Regression of species richness biodiversity indicator from satellite observations and environmental parameters

MALIS course project, fall semester 2020

Cosimo Chetta, Vincenzo Madaghiele

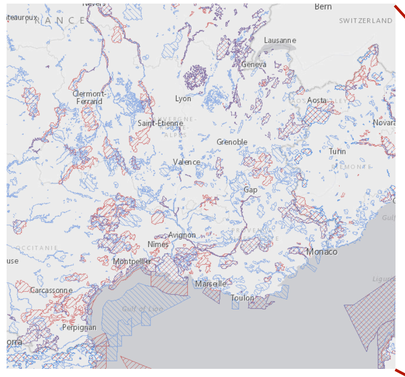
EURECOM
Sophia Antipolis

Dataset and features

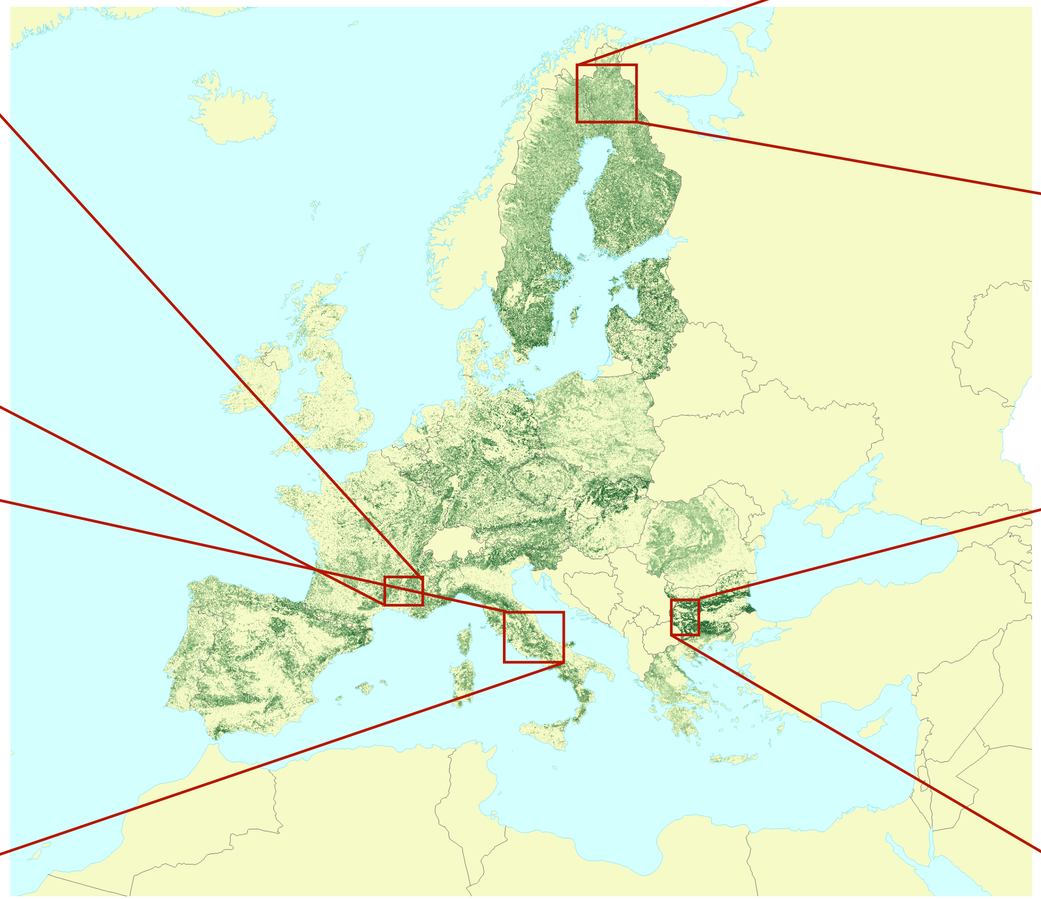
Features

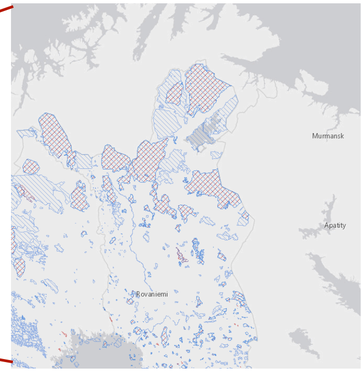
Regression index

Chosen areas

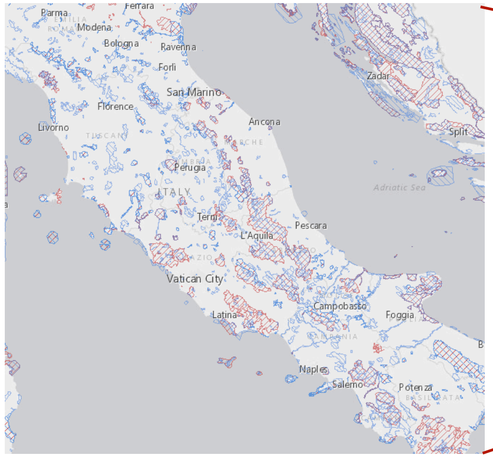


France, N 44.4 - 44.8° E 3.6 - 4.5°

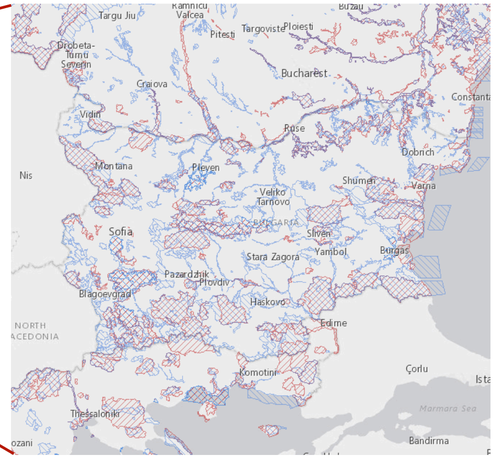




