Human-Caused Wildfire Ignition Probability Estimation

M.Sc. in Stochastics and Data Science

Supervisor:
Marco Grangetto

Co-supervisor:
Andrea Bragagnolo

Candidate:
Simone Genovese

Introduction

Background and motivations

FOREST FIRES REPORT*

85–90%

of worldwide wildfire ignitions caused by human activities

22 out of 27

EU countries that experienced wildfire events in their territory

Over 5500 km²

burnt EU land of which 1000 km² are in protected areas

*statistics updated to the end of 2021. In the same year, the EU promoted the **FirEUrisk** project with the purpose of improving the wildfire management s by including the socio-economic circumstances that affect the occurrence of extreme wildfires as well as the biophysical conditions

THE COMPONENTS OF RISK

Danger

Ignition
Propagation

Exposure

Extent
Asset types

Vulnerability

Fragility
Resilience

THE COMPONENTS OF RISK

Danger

Target

Ignition

Propagation

Exposure

Extent

Asset types

Vulnerability

Fragility

Resilience

THE STRATEGY

1. Identify the study region, collect the target data and define the model output
2. Identify and collect the predictive data including time and geographical references
3. Process the data to fit the model's input requirements
4. Select the best predictive features and models
5. Evaluate model performance and interpretation

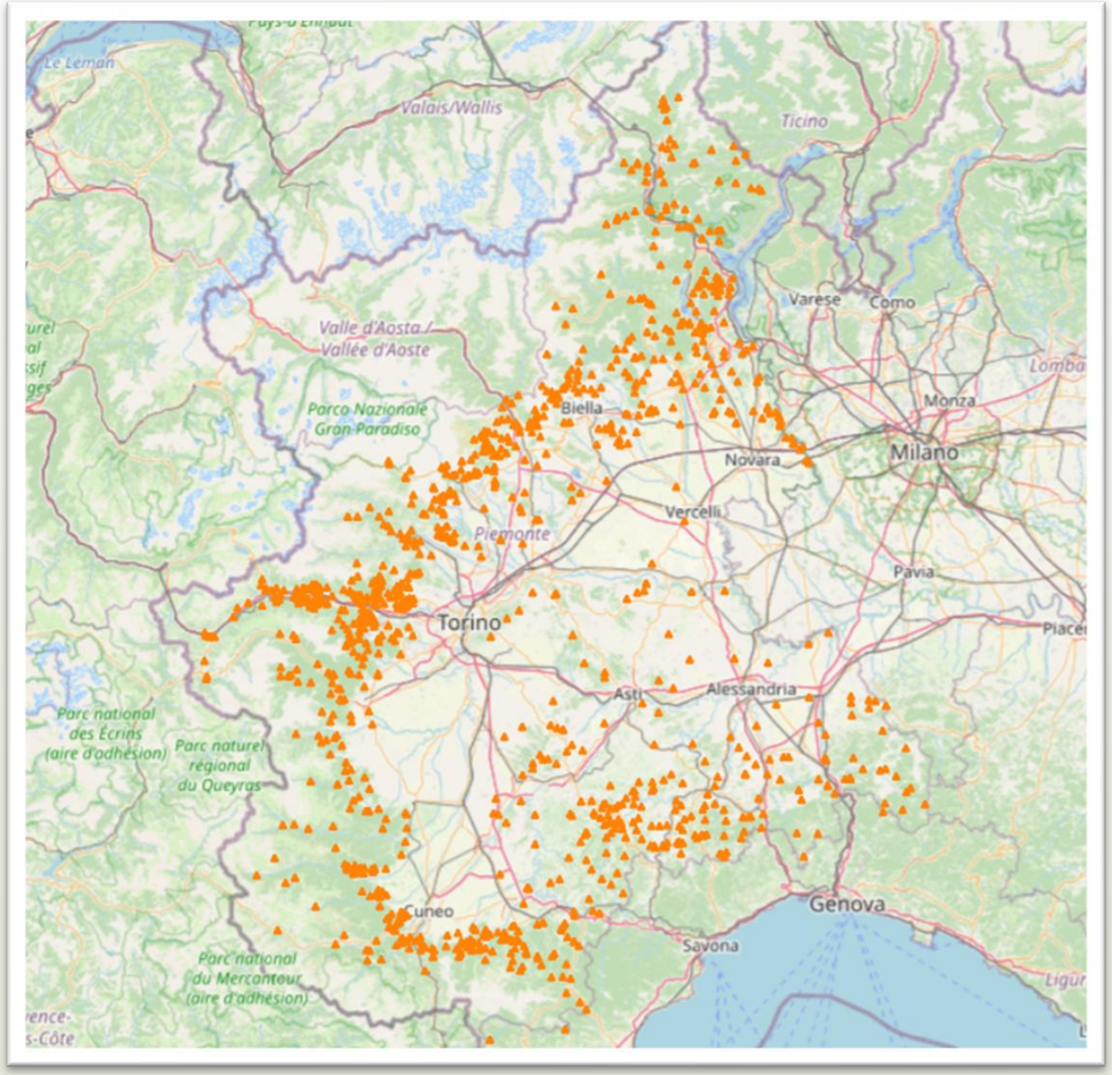
Method

From data to models

Wildfires database

Since 2016 we have 1558 reported wildfires with their space and time coordinates.

PROBLEM: how to define the pointwise ignition probability?

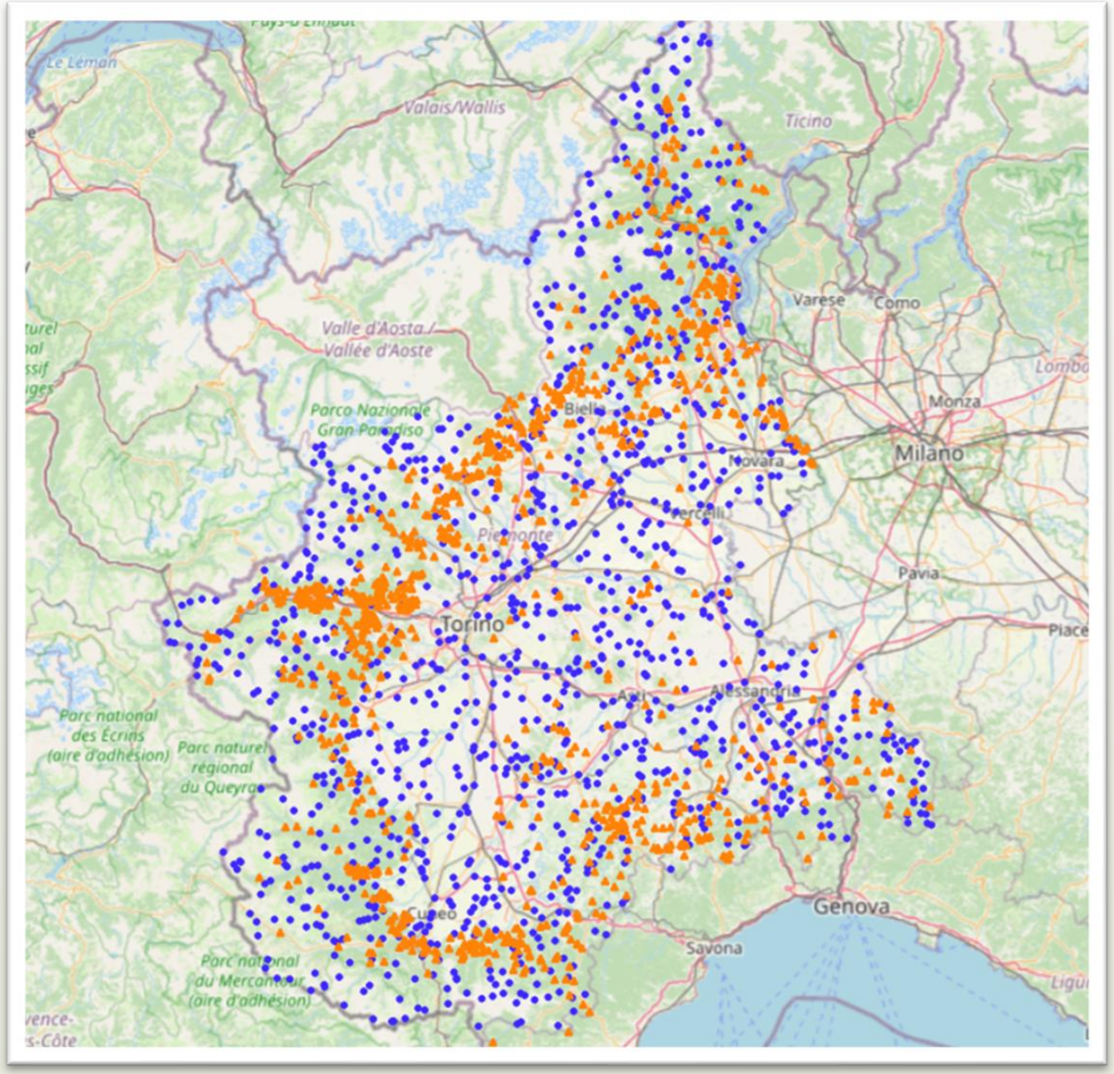


Wildfires database

Since 2016 we have 1558 reported wildfires with their space and time coordinates.

