A Walk-Through on Machine Learning Techniques for Sentinel Big Data Fusion

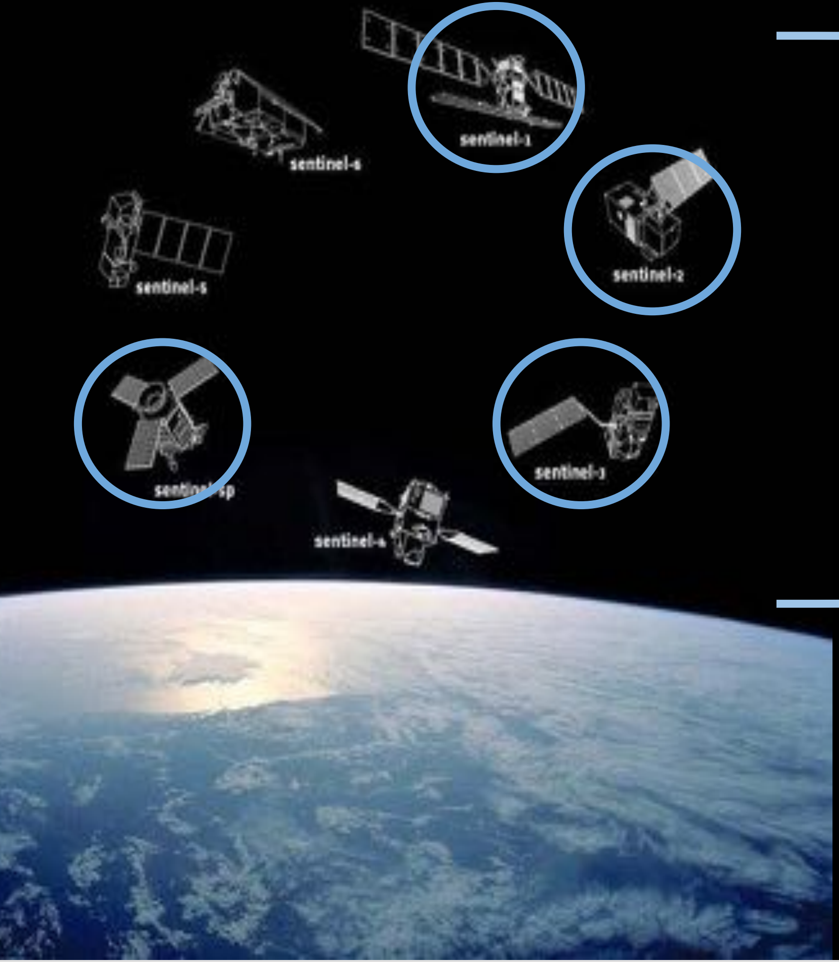
Sara Aparício

ESA ESRIN, EOP-Φ

Φ-Week - 12th November 2018

@_SaraAparicio_

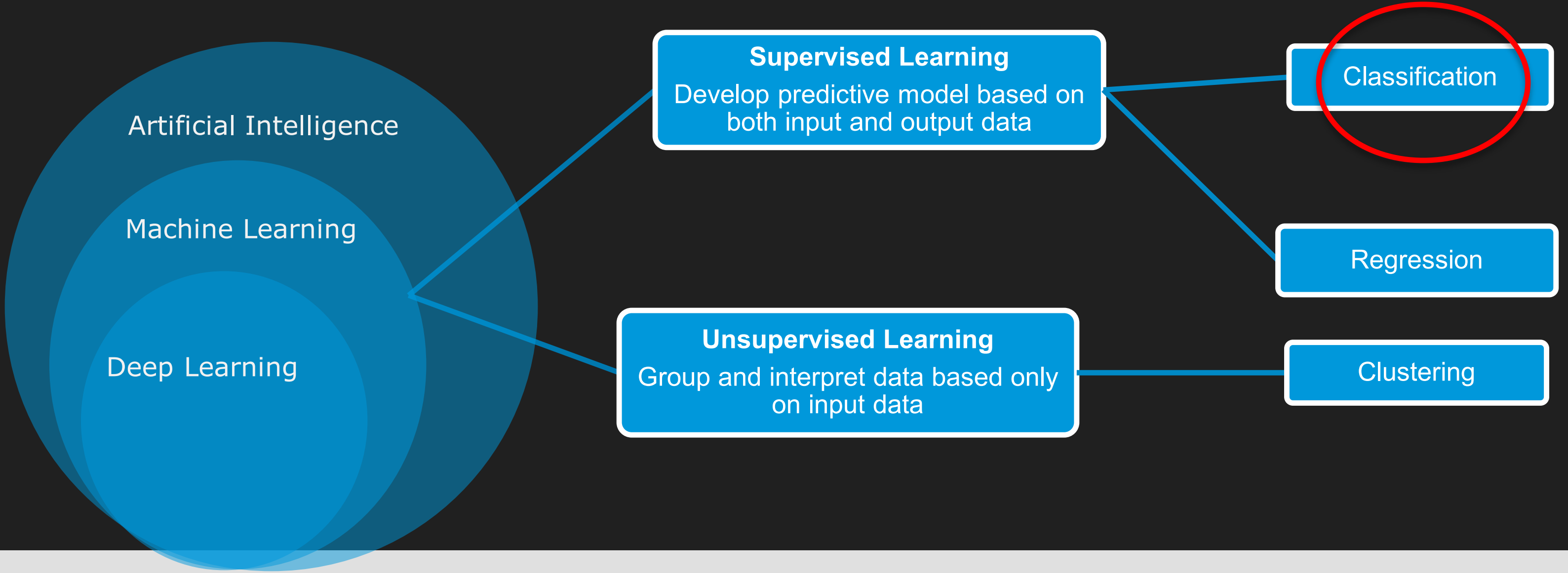
The world’s most comprehensive suite of dedicated EO missions.



