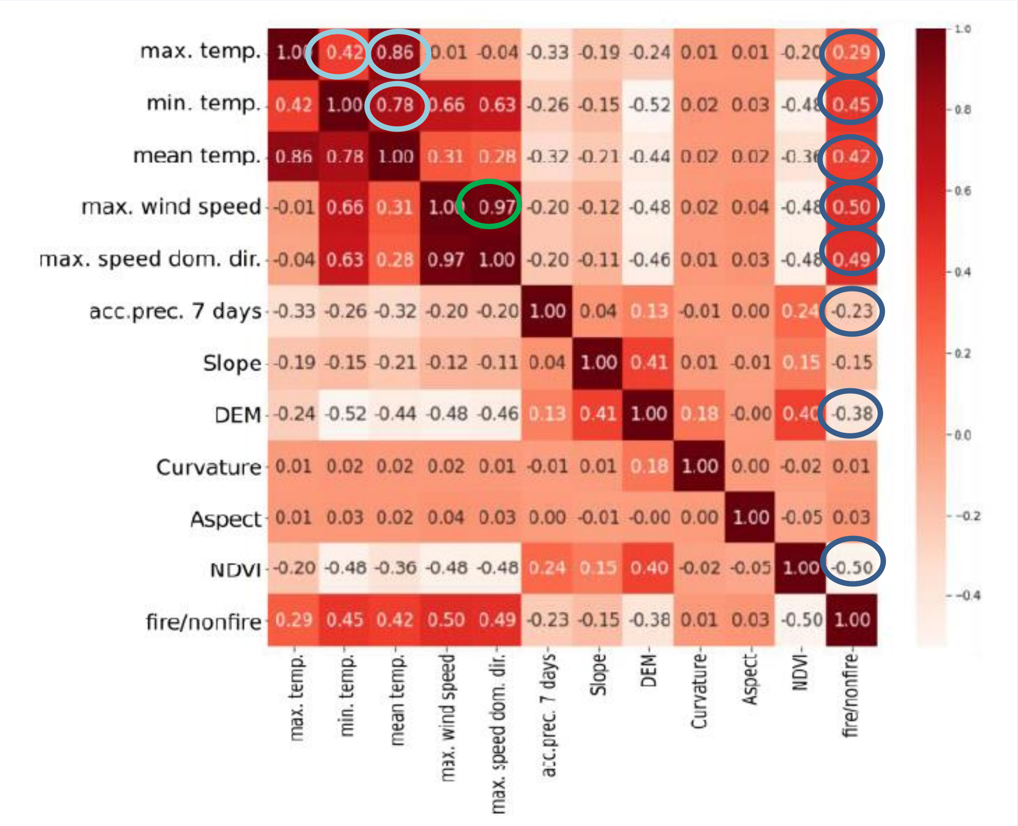
Product	Source	Spatial resolution	Temporal resolution	Features
Wind (u-comp, v-comp)	ERA5-Land	9 km	hourly	Dominant dir., Max speed of the dom. Dir., Max wind speed, Wind dir of the max speed
Temperature	ERA5- Land	9 km	hourly	Max temperature, Min temperature, Mean temperature
Precipitation	ERA5- Land	9 km	hourly	7-day accumulated precipitation
NDVI	NASA -MODIS	500 m	8-day	NDVI
DEM	Copernicus	25 m	Static	DEM, slope, aspect, curvature
Corine	Copernicus	100 m	6-year	CLC
“fire”/”non-fire” cells	NOA, EFFIS, NASA	500 m	Daily	

System's architecture



Feature Ranking with Spearman's correlation

- ✓ A correlation method based on Spearman's rank correlation coefficient has been used for evaluating the dependencies of the various feature combinations. Spearman's correlation can detect linear and non-linear monotonic relations.
- ✓ Not surprisingly the observations on this heat map indicate that the wind speed, the temperature, the NDVI and the elevation are ranked as the more influencing features for "fire"/"non fire" class prediction.
- ✓ The NDVI, the DEM elevation and the accumulated rainfall have inverse correlations with the dependent variable ("fire" / "non-fire")
- ✓ Finally the two wind speed and the three temperature features are highly correlated one another.



Feature Ranking with SFS, RF Impurity and permutation importance

Ranking	SFS AUC	RF Impurity	Permutation importance
1	NDVI	max. speed of the dom. dir.	NDVI
2	maximum wind speed	NDVI	max.temp.
3	max.temp.	max.temp.	maximum wind speed
4	CLC	maximum wind speed	mean temp.
5	acc.prec. of past 7 days	min. temp.	max. speed of the dom. dir.
6	DEM	mean temp.	DEM
7	Aspect	acc.prec. of past 7 days	acc.prec. of past 7 days
8	Curvature	DEM	wind dir. of the max. speed
9	Slope	wind dir. of the max.speed	dominant wind direction
10	min. temp.	dominant wind direction	CLC
11	max. speed of the dom. dir.	CLC	Aspect
12	wind dir. of the max.speed	Aspect	Curvature
13	mean temp.	Slope	min. temp.
14	dominant wind direction	Curvature	Slope

NDVI, wind speed and temperature features are ranked in the first three places.