PROBLEM: how to define the pointwise ignition probability?

1. Past fires are set as *1 probability cases* ●
2. Sample an equivalent set of random points to represent the *0 probability cases* ●
3. **Interpolate** the values with a continuous function



Kriging

Let $\boldsymbol{x}_i = (x_i, y_i)$ be the n data points with target probabilities $p(\boldsymbol{x}_i)$.
Define the random variable P as the linear combination of the p_i

$$\boxed{P(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \, p(\boldsymbol{x}_i)}$$

Kriging

Let $\boldsymbol{x}_i = (x_i, y_i)$ be the n data points with target probabilities $p(\boldsymbol{x}_i)$.
Define the random variable P as the linear combination of the p_is

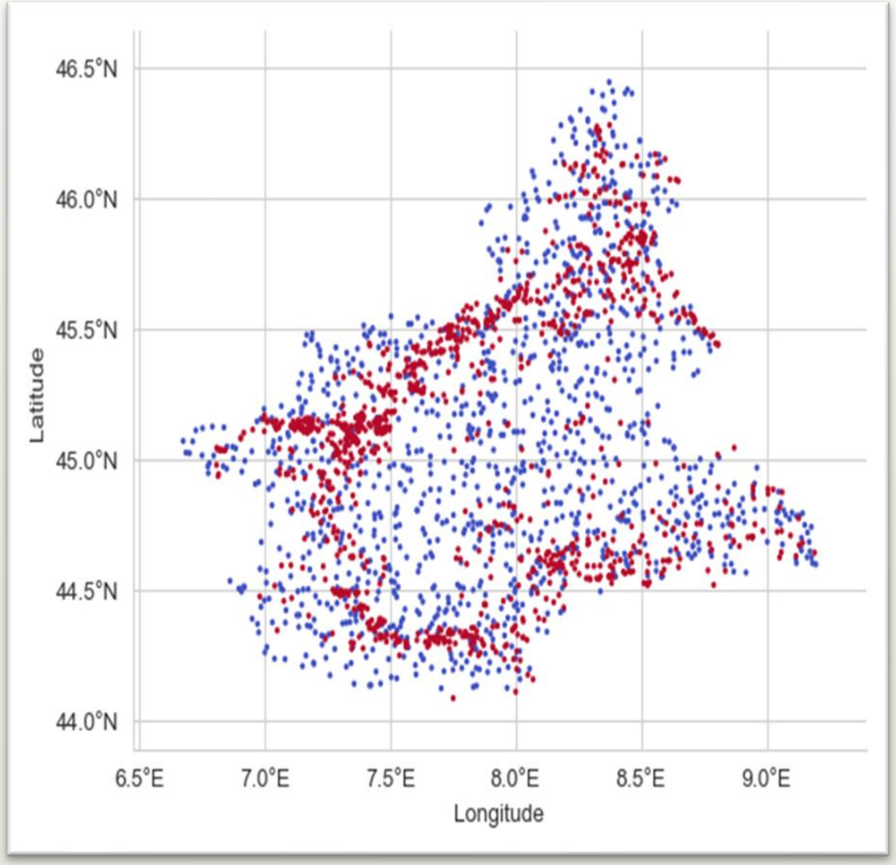
The weights $\lambda_i \in \mathbb{R}$ are defined by the constraints generated by the conditions estimator $\hat{P}$ for each point $\boldsymbol{x}_0$:

$$\boxed{P(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \, p(\boldsymbol{x}_i)}$$

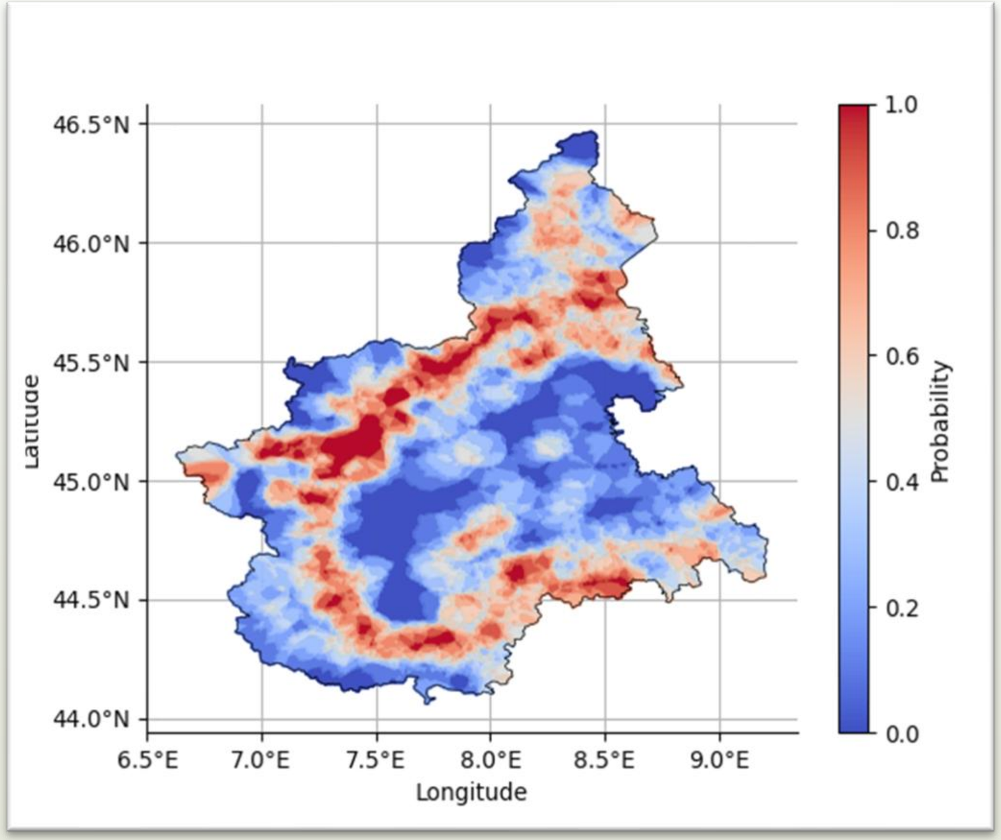
- Unbiased $\quad \mathbb{E}\left[\hat{P}_0 - p(\boldsymbol{x}_0)\right] = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \lambda_i = 1$
- Optimal $\quad \mathbb{E}\left[\left(\hat{P}_0 - p(\boldsymbol{x}_0)\right)^2\right] = -\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma\left(\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|\right) + \sum_{i=0}^{n} \lambda_i \gamma(\|\boldsymbol{x}_i - \boldsymbol{x}_0\|)$

$\gamma(\cdot)$ model of the *variogram* (spatial autocorrelation function)

INTERPOLATED PROBABILITY MAP



Kriging



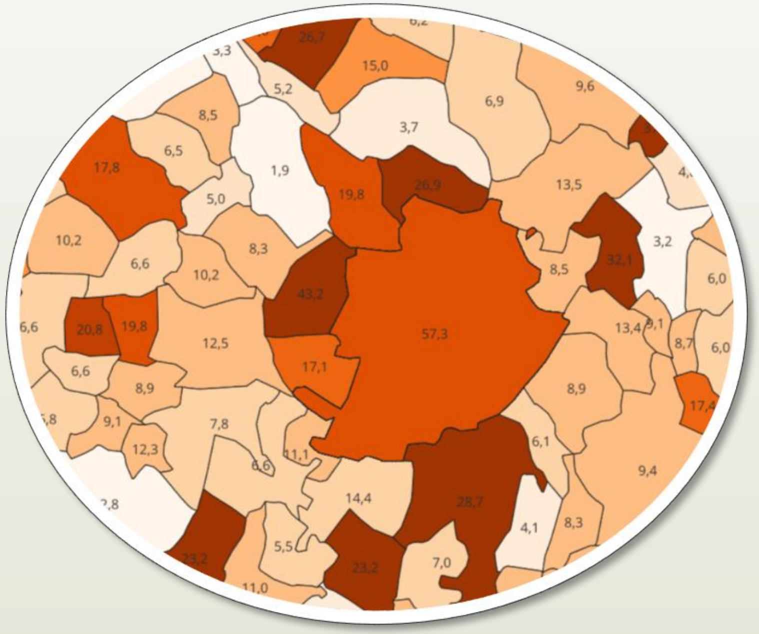
PREDICTIVE FEATURES - 1

Through the map of the **geographical boundaries** of the townships' administrative areas we can assign the variable values to each data point.