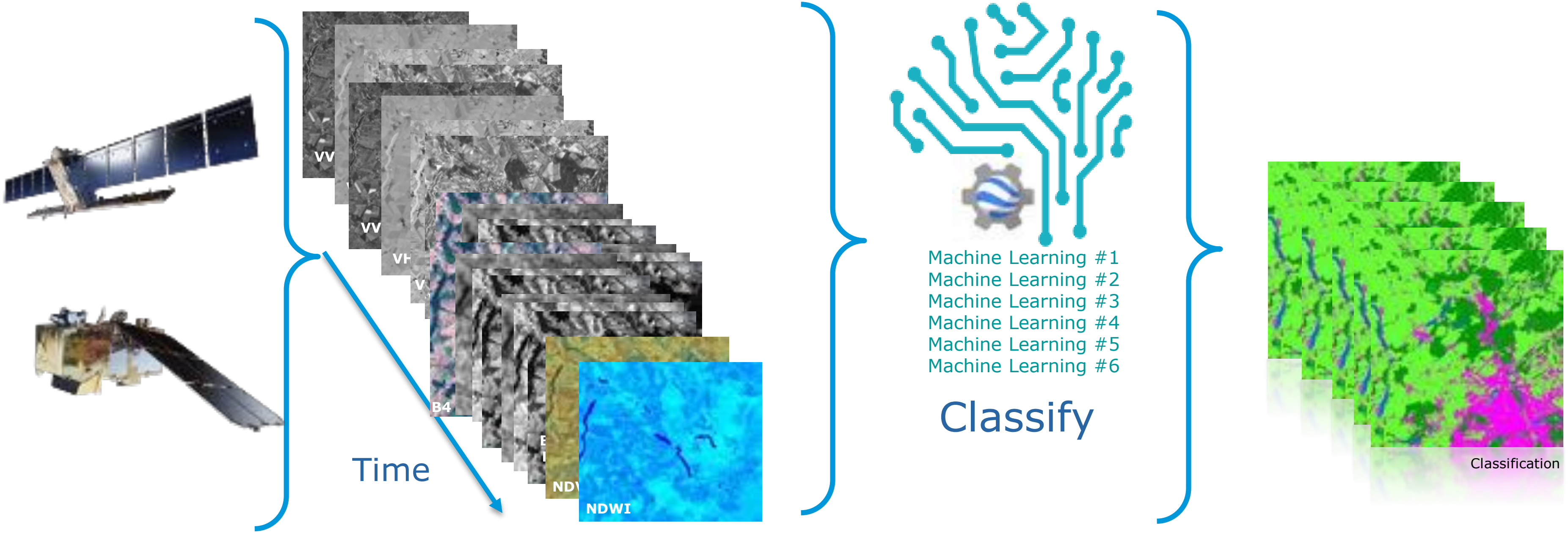
European Space Agency

Major trends

a) The advent of cloud computing & increase computing power,
b) proliferation of open-access satellite data streams,
c) growing use of machine-learning algorithms



