Task 2.8 "Augment precision in MLs detection"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823805

Georgios Spanos, AUTh

Contents

Introduction and Objectives

- Task 2.3 detection methodology:
  - 19 factors/layers (environmental – socioeconomic) acquired and analyzed
  - based on 2017-’18
- Task 2.8 detection methodology
  - Time series analysis using Earth Observation data
  - Accuracy assessment
  - Evaluate the potential of the algorithm in a case study

Data

- **Sentinel-1**
  (GRD)
- **Sentinel-2**
  (SR aka Level 2A)

} image collections (Google Earth Engine)

- **Task 2.3 Hard Layers**
- **Sentinel-2 Global Land Cover (S2GLC) 2017 product**
- **Task 2.3 "ML_Hard_Thresholds" layer**
- **Task 2.4 validation polygons**

Workflow development

Supervised classification

Random Forest classifier

- unexcelled in accuracy
- efficient on large data bases
- handles thousands of input variables
- few tuning parameters
- no overfit
- classification variables importance estimation
- internal unbiased estimate of the generalization error (Out-Of-Bag Error Estimate

Sources: Oshiro et al. 2012, Kruppa et al. 2013, Breiman 1999

Workflow development

Training Data

- Forest
- Croplands
- Impervious

Hard Layers

- Water → S2GLC
- ML → ML_Hard_Thresholds

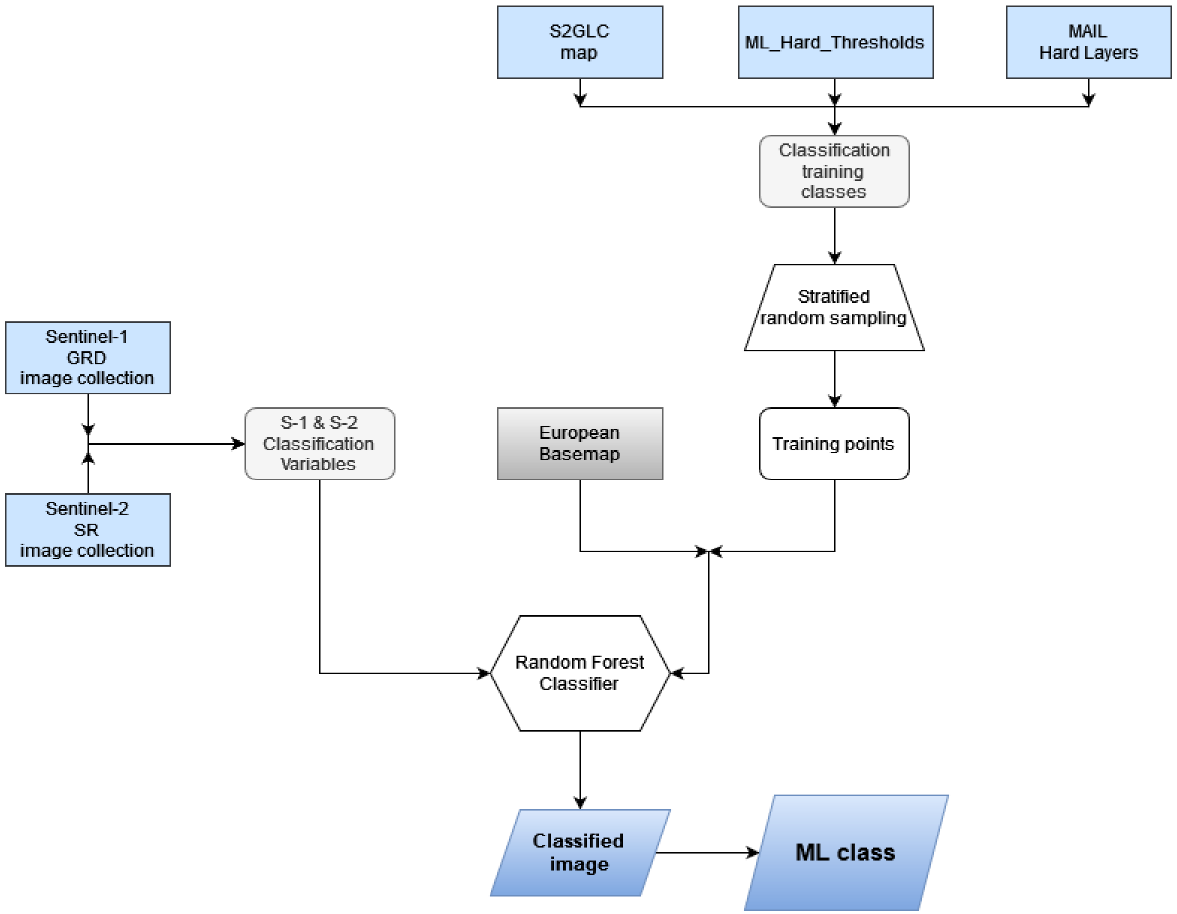
European training image

Workflow development

Classification Tool

User inputs:

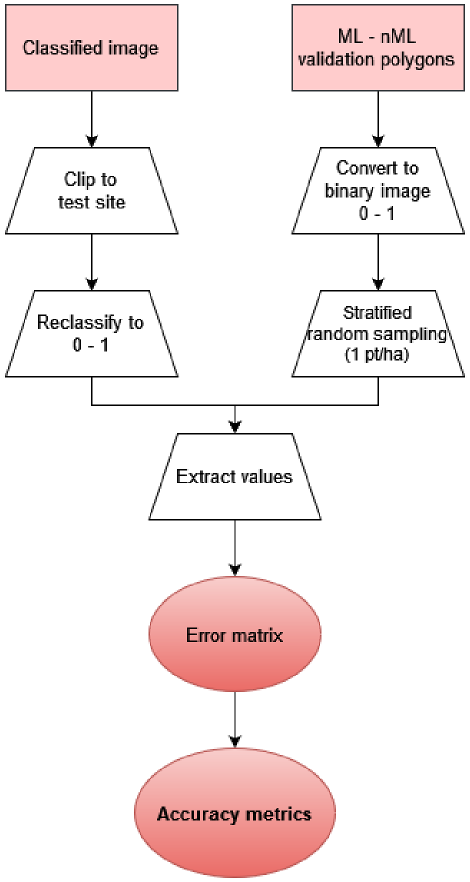
- AOI
- Date
  (>2 years range)



Workflow development

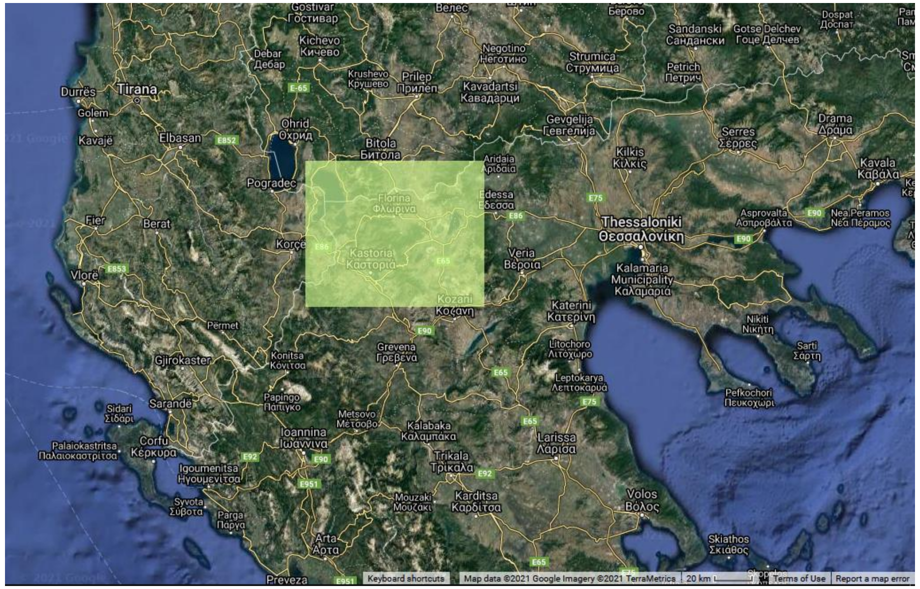
Accuracy Assessment

- Evaluate the performance of the MLs detection methodology
- Comparable with the Task 2.4 “Accuracy Assessment”
- Task 2.4 Validation polygons