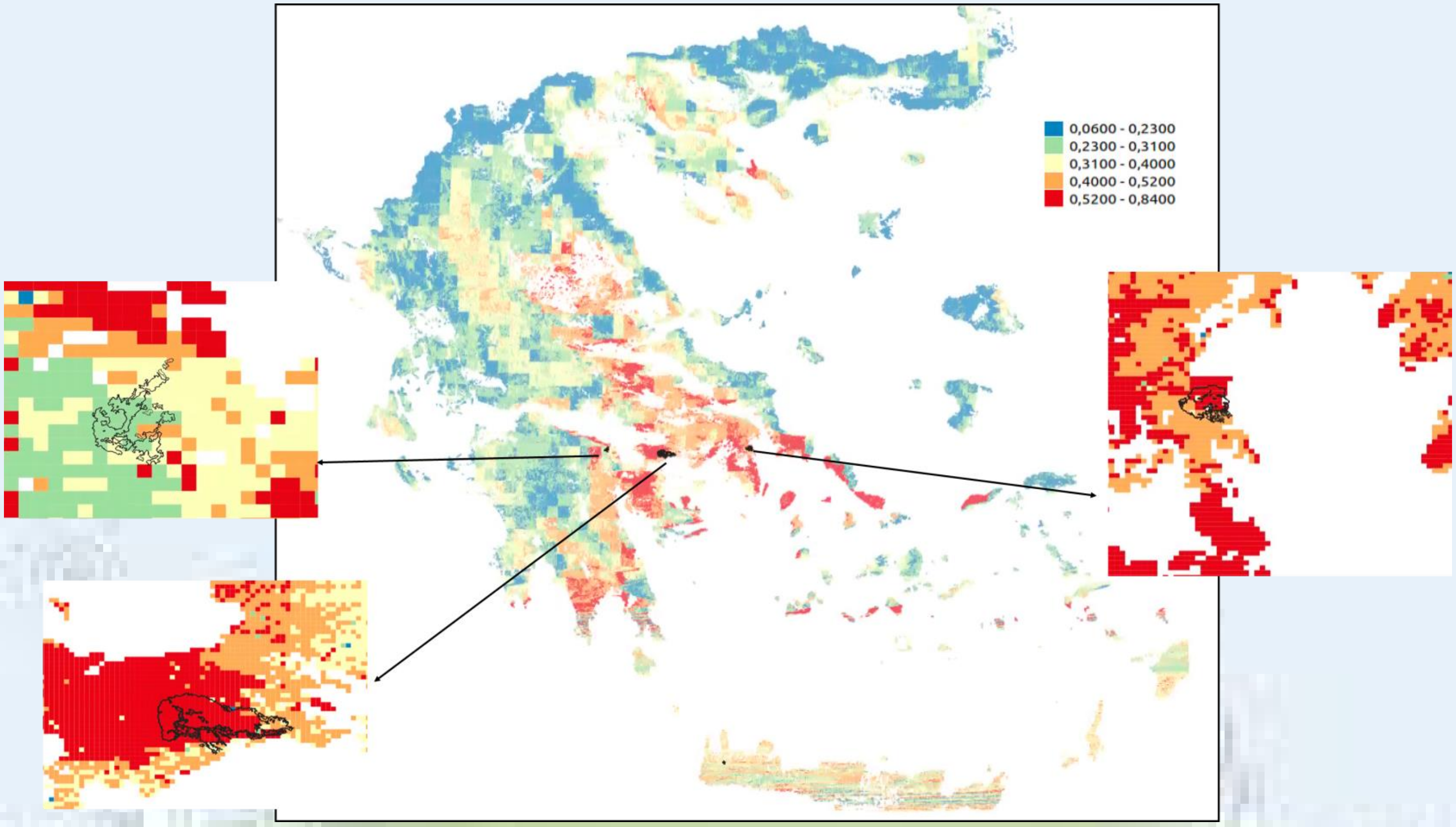
Random Forest Results

RF on shuffled feature set split 90% for training and 10% for validation achieved scores for the Recall of “fire” class above 92%

Training/Test set	Class	Precision	Recall	F1-Score
Training/Test with shuf-fling	Fire	0.94	$\boxed{0.92}$	0.93
	Non fire	0.92	0.94	0.93
k-fold cross validation (folds include entire days)	Fire	0.75	$\boxed{0.77}$	0.76
	Non fire	0.77	0.75	0.76

The best models of the 10-fold cross validation scheme under the rule that the cell samples of a specific day cannot be distributed in more than one fold achieved a mean Recall score of above 77%

Random Forest prediction on entire day



Prediction on test sets of entire day. Black polygons show the actual fires burn scar mapping

Implementation of a Random Forest classifier to examine wildfire predictive modeling in Greece using diachronically collected fire occurrence and fire mapping data

Thank you for watching!

Questions?