EAI Simutools 2023

Wildfire Risk Mapping based on Multi-Source Data and Machine Learning

Work completed at the UMR CNRS SPE 6134 , University of Corsica

Ghinevra Comiti, Paul-Antoine Bisgambiglia, Paul Bisgambiglia

The GOLIAT project

Group of Tools for Fire Fighting and Regional Planning.

Multidisciplinary project (biology, physics, computer sciences, economics, history, etc.).

Put scientific knowledge at the service of field professionals (fire fighters, foresters, etc.), engage in mediation with the population.

Office National des Forêts

Università di Corsica Pasquale Paoli

Aix*Marseille université Initiative d'excellence

arobase.fr

Università di Corsica Pasquale Paoli | Laboratoire Sciences pour l'Environnement UMR 6134 SPE | cnrs

Cullettività di Corsica Collectivité de Corse

Our mission

Creating a forest fire risk map for Corsica.

Our map should be generated for a given day, and associate a numerical fire risk to each part of the island.

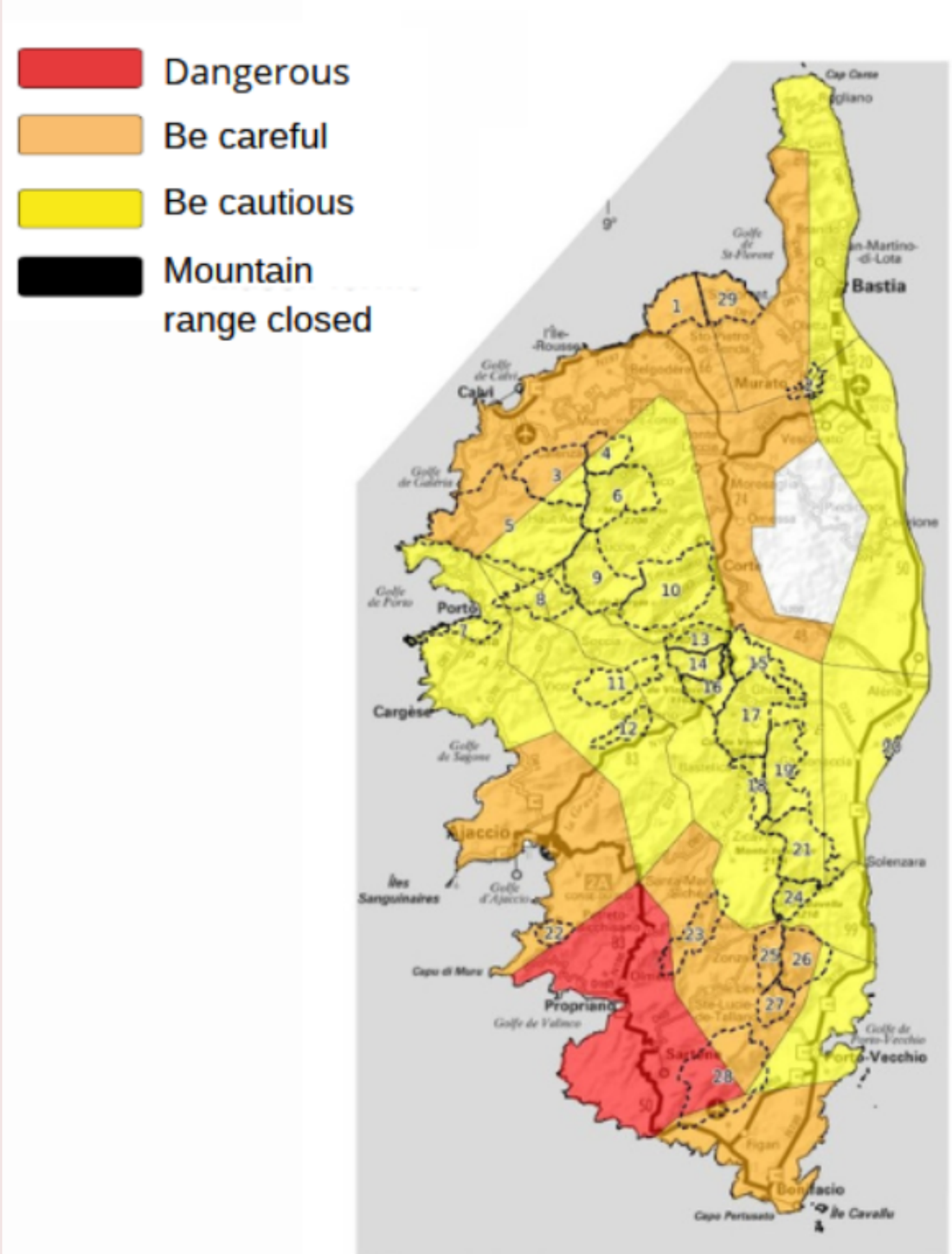
Such a map already exists, but has lots of flaws. Our goal is to improve that.