Finland, N 67.0 - 69.0° E 25.0 - 28.0°



Italy, N 40.5 - 43.0° E 12.5 -15.5°



Bulgaria, N 41.0 - 43.5° E 22.0 - 24.0°

EEA geospatial data catalogue, Richness of forest-related species and habitats indicator 2012 dataset, Nov. 2018

Cosimo Chetta, Vincenzo Madaghiele

EURECOM

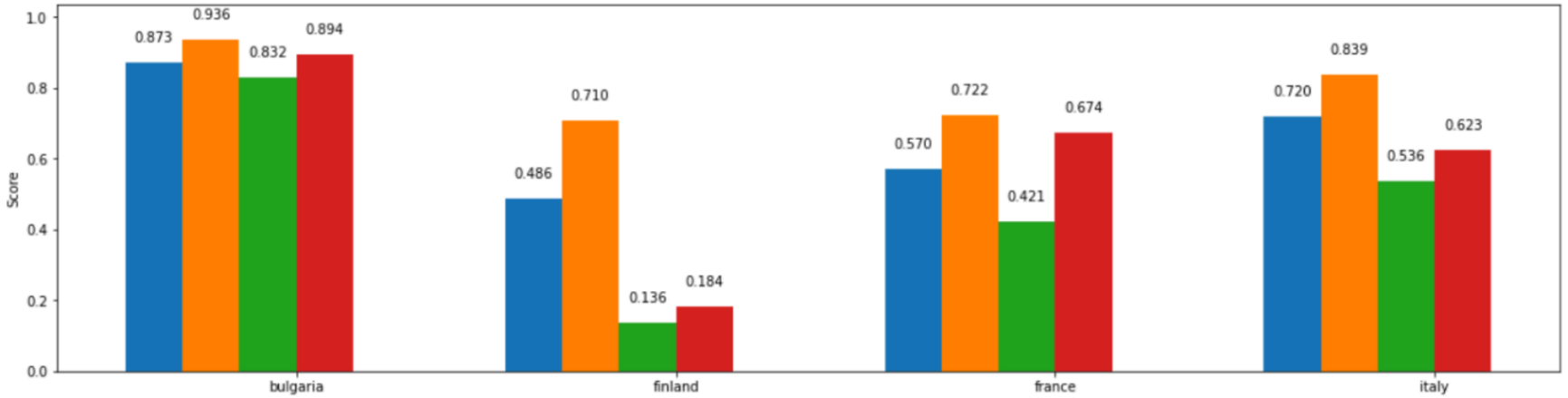
Data cleaning

Handling invalid data and pre-processing

Region	*Remove*	*Mean*	*Closest mean*	*knn imputer*
Bulgaria	0.918	0.925	0.925	0.923
Finland	0.713	0.717	0.717	0.717
France	0.732	0.742	0.742	0.735
Italy	0.803	0.832	0.835	0.827