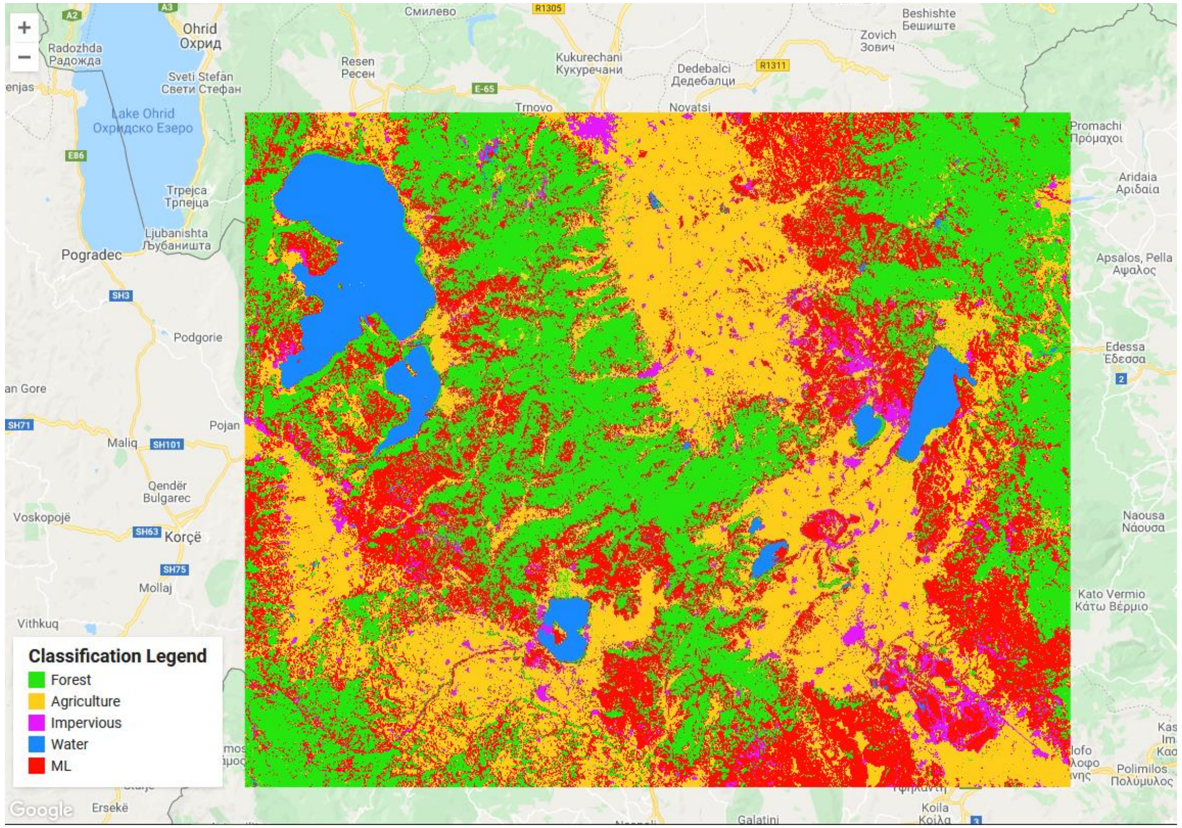


Results

Example Area

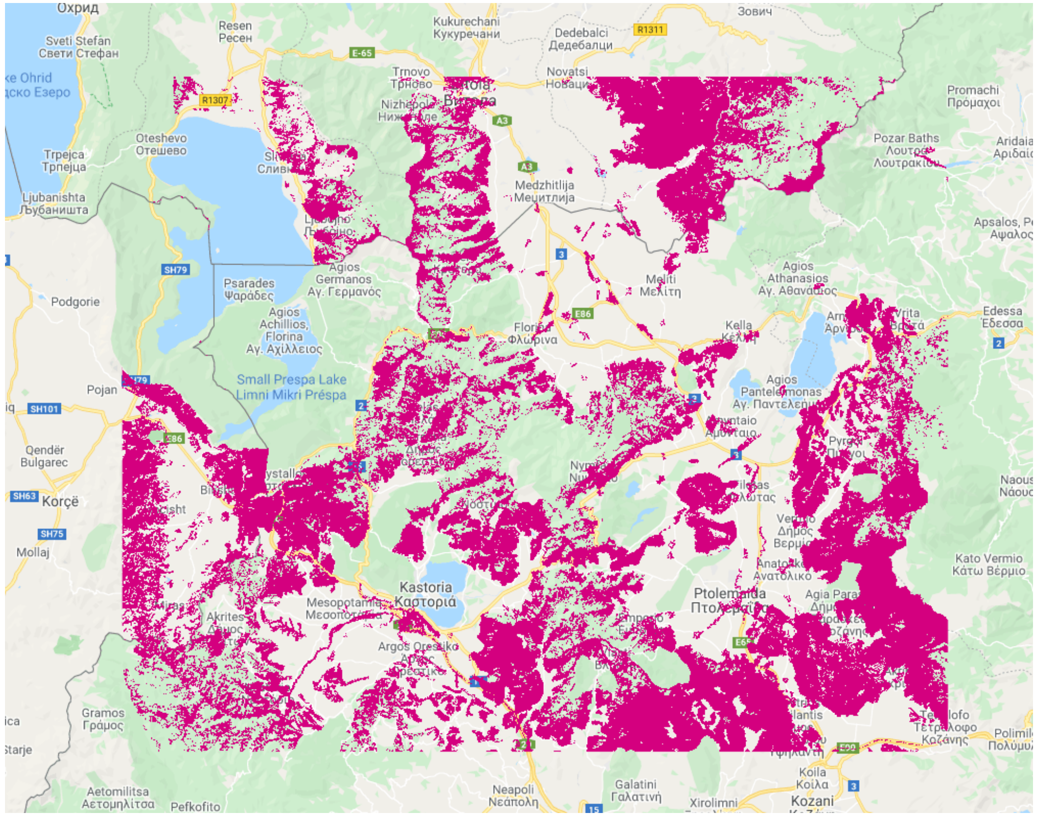


LC classification

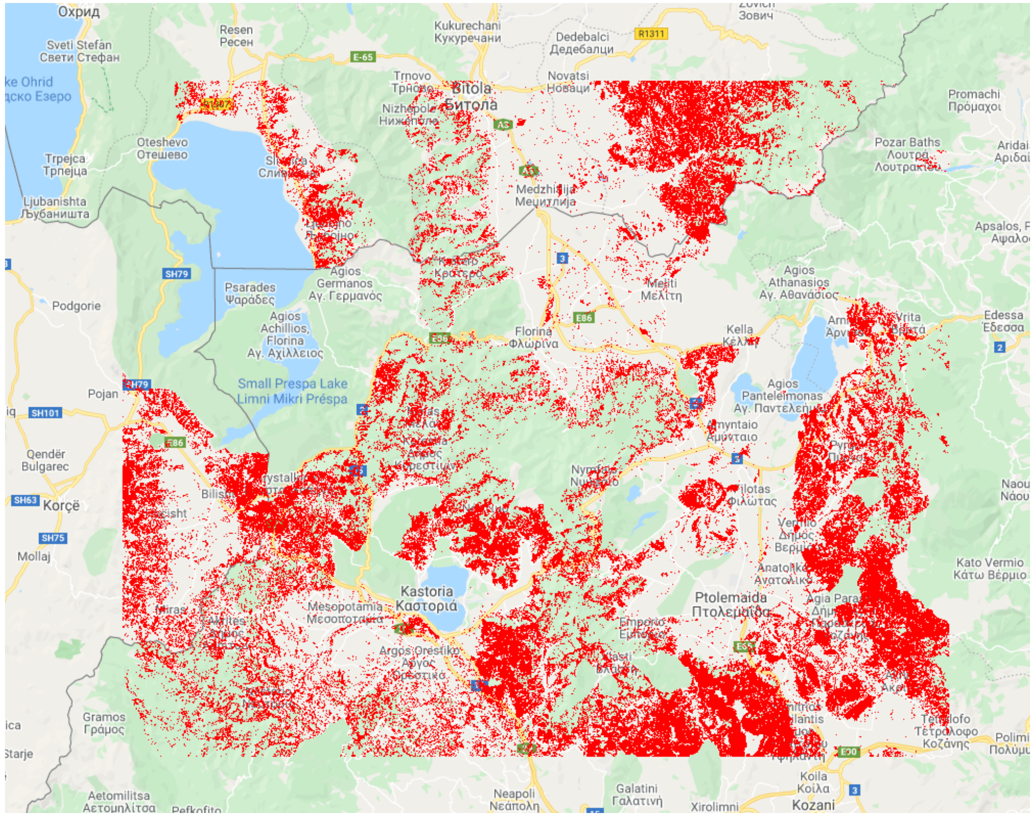


Results

MLs in 2.3



MLs in 2.8



Results

Accuracy Assessment

- Greece & Spain small decrease