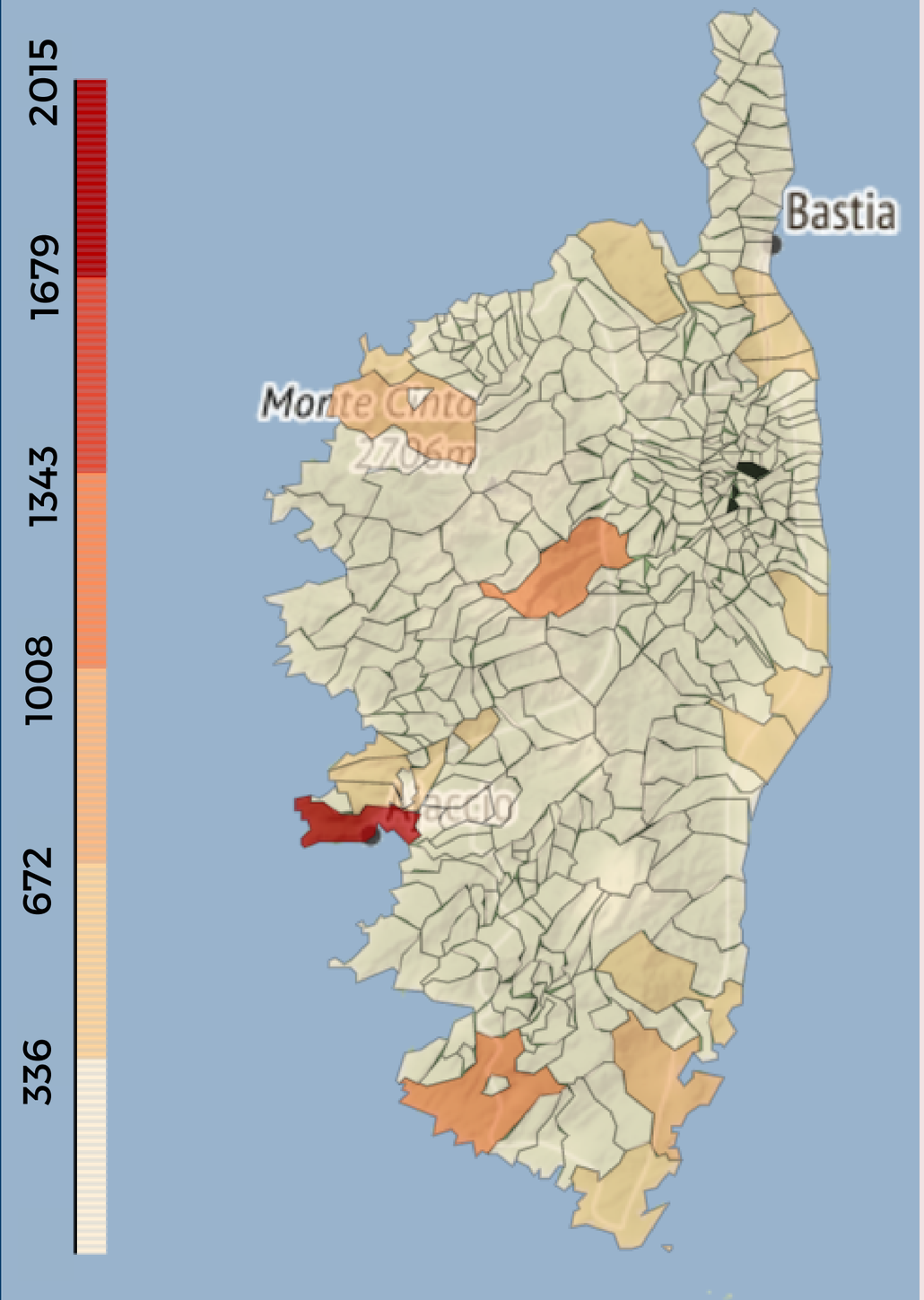
The existing map

Is published by the Corsican prefecture every day during fire season.