Sentinel data fusion with Machine Learning Techniques

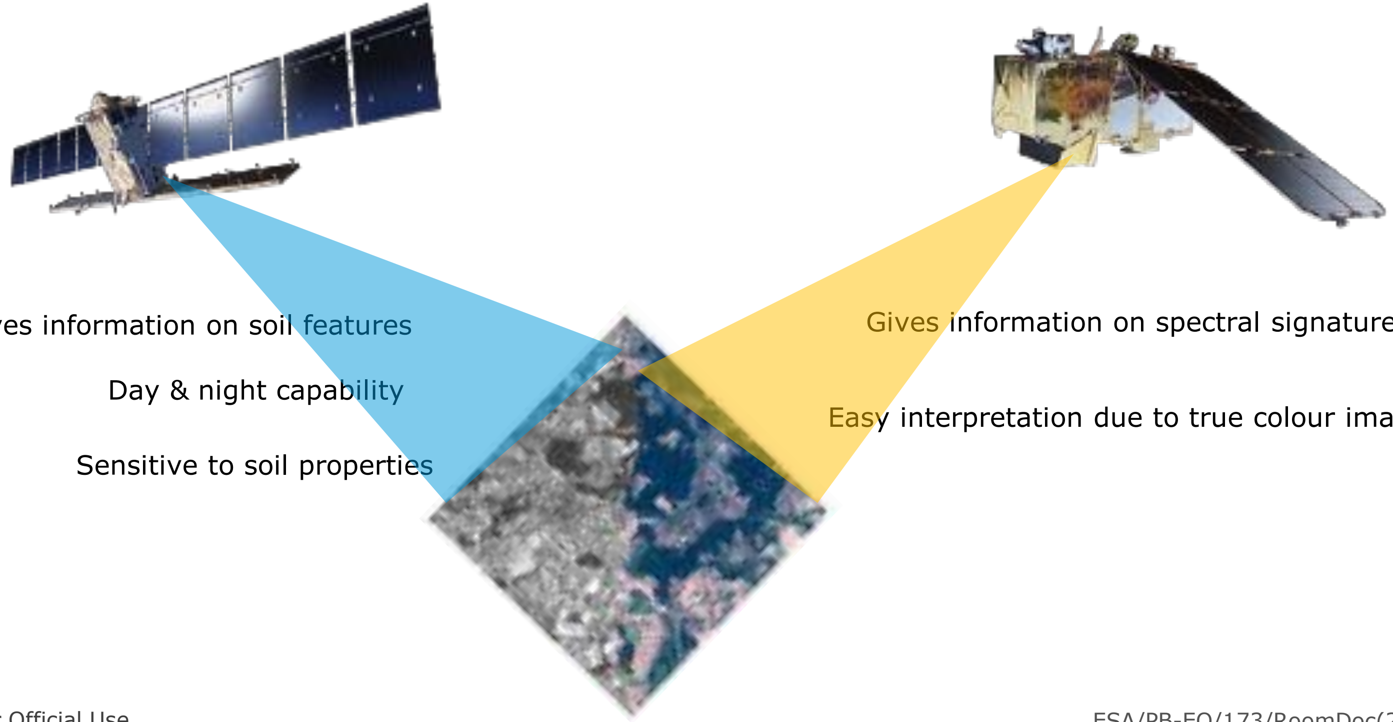


Multisensor/multitemporal data

Best technique?
Best band selection?

Motivation I – Merging different datatypes

A promising direction of machine learning in Earth Observation is its pairing with data fusion



Motivation 2 – AI4EO the white paper


Towards a European AI for Earth Observation
Research & Innovation Agenda
Proceedings of a workshop at ESA Φ-lab

This document addresses elements of what needs to be done at European level, some … potential of artificial i… data and the data … pply chain and vice versa.