Agriculture
Education
Incomes

Employment
Weather
Roads

Fragility indexes
Flora
Demographic data

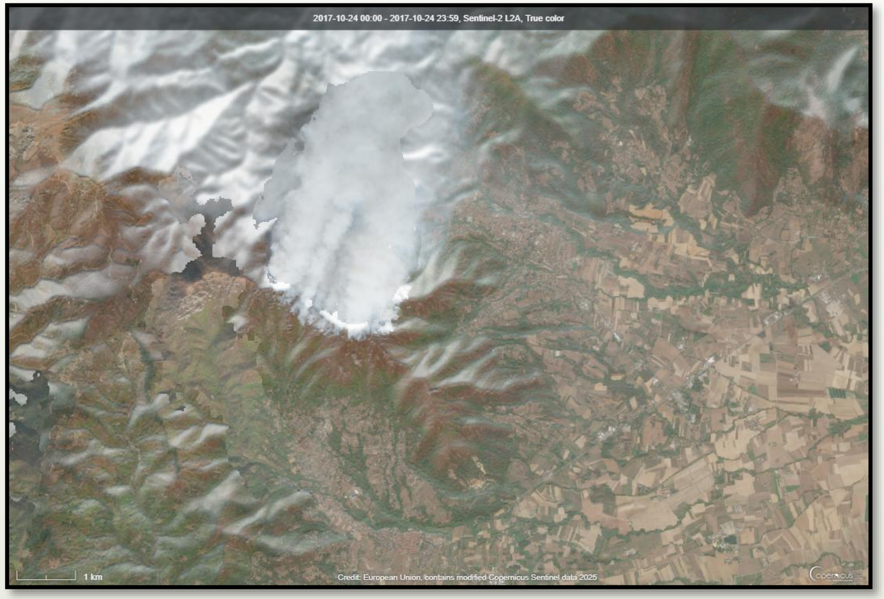


Land consumption percentages in the province of Turin

PREDICTIVE FEATURES - 2

Sentinel-2 remote sensing
3 images, 13 bands, 10 meters resolution

PROBLEM: how to include images in the model?



Cumiana/Cantalupa
24/10/2017

PREDICTIVE FEATURES - 2

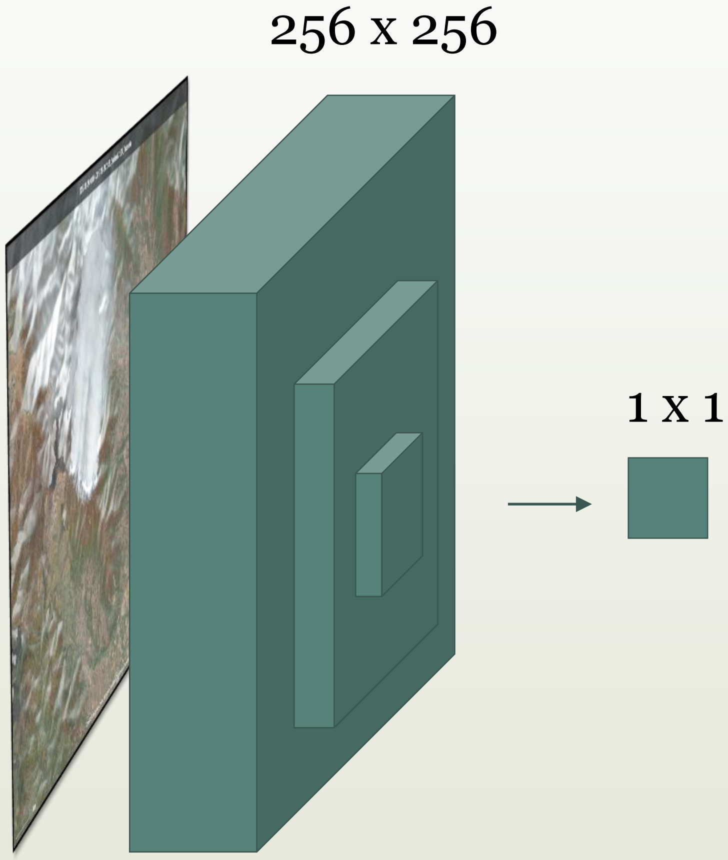
Sentinel-2 remote sensing
3 images, 13 bands, 10 meters resolution

PROBLEM: how to include images in the model?

Halve the size of the image until only 1 pixel is extracted

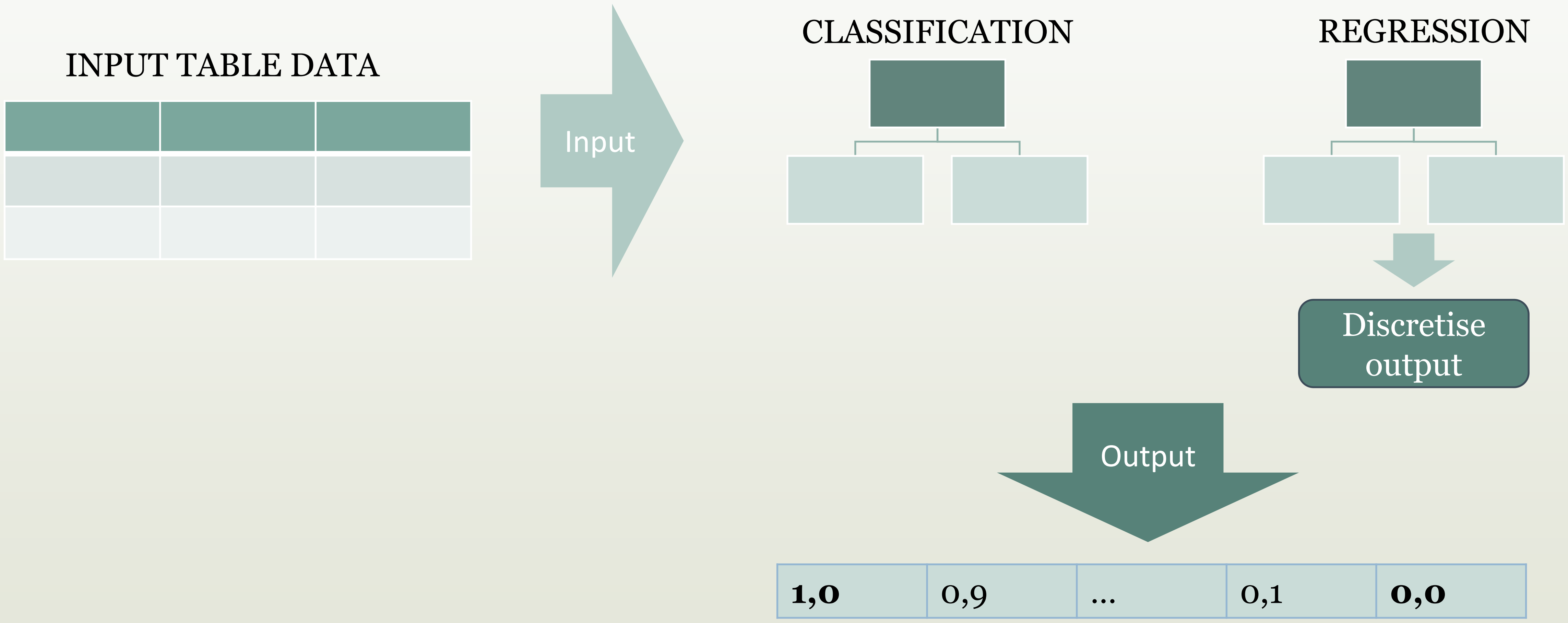
Process the images through a CNN architecture:

- kernel size 3
- stride 2
- padding 1



Note: the CNN's weights will not be trained!