Isn’t very precise.

Is exclusively based on meteorological data.





Fire data

For each fire in our database, we gathered the following data :

- A unique number identifying the fire
- The date and time of the fire
- Department
- Municipality
- Area, perimeter
- Cause of the fire (when it is known)
- Type of fire (forest fire, residential fire...)

Other types of data

Meteorological data :

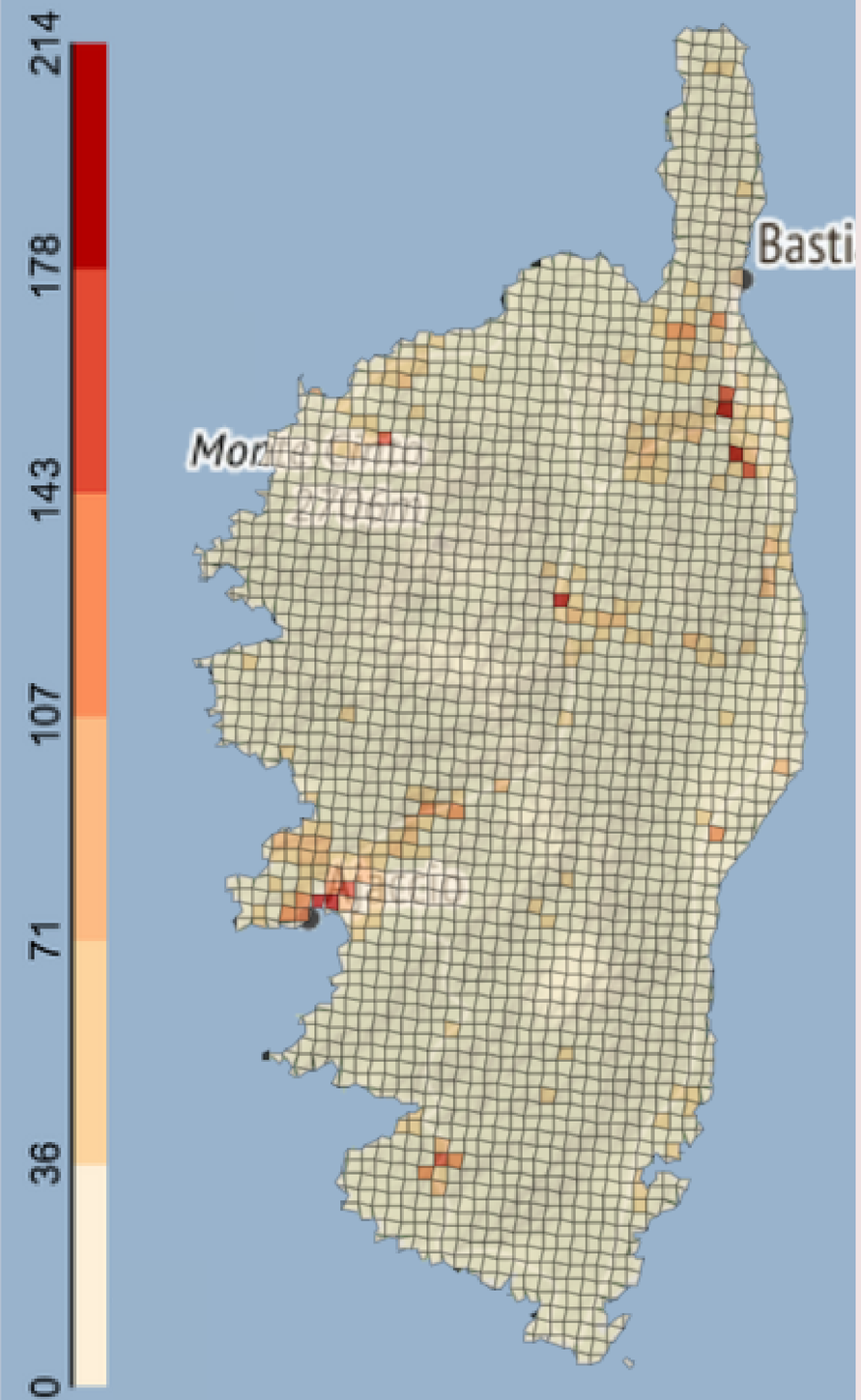
- Temperature of the day
- Wind speed
- Soil humidity