It captures the recommendations of a community-led AI4EO workshop held at ESA/ES… list in Annex 1). The aim … ssess progress in the de… techniques to the world of EO and to explore the potential value of a concerted **Research and Innovation (R&I)** effort on this topic at **European l…** … be a *dynamic* report captur… community. It will be revi… workshop at ESA/ESRIN on November 14th, 2018…….(….)

Classification/ Recognition
- updating land-cover maps

Detection
- large scale in automatic basis
- use SAR in machine learning

Data Fusion
- Merging diverse EO data

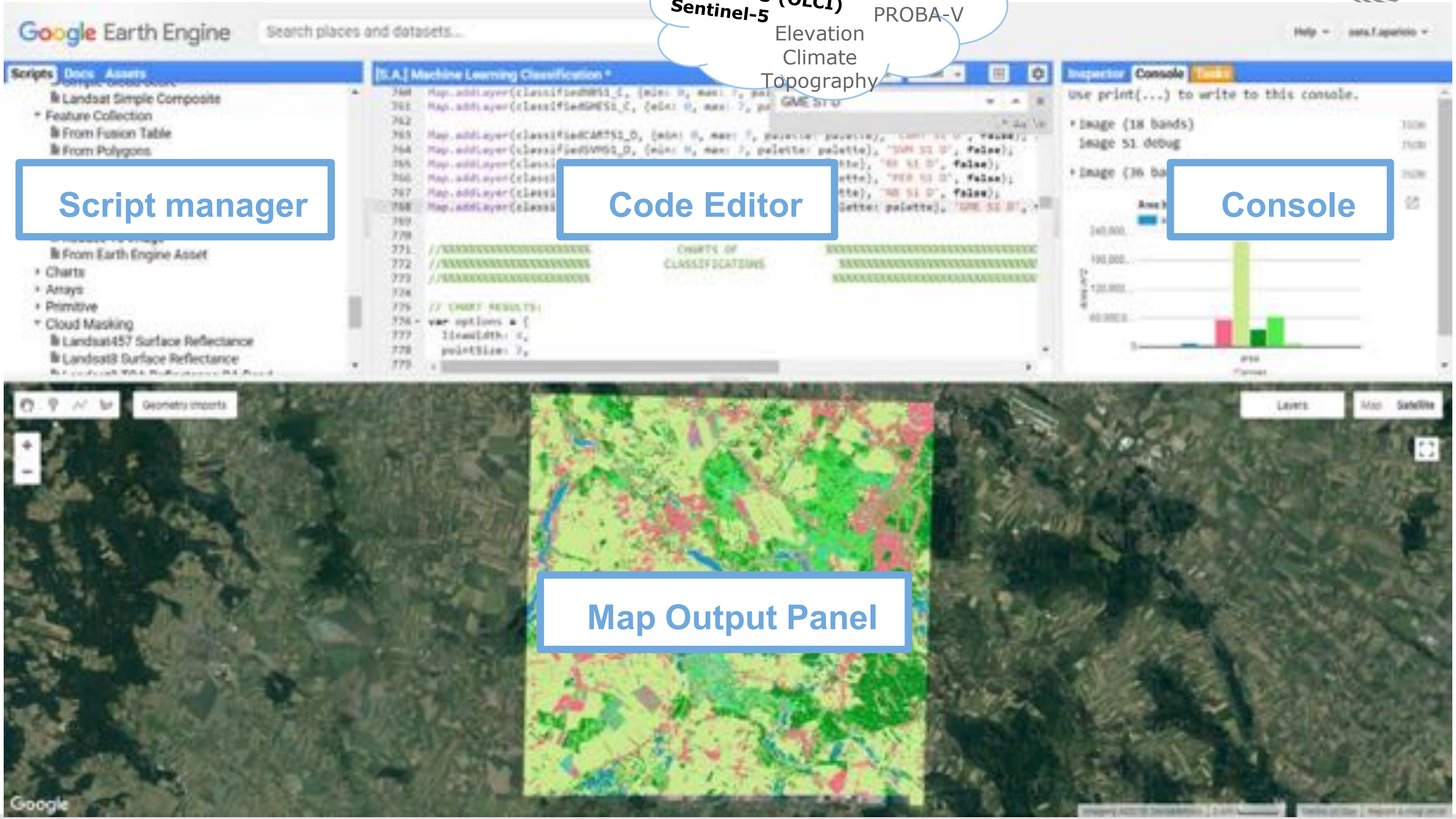
Objectives/Goal

1. To develop **a pixel-based classification, reproducible, scalable with a machine learning-based approach of large-area mapping**/land cover of high resolution (10m) based on a multi-sensor & multi-temporal approach;

2. To evaluate the **additive value** of open-access satellite optical and radar variables, processed using cloud computing, to a topographic baseline model.