RANDOM FOREST MODELS



SAMPLING METHODS

Sample A

Generate further random points and assign the Kriged map value

4226 observations

Sample B

Balanced sample from the Kriged map

11000 observations

Sample C

Balanced sample from each seasonal Kriged map

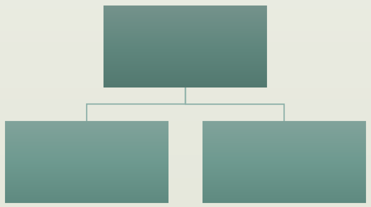
13000 observations

Train/Validation – 70%/15% | Test – 15%

FEATURE SELECTION

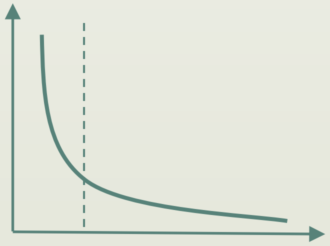
RFE selection

Recursive elimination on Random Forest algorithms



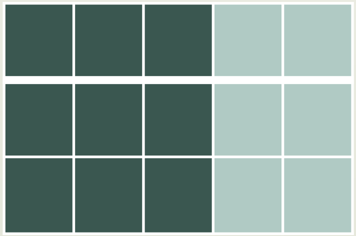
PCA reduction

Leverage on the high correlations among similar variables



Human-related filter

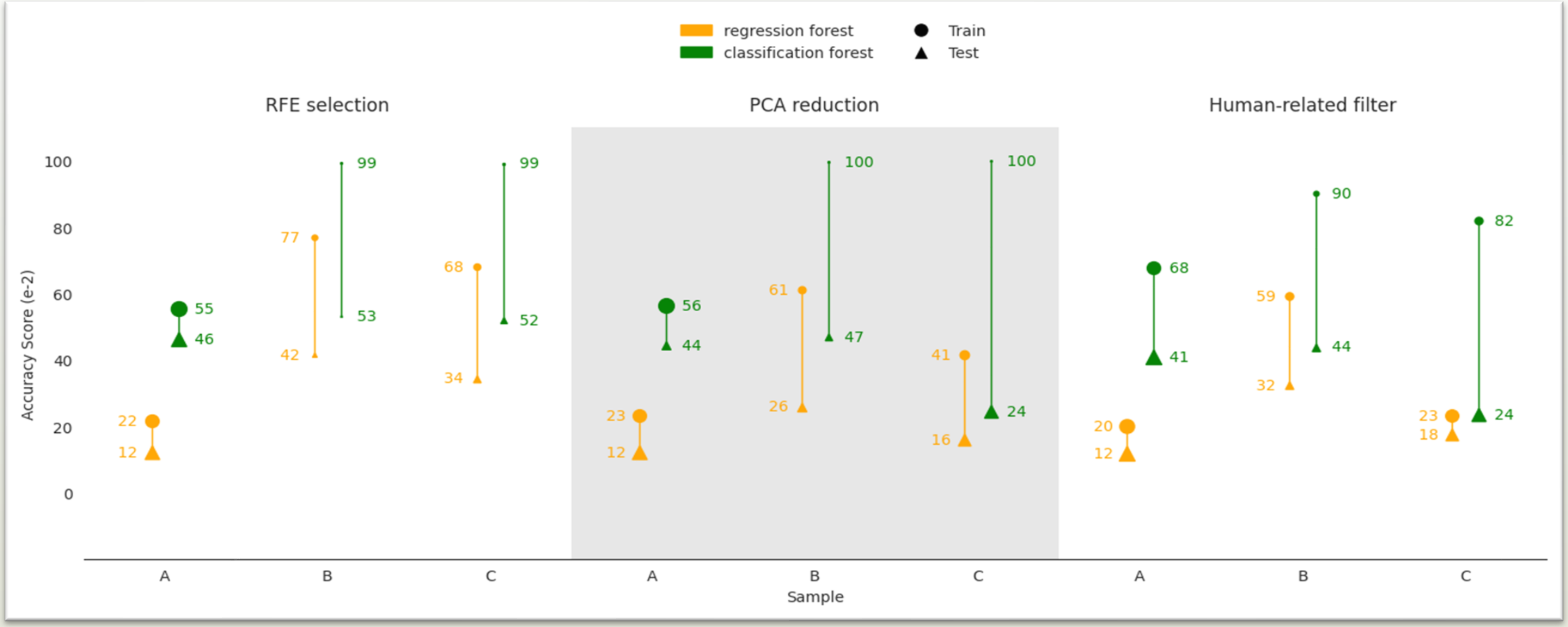
Focus only on the human impact



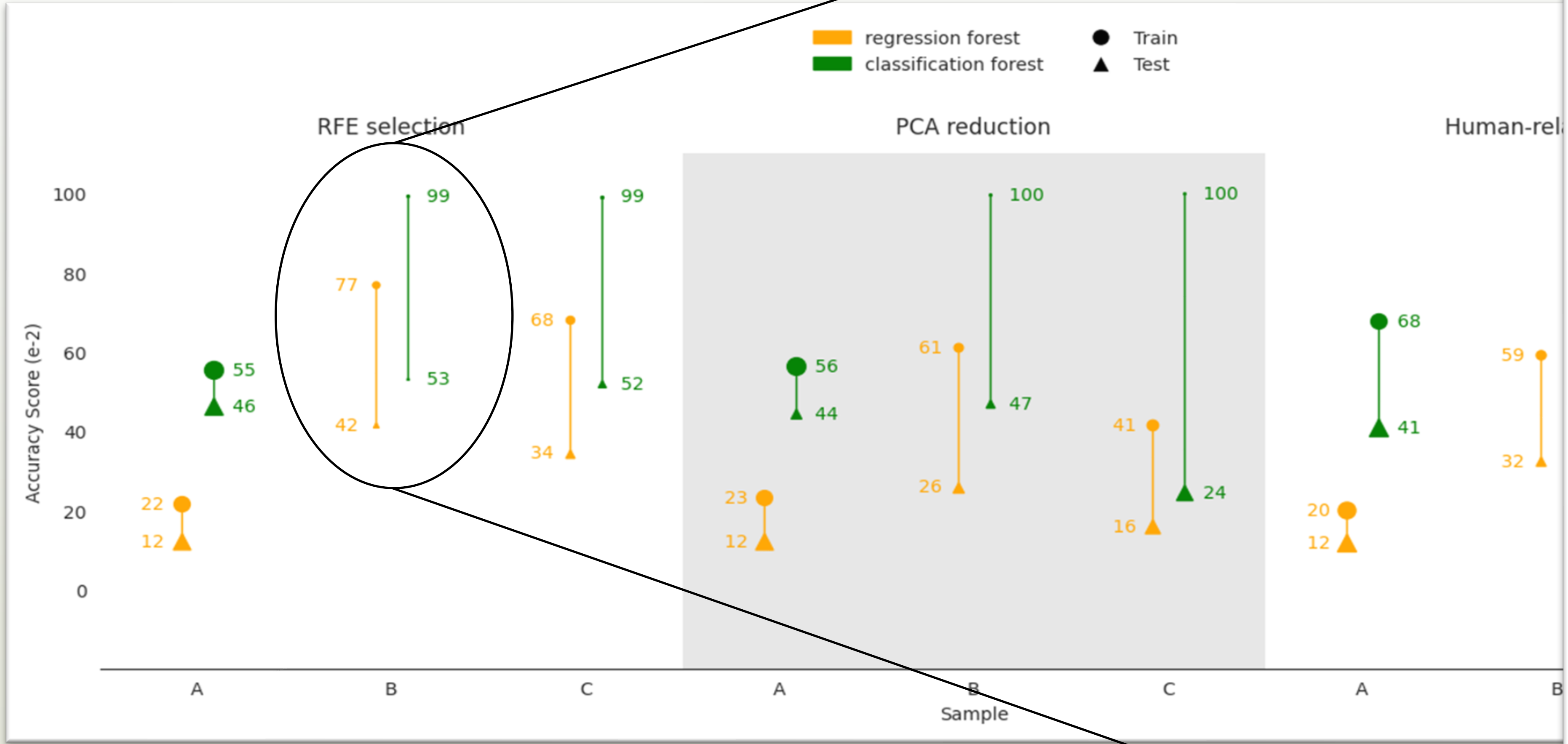
Results

Scores and features importance

PERFORMANCES



PERFORMANCES



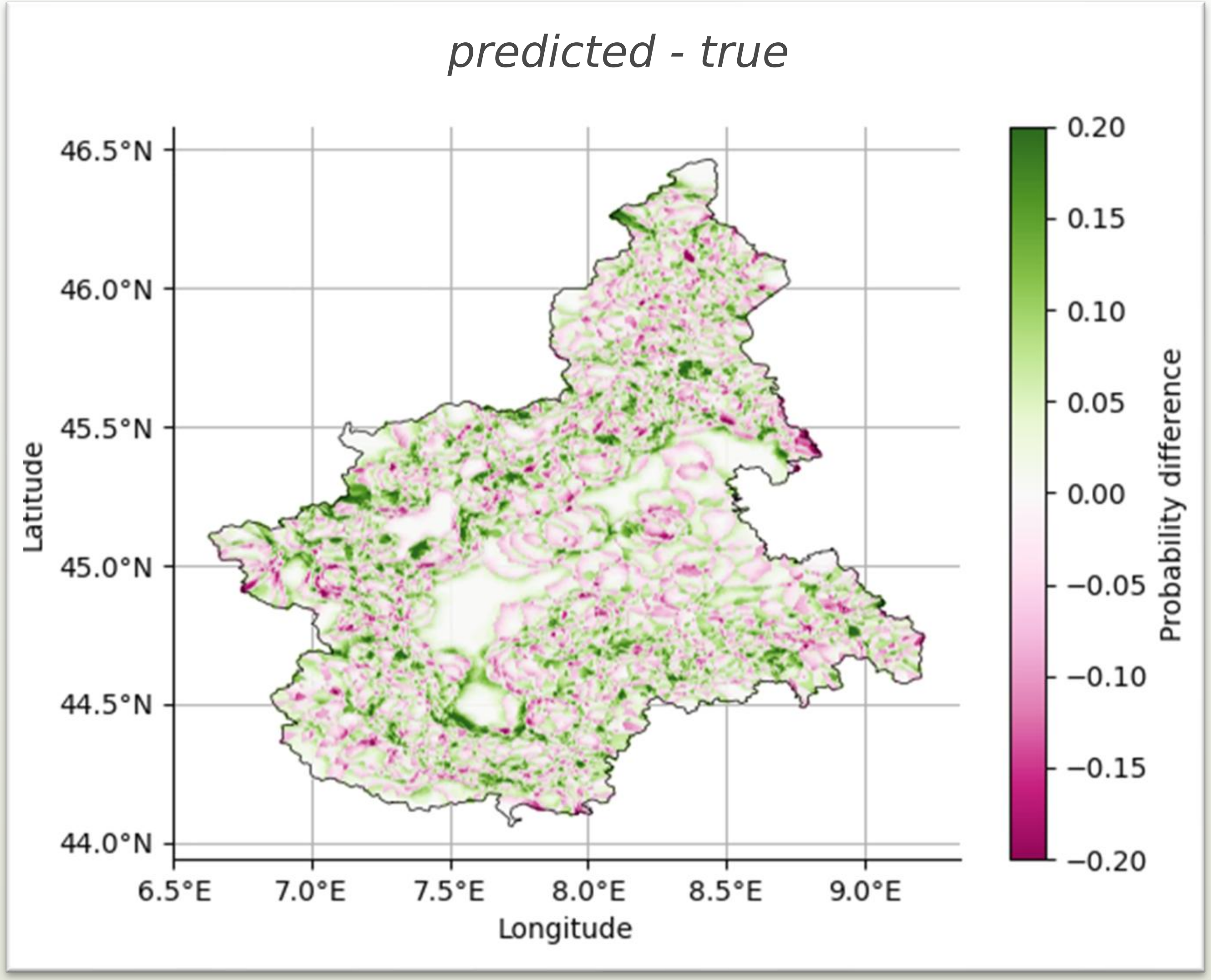
We analyse the best performing model's results