Land data :

- Land use
- Road density
- Vegetation type
- Altitude

Socioeconomic data :

- Unemployment rate
- Day of the week



Map

To generate a fire risk map, we needed to define a scale. To this end, we devided the territory into a grid. We decided to use a grid that’s already used by the operationals to locate fire, the DFCI (Defense of the forest against fires) grid, which is 2x2km.

Each fire is located on that grid.

Creating a dataset

To create a complete dataset, we must consider both the fires and the absence of fire. Each row of data consists of the data previously stated, a localisation and date, and a boolean “fire”.

We plan of training a classification model on this data, and caracterize each grid square, each day, as “fire square” or “no fire square”.

27 500 fires on our dataset, and 27 500 random “no fire” days were selected for training, to obtain a balanced dataset (results were constants for different sets of random “no fire” days)

Potential models

- Le Random Forest

Modelling temporal variation of fire-occurrence towards the dynamic prediction of human wildfire ignition danger in northeast Spain, Yago Martin et al.

- Le Boosted Regression Tree

Application of Bagging, Boosting and Stacking Ensemble and EasyEnsemble Methods for Landslide Susceptibility Mapping in the Three Gorges Reservoir Area of China, Xueling Wu et al.

- L'Adaboost

GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping, Shruti Sachdeva et al.

The chosen one : Random Forest

Better results among the tested models : AUC-ROC > 0.8.
But also, has very interesting caracteristics :

- Robust to noise and outliers
- Less prone to overfitting
- Very easy to turn predictions into probabilities, which will be useful to create a hazard map that’s not too binary