- Germany stable
- Poland slight increase

Country	Overall Accuracy		Kappa		OOB Error Estimate
Task	2.3	2.8	2.3	2.8	
Greece	*0.715*	*0.614*	*0.41*	*0.23*	*0.26*
Spain	*0.828*	*0.731*	*0.65*	*0.43*	*0.23*
Germany	*0.606*	*0.543*	*0.04*	*0.03*	*0.21*
Poland	*0.909*	*0.942*	*0.64*	*0.77*	*0.25*

Notes

- Temporal aspect
- Quality of input data

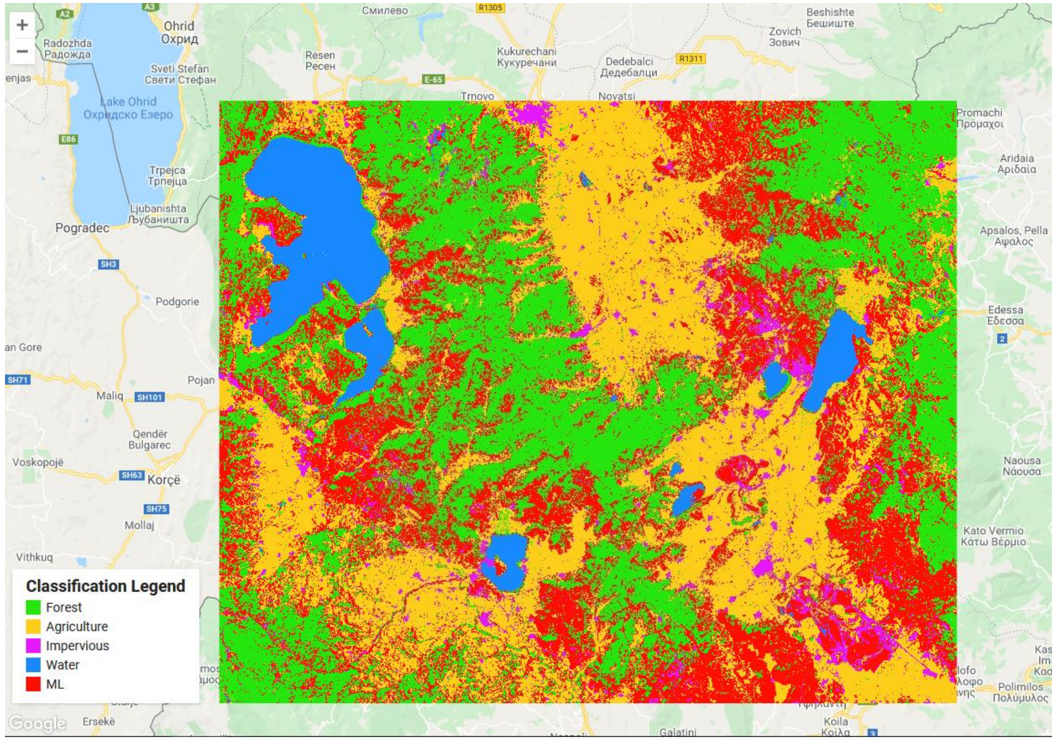
Case Study

- What is the potential of the algorithm should data of higher quality are available for the ML class?