Models comparison

R^2 score of the model for each region for a 4-fold cross validation



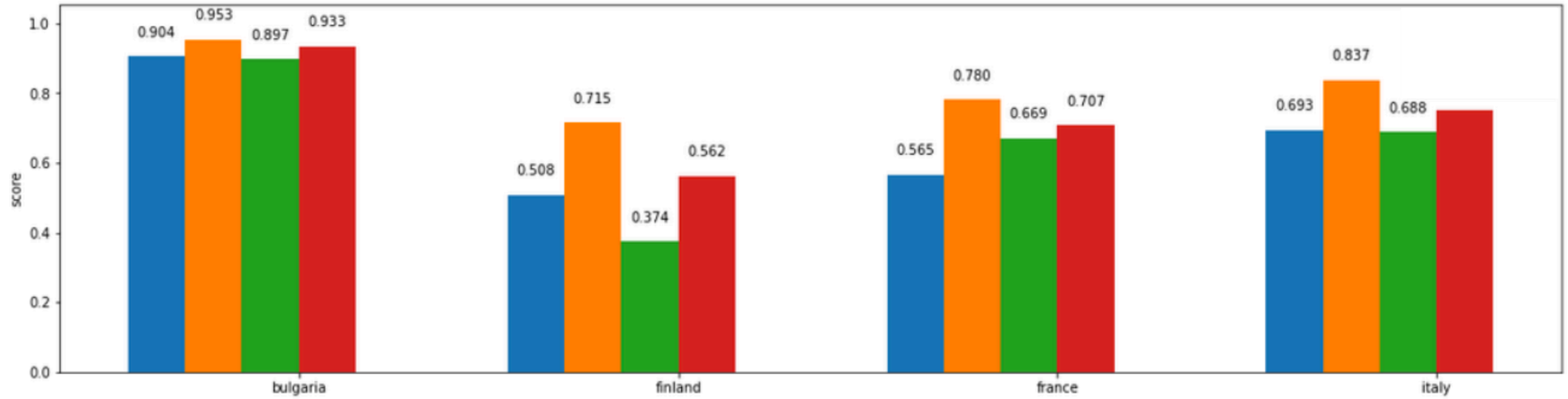
Decision Tree

Random Forest

Support Vector Machine

Neural Network

R^2 score of the model for each region for a 4-fold cross validation on log(species richness)