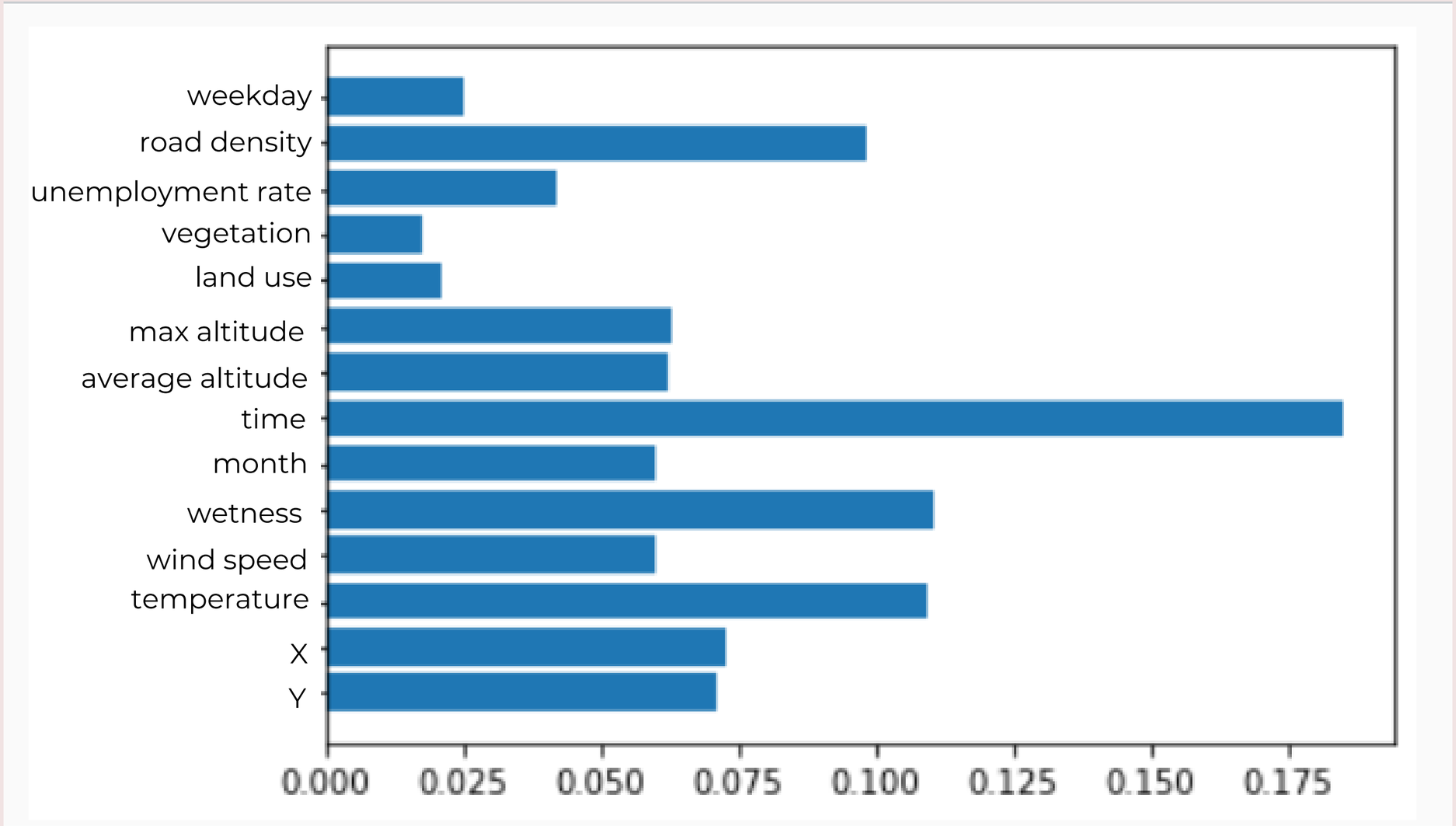
Evaluating the model

Use of the AUC-ROC Curve

Minimization of the false negatives is very important in this project, more than the potential false positives. Because of that, we want to maximize recall, which is a metric that measures the proportion of actual positive instances correctly identified, out of the total number of actual positive instances.

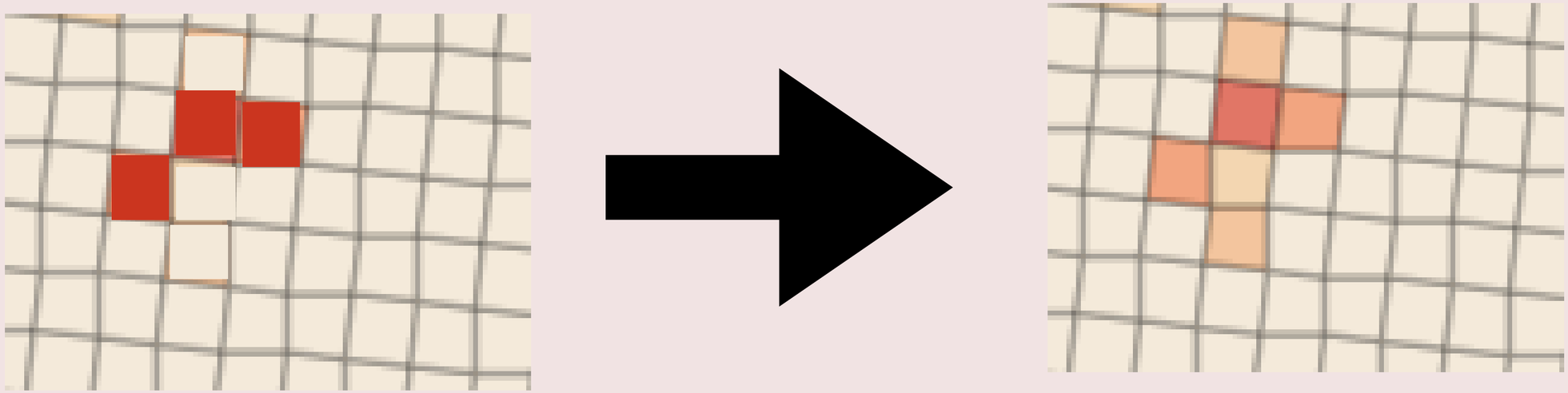
$$\text{Recall} = \frac{\text{True positive}}{\text{True positive + False negative}}$$