Distribution of the residuals

- The overestimations are heavier and more scattered than the underestimations

- With no error tolerance the misclassification percentage is 91.72%

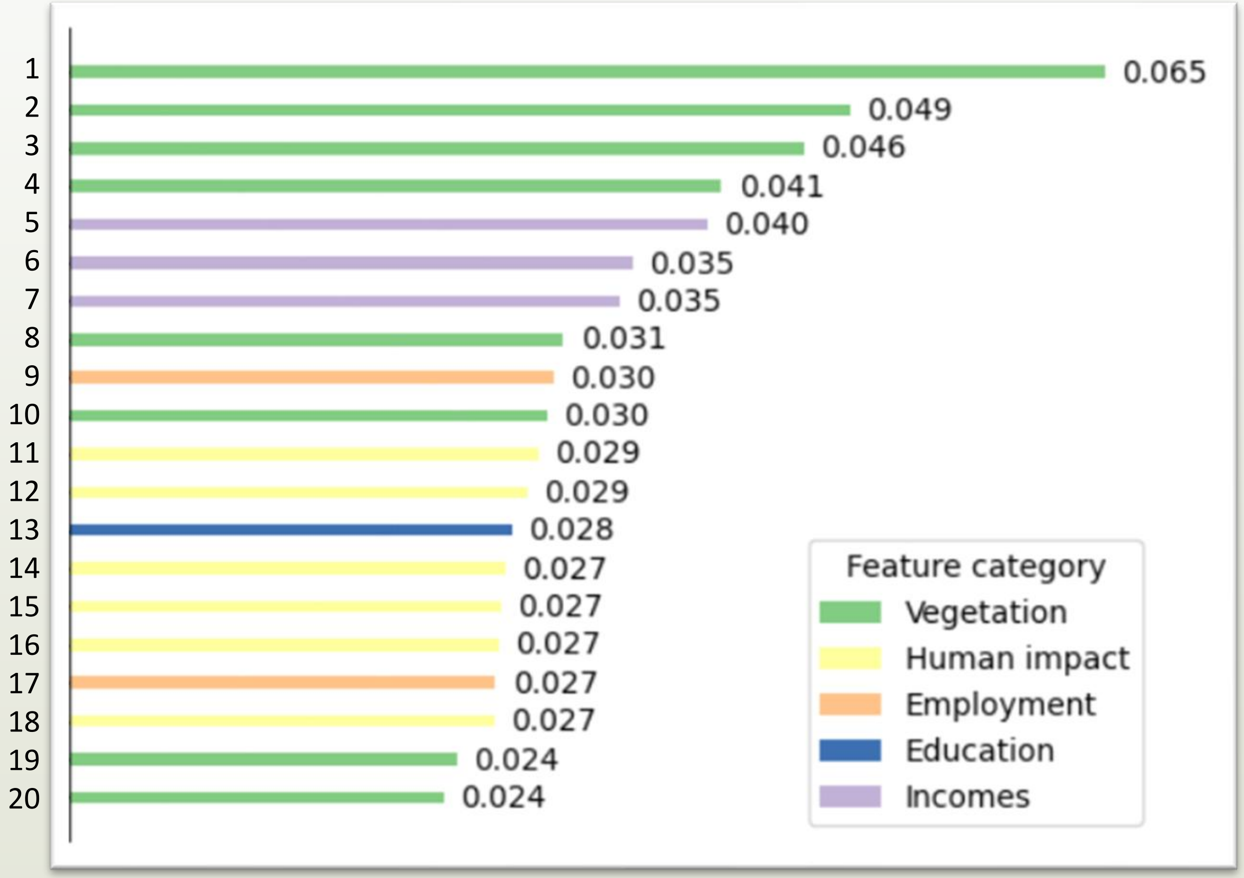
- With a tolerance of 10% the misclassification percentage reduces to the 11.52%



Feature importance

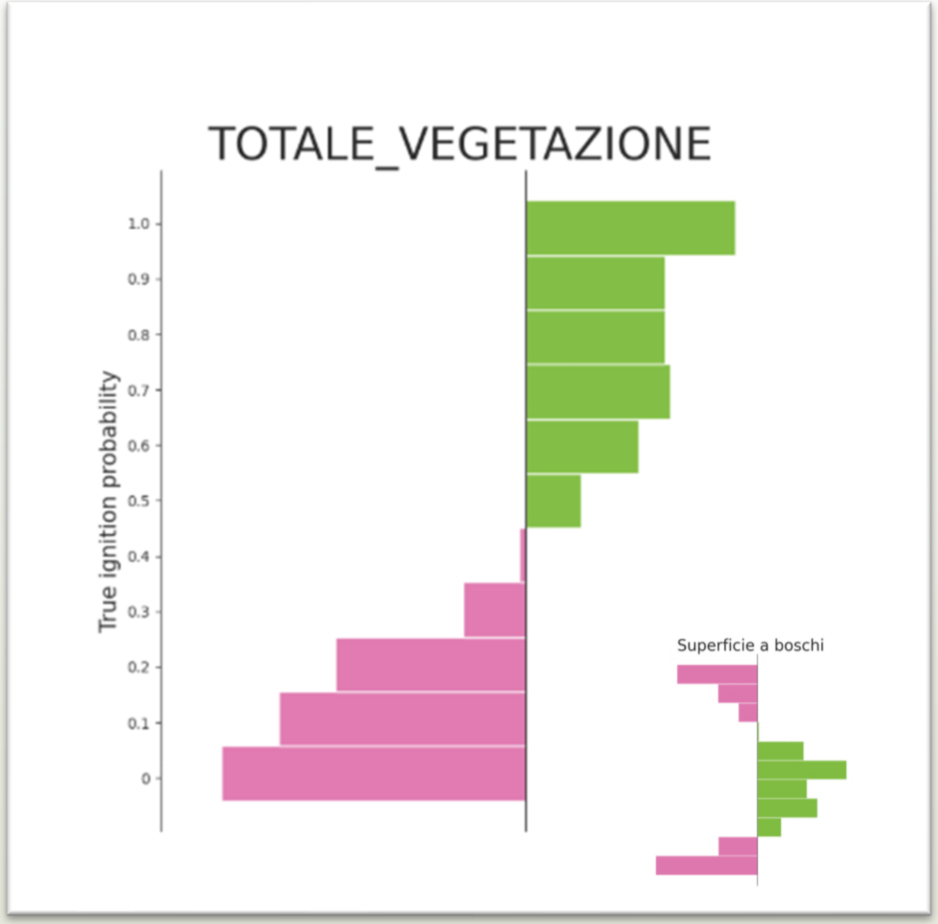
- Vegetation coverage best discriminates the target probabilities
- Socioeconomic variables describe the population's characteristics

The class distributions of the most important variables allows target profiling



Vegetation profiles

- No vegetation
  - ⇒ low probabilities
- Large controlled woods
  - ⇒ high probabilities
- The more common vegetation types overlay the past ignitions
- The sparser vegetation types make the model overfit on the training probabilities



Variable class means with respect to the overall variable mean

MORE INSIGHTS

Incomes

Higher probabilities appear in areas with many residents in the middle-to-upper wealth ranges, as well as in zones where there are not enough people to reach the average wealth level of Piedmont's townships.

Education

Past fires ignited most in administrative areas that register the lowest percentages of people with low education, surrounded by people with the highest ratios of this index.

Fragility indexes

The inhabitants that produce little rubbish but has high motorization emissions are more prone to ignite.

CONCLUSIONS

The method proved to be a powerful tool for identifying the most useful variables to discriminate the interpolated probabilities, despite its low accuracy.

The limitations of this method rely on the dependence on the interpolation method of the target and the interdependence of human-related features.

Possible improvements

- Predict the number of ignitions
- Filter high correlated variables
- Contrastive Learning

Thank you!