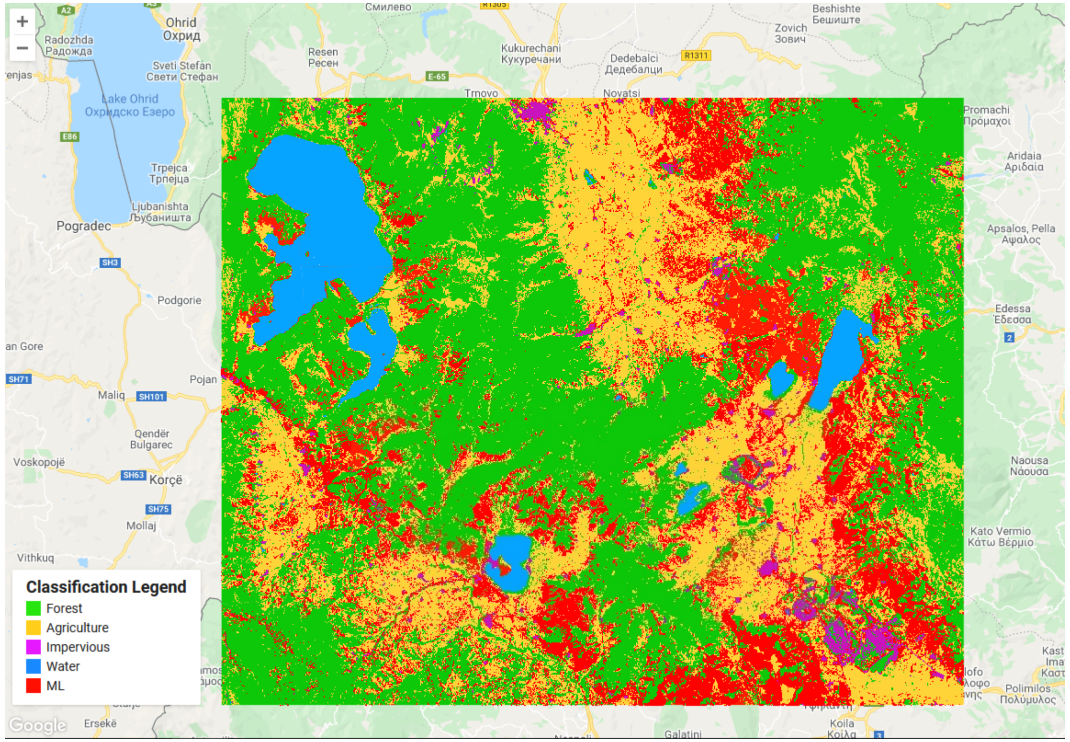
- Employ 2.4 validation polygons
    - Some used for training of the ML class
    - Rest used for validation