Features’ importance



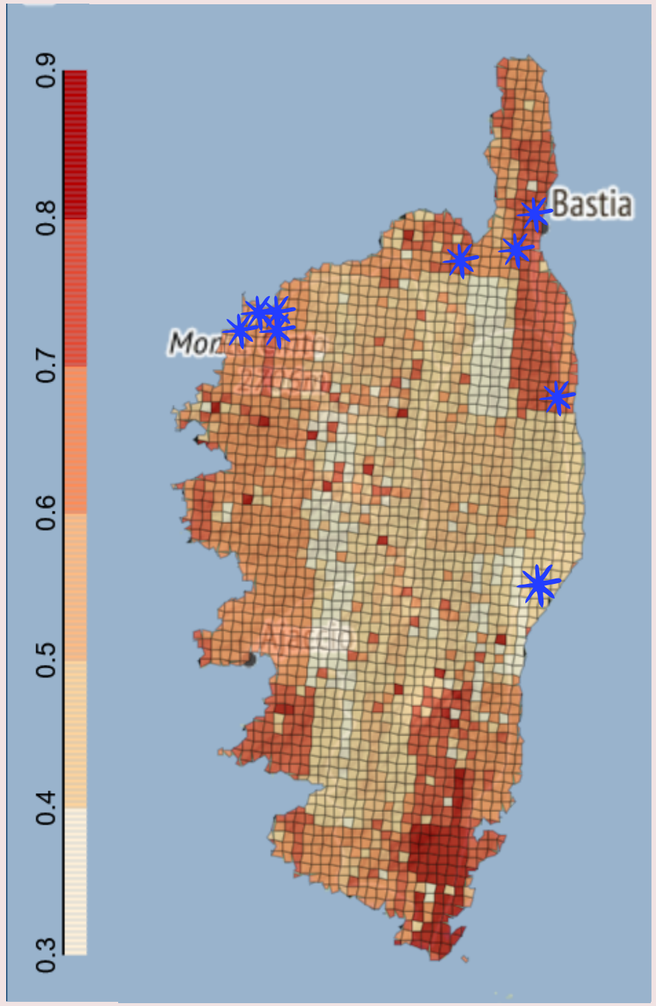
From classification to probabilities

The proposed “fire/not fire” classification is efficient but very binary. We wanted our risk map to be more nuanced. We turned this binary classification into a classification probability to create our output. It’s very easy due to the way ensemble learning models are built (vote of weak learners).