References - 1

[1] Emilio Chiuveco et al. (2023): Towards an Integrated Approach to Wildfire Risk Assessment: When, Where, What and How May the Landscapes Burn. Fire, 10.3390

[2] Simona Barbarino et al. (2012): Gli incendi boschivi nelle Alpi: Conoscenza, previsione e cooperazione per difendere il nostro patrimonio forestale, Arpa Piemonte

[3] https://environment.ec.europa.eu/topics/forests/forest-fires_en

[4] FIREURISK - DEVELOPING A HOLISTIC, RISK-WISE STRATEGY FOR EUROPEAN WILDFIRE MANAGEMENT, DOI: 10.3030/101003890

[5] C.E. Van Wagner (1974): Structure of the Canadian forest Fire Weather Index. Department of the environment, Canadian Forestry Service, Publication No. 1333

[6] https://forest-fire.emergency.copernicus.eu/

[7] Andrina Gincheva et al. (2024): A monthly gridded burned area database of national wildland fire data. Scientific data, 10.1038

[8] Giovanni Laneve, Marco Di Fonzo, Valerio Pampanoni, Ramon Bueno Morles (2024): Progress and Limitations in the Satellite-Based Estimate of Burnt Areas. Remonte sensing, 10.3390

[9] Sergi Costafreda-Aumedes, Carles Comas, Cristina Vega-Garcia (2017):Human-caused fire occurrence modelling in perspective: a review. In-ternational Journal of Wildland Fire, 10.1071

[10] Mario Elia et al. (2020): Estimating the probability of wildfire occurrence in Mediterranean landscapes using Artificial Neural Networks. Environmental Impact Assessment Review, 10.1016

References - 2

[11] Mariana Dondo Bühler, Mónica De Torres Curth, Lucas Alejandro Garibaldi (2013): Demography and socioeconomic vulnerability influence fire occurrence in Bariloche (Argentina). Landscape and Urban Planning, 10.1061

[12] Pedro Almeida, Isilda Cunha Menezes, Ana Isabel Miranda (2024): A Human Behavior Wildfire Ignition Probability Index for Application to Mainland Portugal. Fire, 10.3390

[13] Annalie Dorph, Erica Marshall, Kate A. Parkins, Trent D. Penman (2022): Modelling ignition probability for human- and lightning-caused wildfires in Victoria, Australia. Natural Hazards and Earth System Sciences, 10.5194

[14] Jiangxia Ye et al. (2017): Modeling the spatial patterns of human wildfire ignition in Yunnan province, China. Applied Geography, 10.1016

[15] Rita Durão, Catarina Alonso, Célia Gouveia (2022): The Performance of ECMWF Ensemble Prediction System for European Extreme Fires: Portugal/Monchique in 2018. Atmosphere, 10.3390

[16] https://land.copernicus.eu/en/products/corine-land-cover

[17] Leone D. Mancini Piermaria Corona, Luca Salvati (2018): Ranking the importance of Wildfires' human drivers through a multi-model regression approach. Environmental Impact Assessment Review, 10.1016

[18] Luiz Felipe Galizia, Thomas Curt, Renaud Barbero and Marcos Rodrigues (2021): Understanding fire regimes in Europe. International Journal of Wildland Fire, 10.1071

[19] Pere Joan Gelabert et al. (2025): Assessing human-caused wildfire ignition likelihood across Europe. The EGU interactive community platform, 10.5194

References - 3

[20] Meijer, J.R., Huijbregts, M.A.J., Schotten, C.G.J. and Schipper, A.M. (2018): Global patterns of current and future road infrastructure. Environmental Research Letters, 13-064006

[21] https://www.arpa.piemonte.it/bollettino/bollettino-pericolo-incendi-boschivi

[22] Gilles Louppe (2014): Understanding Random Forest, University of Liège

[23] P. K. Kitanidis (1997): Introduction to Geostatistics: Applications to Hydrogeology. Cambridge University Press

[24] https://geostat-framework.readthedocs.io/projects/pykrige/en/stable/variogram_models.html

[25] Ing. Antonella Vecchio, Dr. Marco Falconi: Approfondimenti di statistica e geostatistica. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici (APAT)

[26] https://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z00000076000000.htm

[27] https://www.publichealth.columbia.edu/research/population-health-methods/kriging-interpolation

[28] https://www.agenziaentrate.gov.it/portale/archivio/modelli-e-istruzioni/modelli-2008-2016/modelli-di-dichiarazione/2007/unico-pf-2007/fascicolo_3/1_istruzioni_per_la_compilazione_dei_quadri_aggiuntivi_al_modello_base/6_istruzioni_per_la_compilazione_del_quadro_rg.html

[29] Caitlyn Reilley et al. (2023): The Influence of Socioeconomic Factors onHuman Wildfire Ignitions in the Pacific Northwest, USA. Fire, 10.3390

The variogram

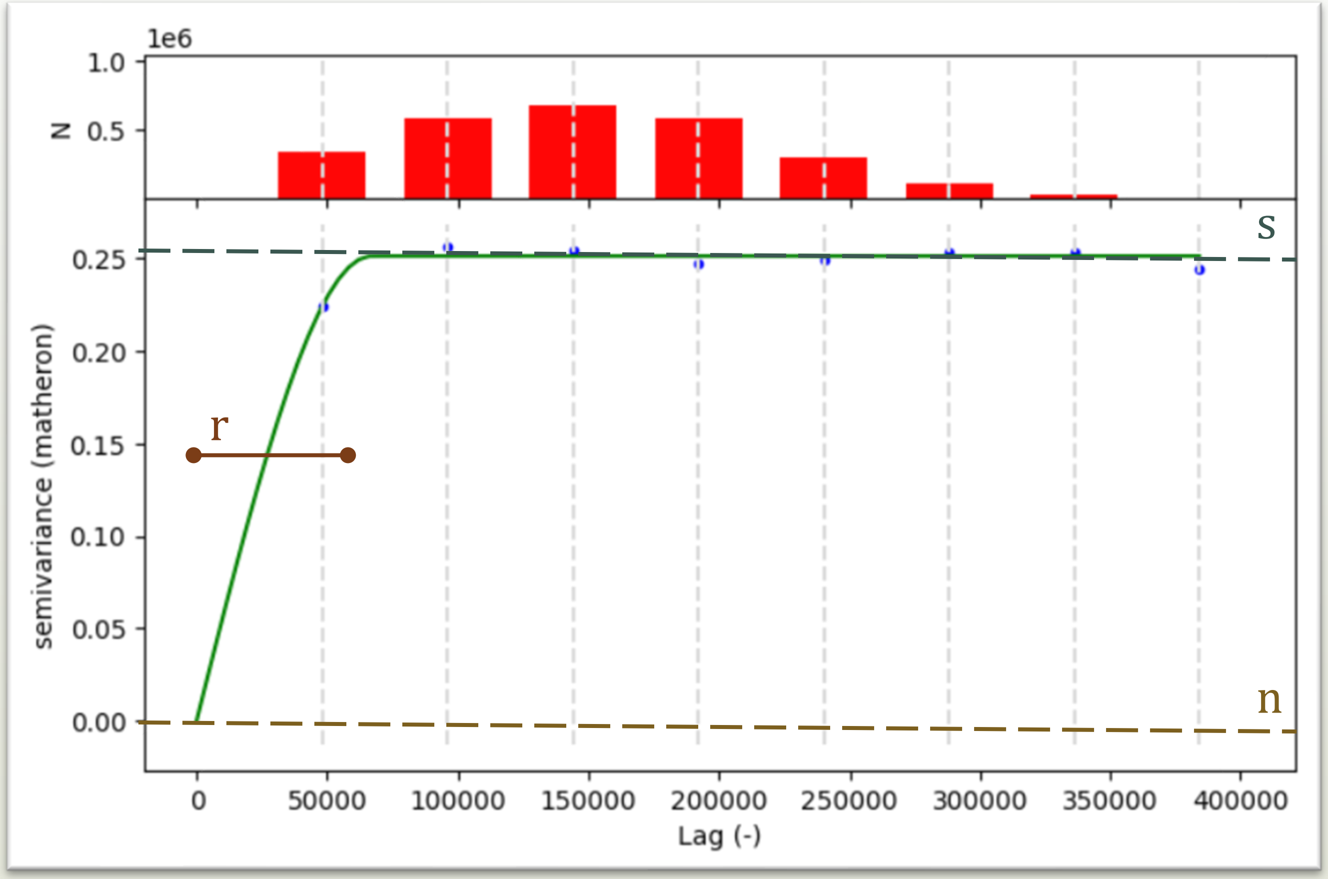
It is a function representing the variance of the values *p* against the spatial distance between the domain points **x**.

The Kriging algorithm requires an approximation model of the variogram.