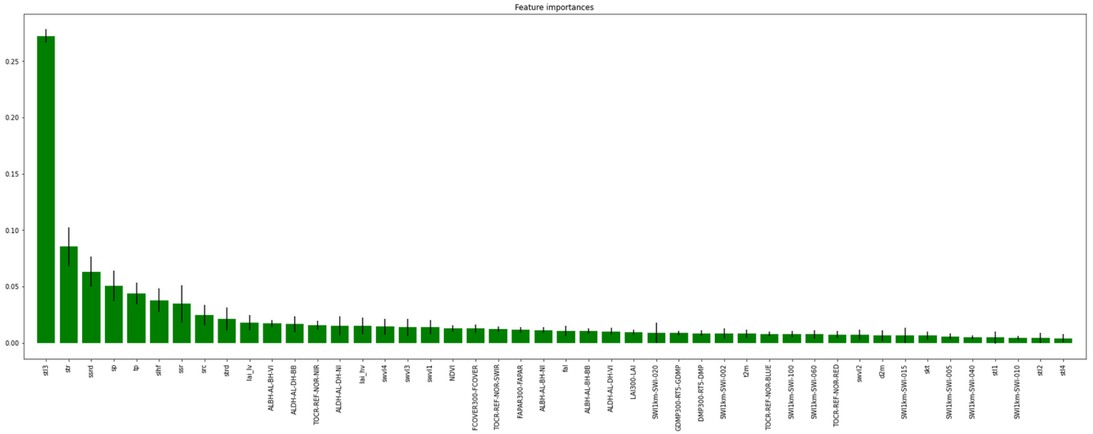


Cosimo Chetta, Vincenzo Madaghiele

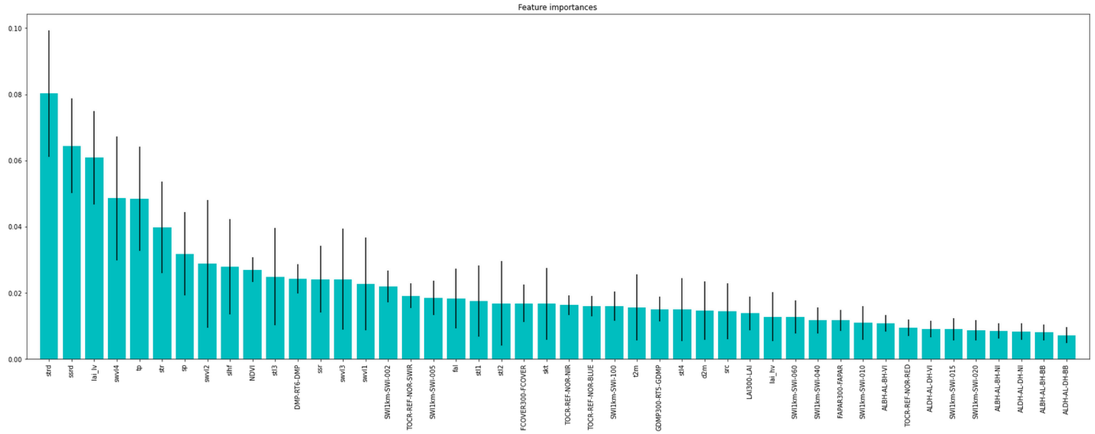
EURECOM

Feature importances

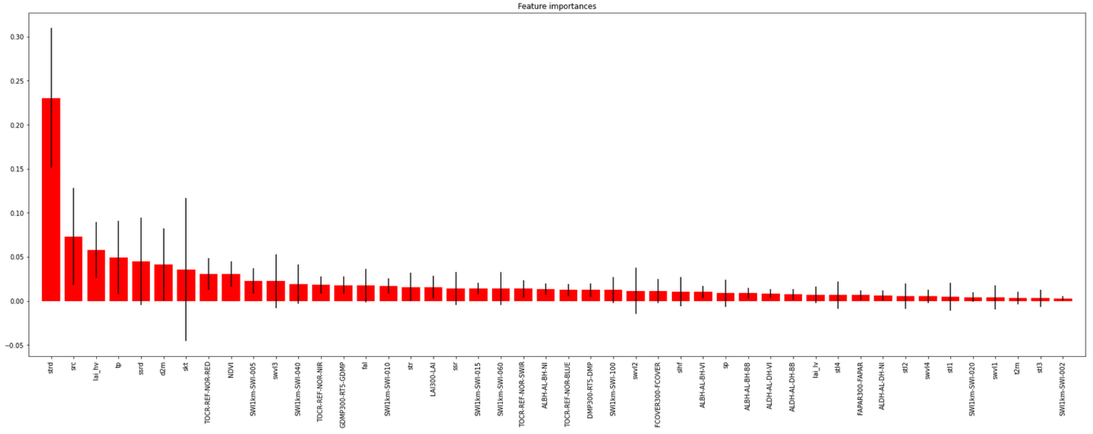
Bulgaria



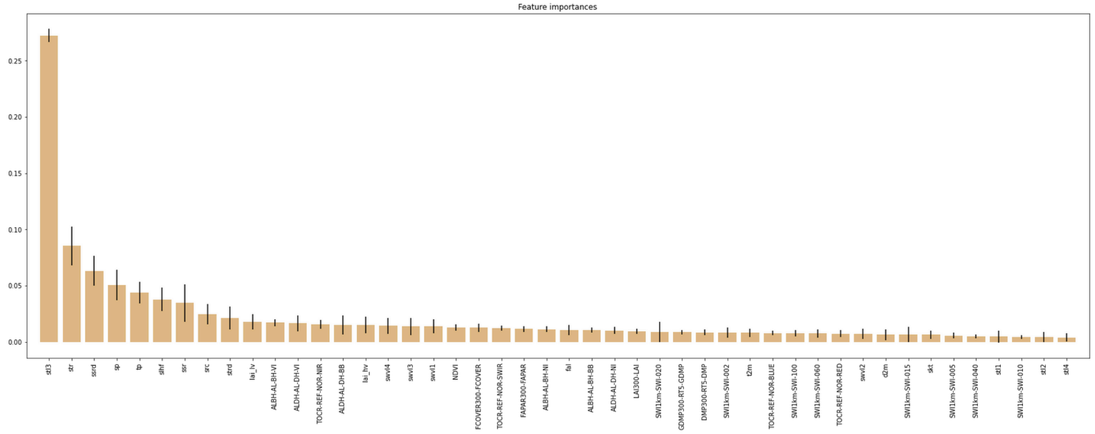
Finland



France



Italy

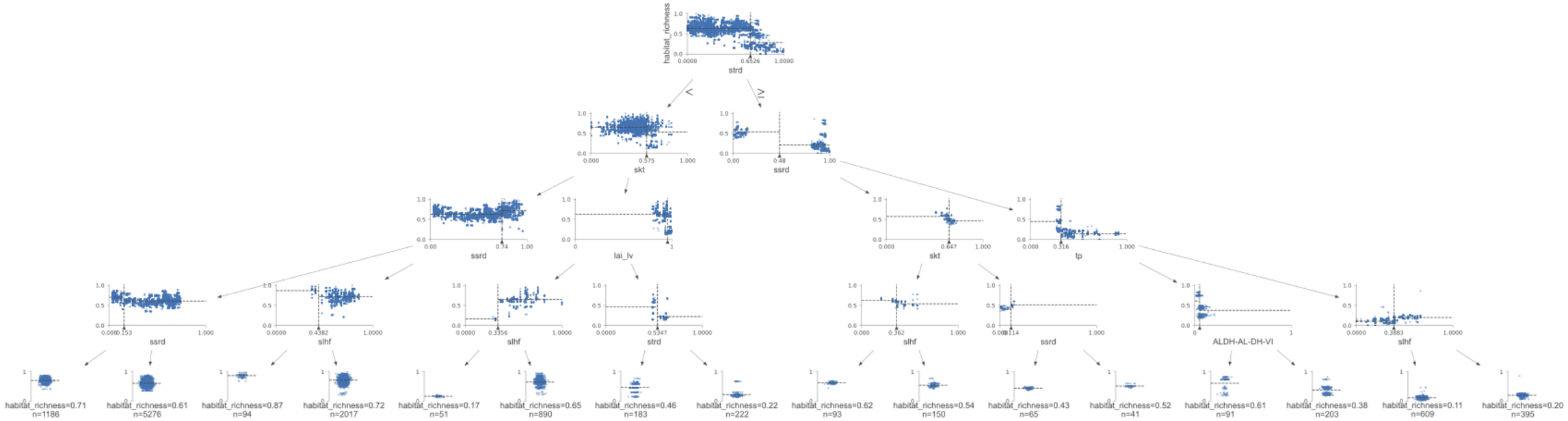


Cosimo Chetta, Vincenzo Madaghiele

EURECOM
Sophia Antipolis

Regression Tree visualization

Insight into the model’s decisions



Cosimo Chetta, Vincenzo Madaghiele

Discussion and future work

Conclusions

- Greater score with respect to the literature
- Only open access data that can be derived by satellites observations
- Ecological insight by analysis of feature importances and regression trees
- Scalability to different habitats

Limitations

- Performance is influenced by area size and internal composition
- Possibility of error propagation from feature derivation
- Need for further validation on different areas and richness indexes

Thank you!

Vincenzo Madaghiele (vincenzo.madaghiele@eurecom.fr)

Cosimo Chetta (cosimo.chetta@eurecom.fr)

Code, datasets and visualizations are available at:
https://github.com/vincenzomadaghiele/Regression-of-biodiversity-indicators