Visualization

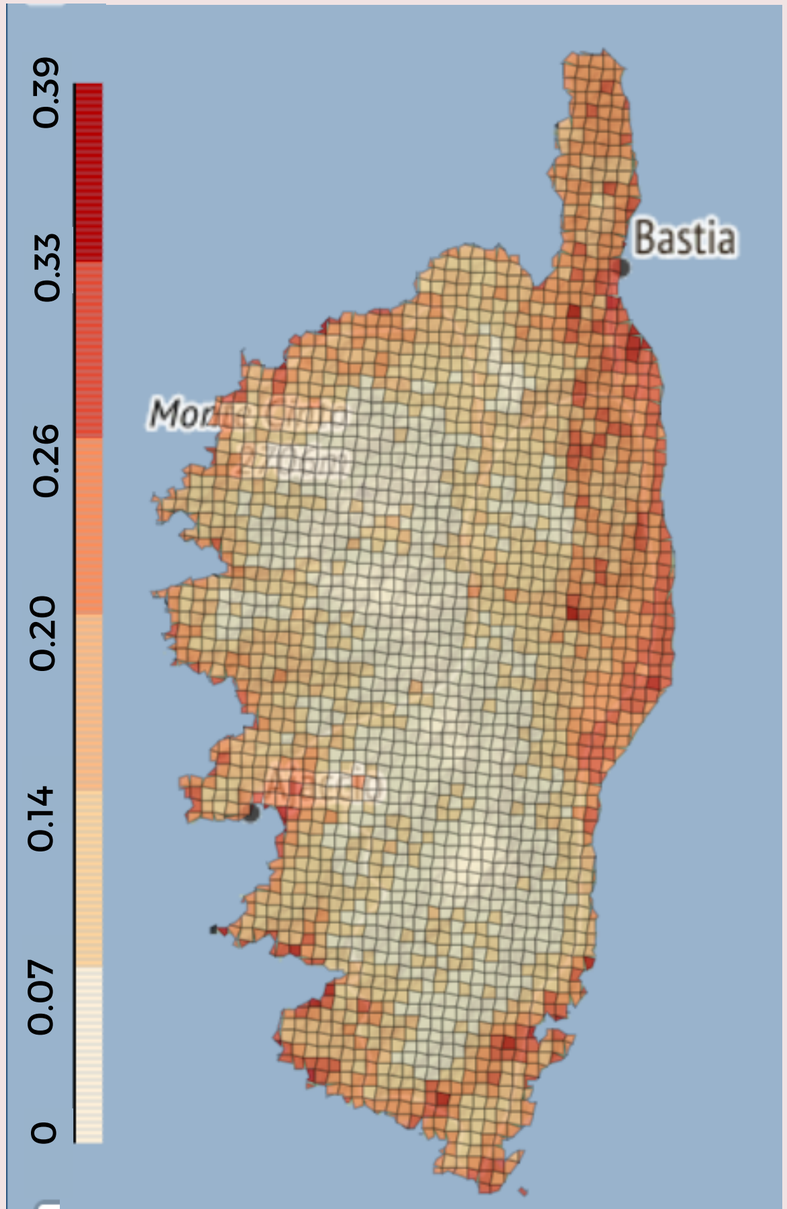
Estimated fire probability



Real fires of the day

08/07/2009	2B	Montegrosso
08/07/2009	2B	Ghisonaccia
08/07/2009	2B	Oletta
08/07/2009	2B	Calvi
08/07/2009	2B	Cervione
08/07/2009	2B	Calenzana
08/07/2009	2B	Olmi-Cappella
08/07/2009	2B	San-Martino-di-Lota
08/07/2009	2B	Olmeta-di-Tuda
08/07/2009	2B	Calvi
08/07/2009	2B	Oletta
08/07/2009	2B	Olmi-Cappella

Map comparison

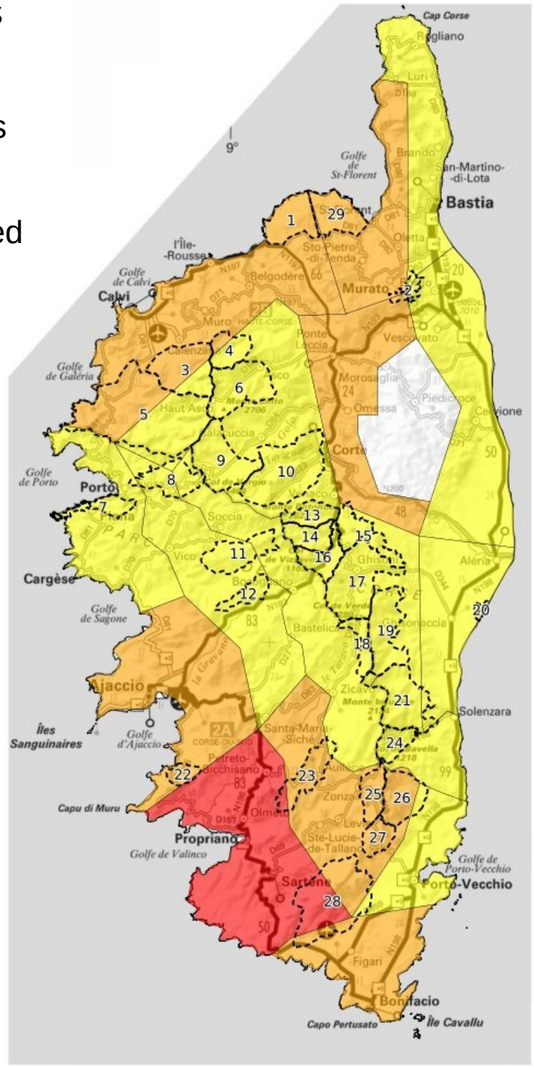


Dangerous

Be careful

Be cautious

Mountain range closed



Perspectives

This tool has a great potential to automatically generate fire susceptibility maps. Those maps can be used by field professional for territorial planning, and by the population to plan outdoor activities.

However, a some improvements are needed to make the map usable in real life. For instance, a phase of method validation and consolidation would be needed. We will also take a closer look at the current map and the predictions and accuracy differences between the two maps.