Spherical variogram model:

$$\gamma(d) = \begin{cases} f(d) + n & d \leq r \\ s & d > r \end{cases}$$

where *f* is a spherical function



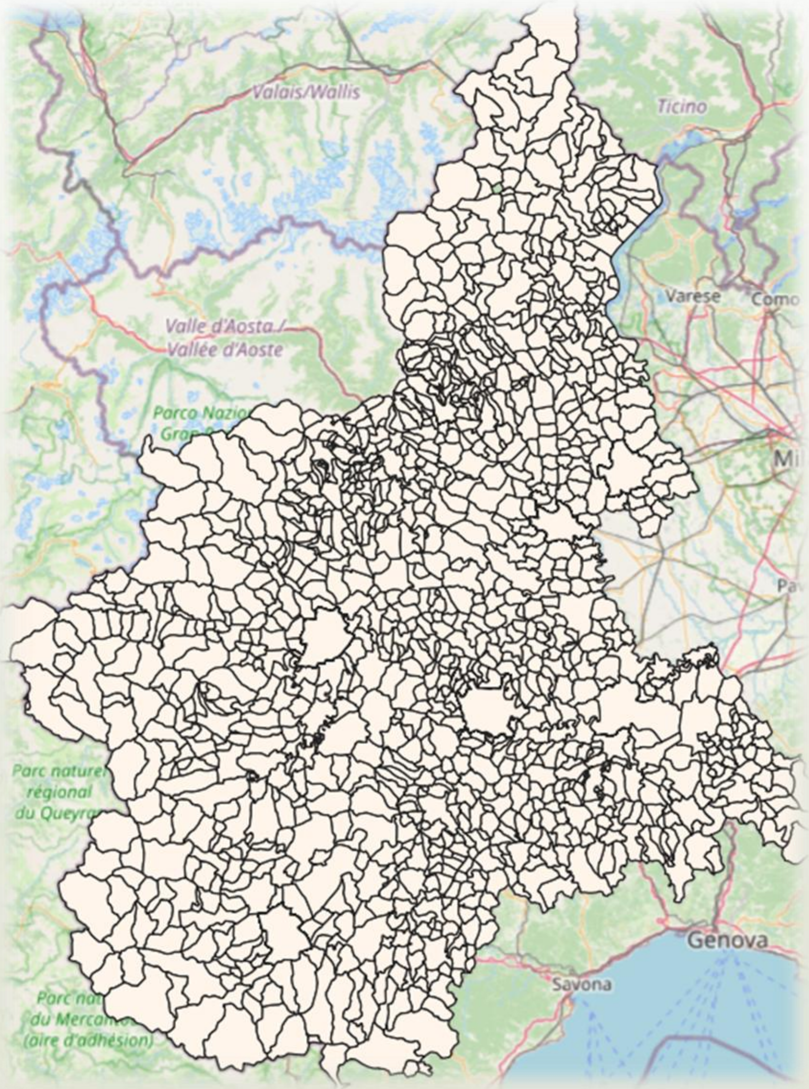
PREDICTIVE FEATURES - 3

The *road coverage map* classifies 5 types of roads.

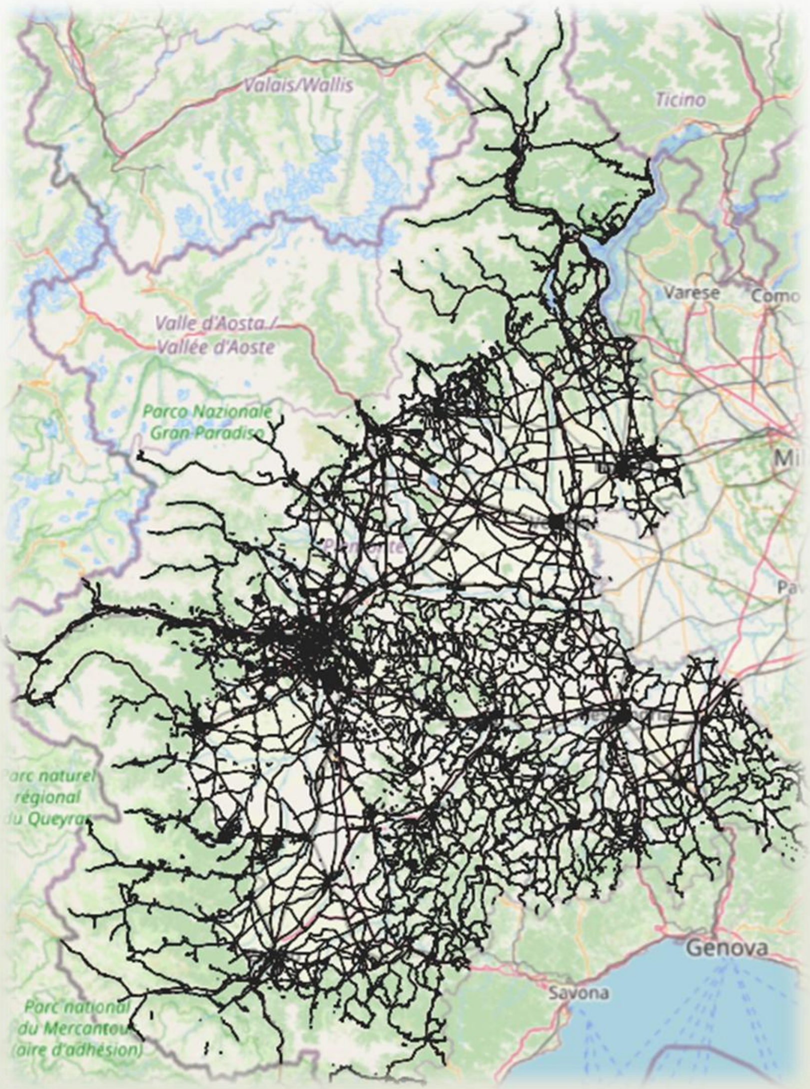
The *vegetation cover map* classifies the flora depending on category, type, planned intervention and priority, management status.

The **coverage percentage** of each type of asset is assigned to each target point within 1 kilometer around them.