Case Study

ML training:
2.3 ML_Hard_Thresholds

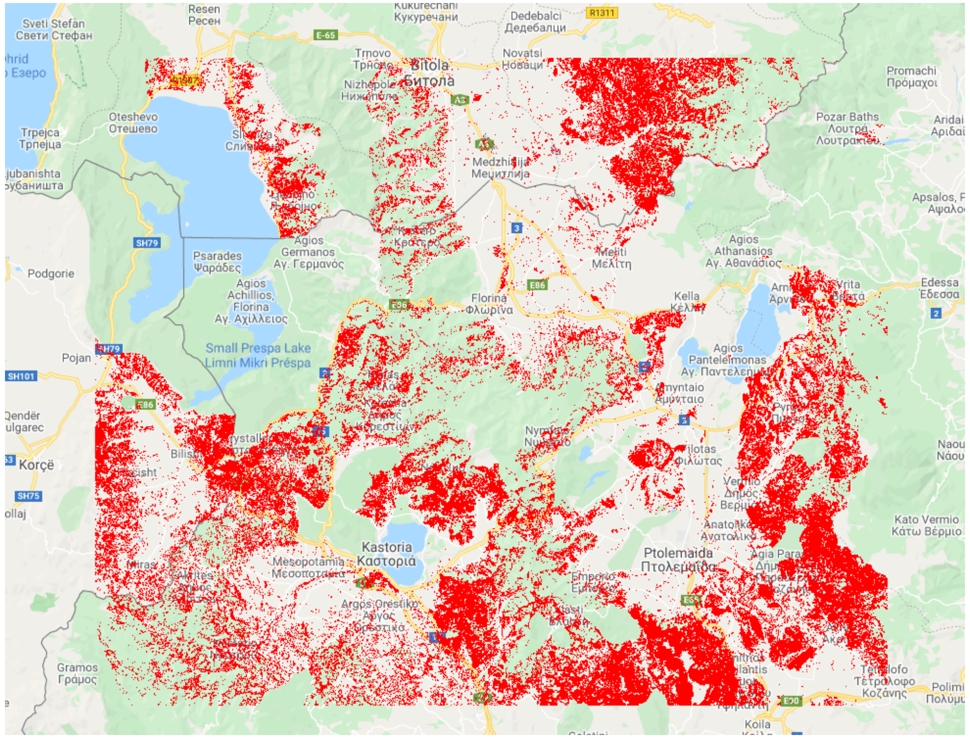


ML training:
2.4 validation polygons

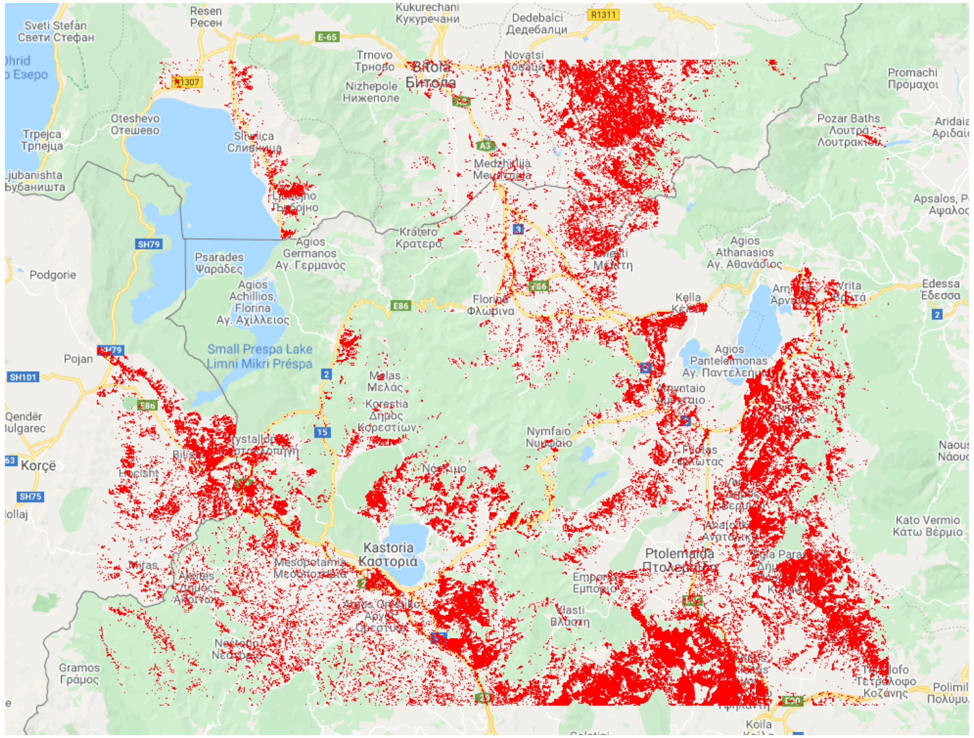


Case Study

ML training: 2.3 ML_Hard_Thresholds



ML training: 2.4 validation polygons



Case Study

Accuracy Assessment

- Increase in accuracy
- Decrease in OOB Error Estimate

Country	Overall Accuracy		Kappa		OOB Error Estimate		Split
Task	2.3	2.8 (1)	2.3	2.8 (1)	2.8	2.8 (1)	
Greece	*0.715*	*0.777*	*0.41*	*0.54*	0.26	*0.16*	*0.1*
Spain	*0.828*	*0.861*	*0.65*	*0.66*	0.23	*0.16*	*0.5*

Notes

- Good input -> good output
- Only small amount of data can suffice

Conclusions

- Identifying MLs with Earth Observation data is a challenging task
  - But possible!
- Major achievements:
  - Preparation time reduced
  - Futureproof analysis
- Tool is capable of providing accurate results.
  - MLs detection precision, augmented.

References

- Breiman, L. (1999). *Random Forests-Random Features*.
- Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, *40*(13), 5125–5131. https://doi.org/10.1016/j.eswa.2013.03.019
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How Many Trees in a Random Forest? *Machine Learning and Data Mining in Pattern Recognition*, *7376 LNAI*, 154–168. https://doi.org/10.1007/978-3-642-31537-4_13

Acknowledgments

- I’d like to thank my MAIL colleagues/friends who supported me for the realization of this Task
  - Namely, Marek, Fernando, Dzhaner, Pablo and Sebastian for the technical support,
  - Ewa for the Literature Review and preliminary analysis, and
  - Michal for the supervision-coordination

Thank you for your attention!

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823805

Georgios Spanos, geospan90@gmail.com