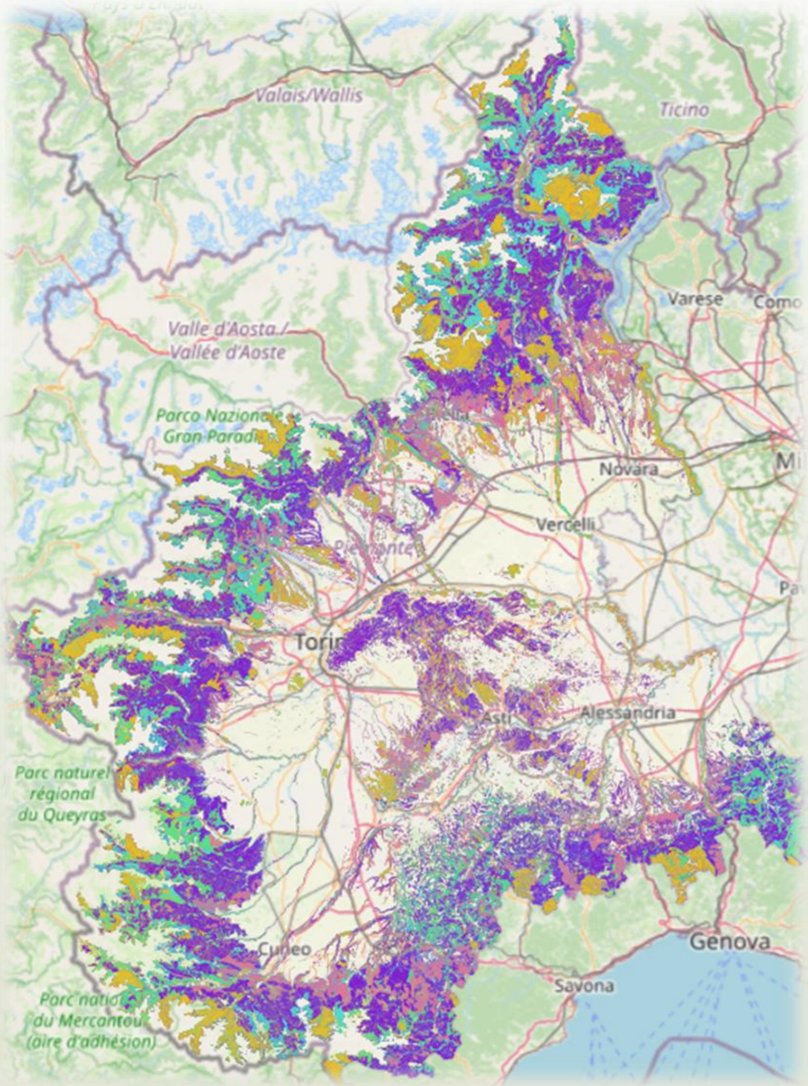
Municipalities boundaries



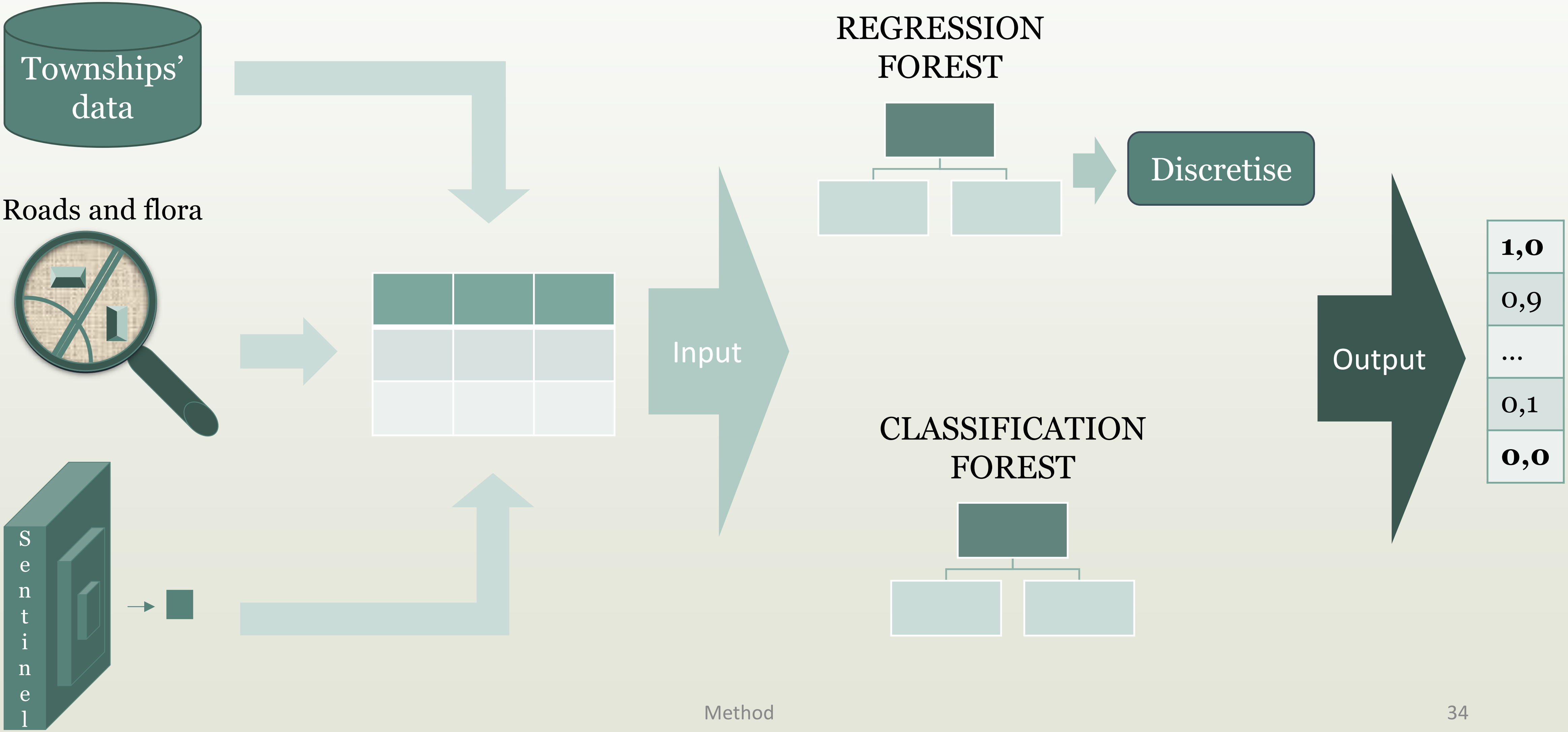
Road web



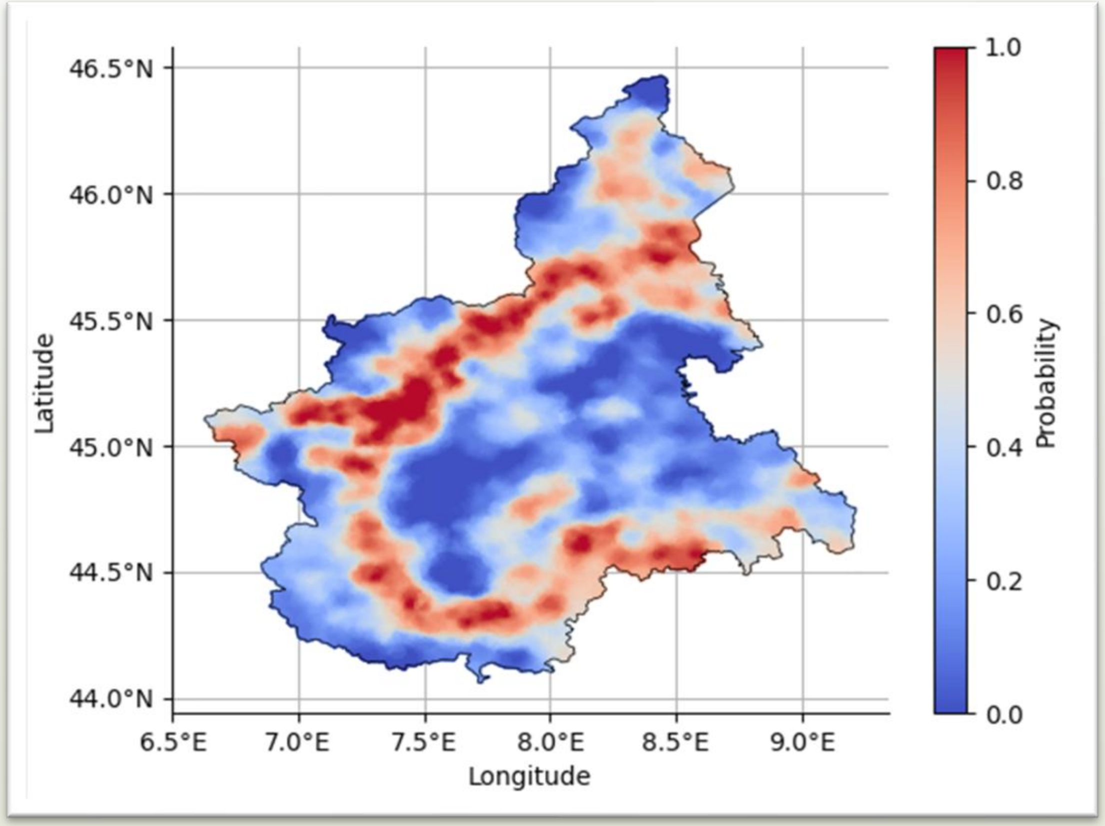
Vegetation coverage



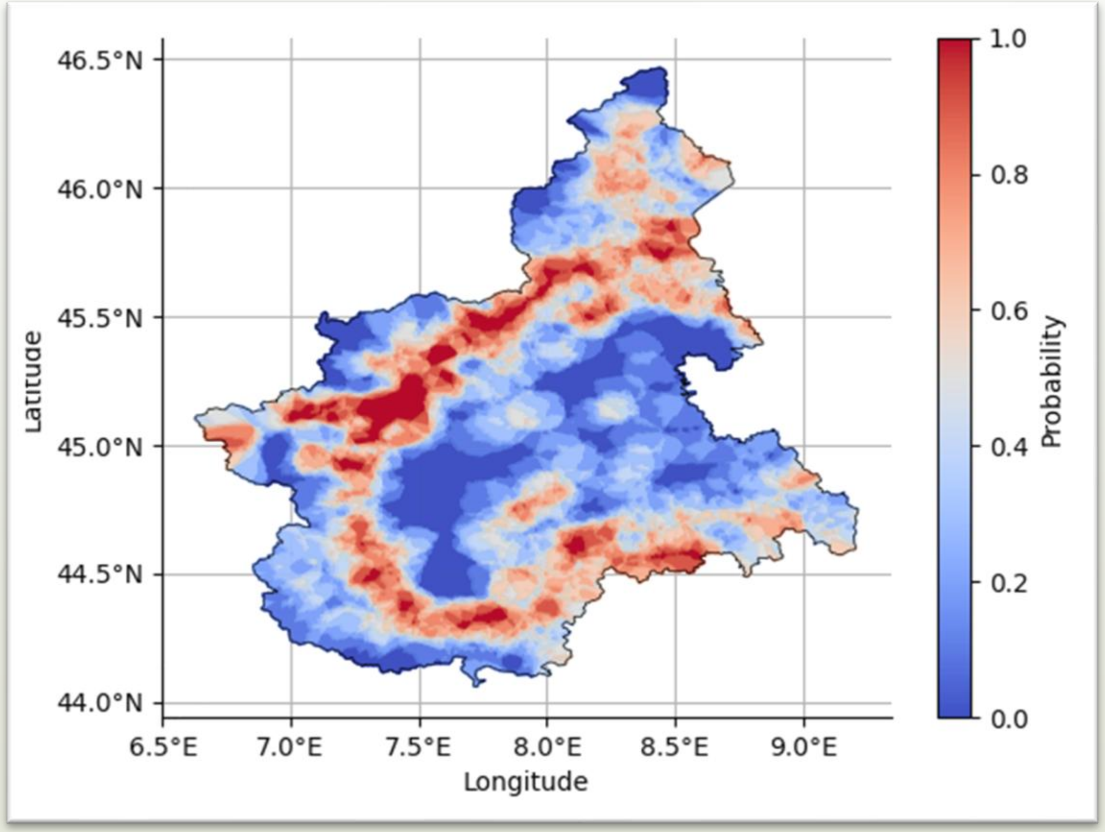
THE MODEL'S ALGORITHM



Kriged map trained on the predicted values



Kriged map trained on the true values



FEATURE SCORES