3. To explore/understanding/address efficiency of Google Earth Engine to effectively execute **big data workflows** using machine learning techniques on Google Earth Engine (and accuracy) for multi-temporal land use mapping.

Goole Earth Engine

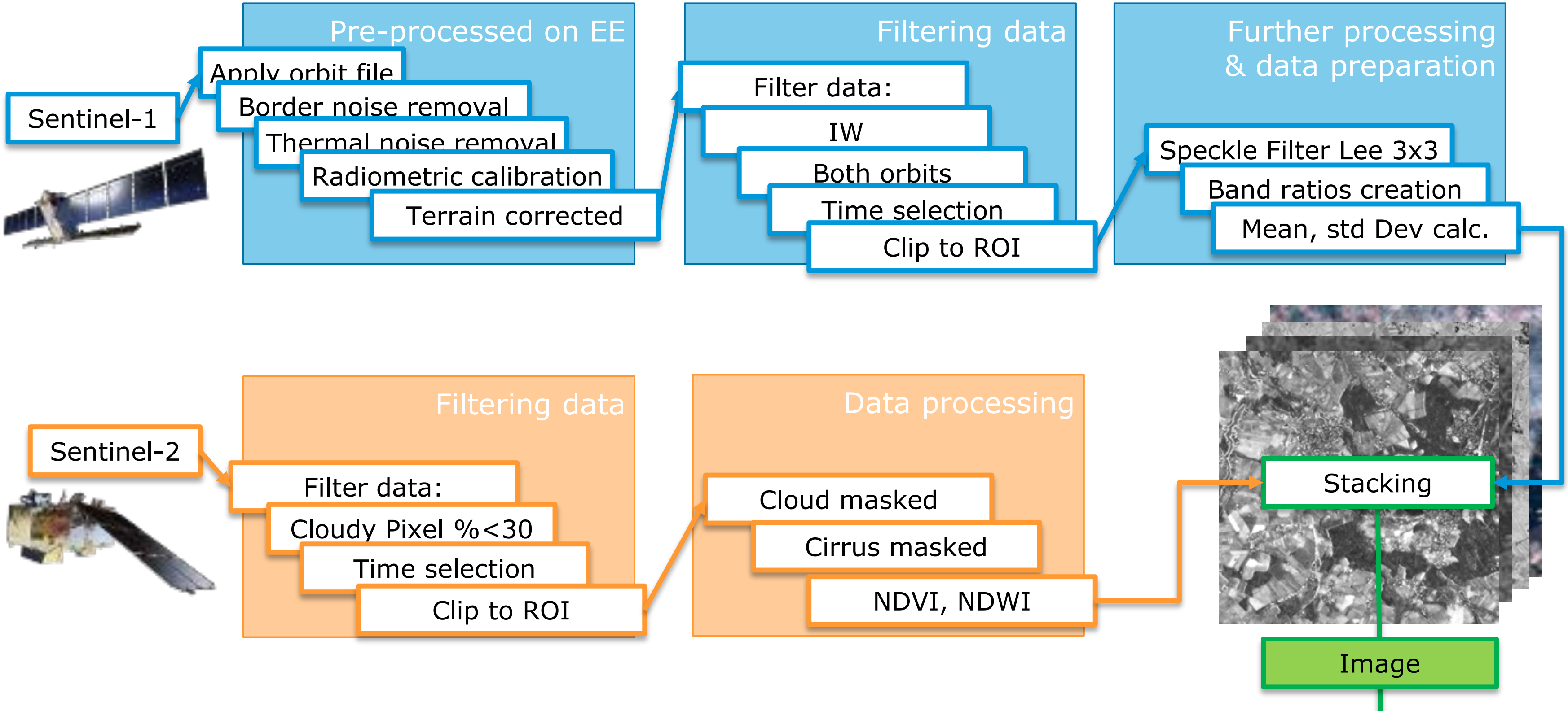


METHODOLOGY

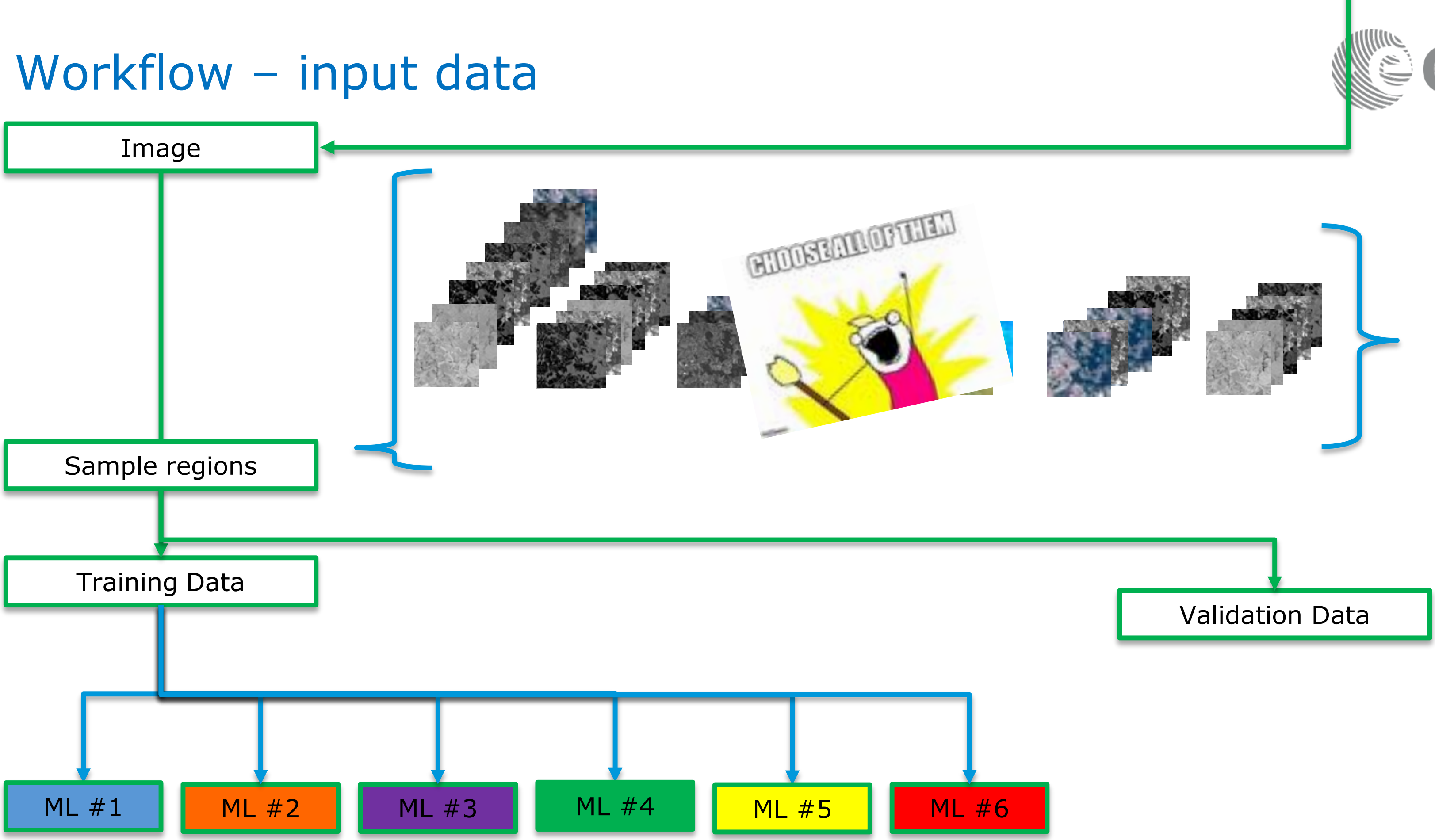
Case study


West Wieljopolska
Landcover classes
Agriculture, Coniferous forest, Mixed Forest, Grassland, Bare soil, Wetland, Urban fabric, Water body.
(Sentinel-2 Natural Color)

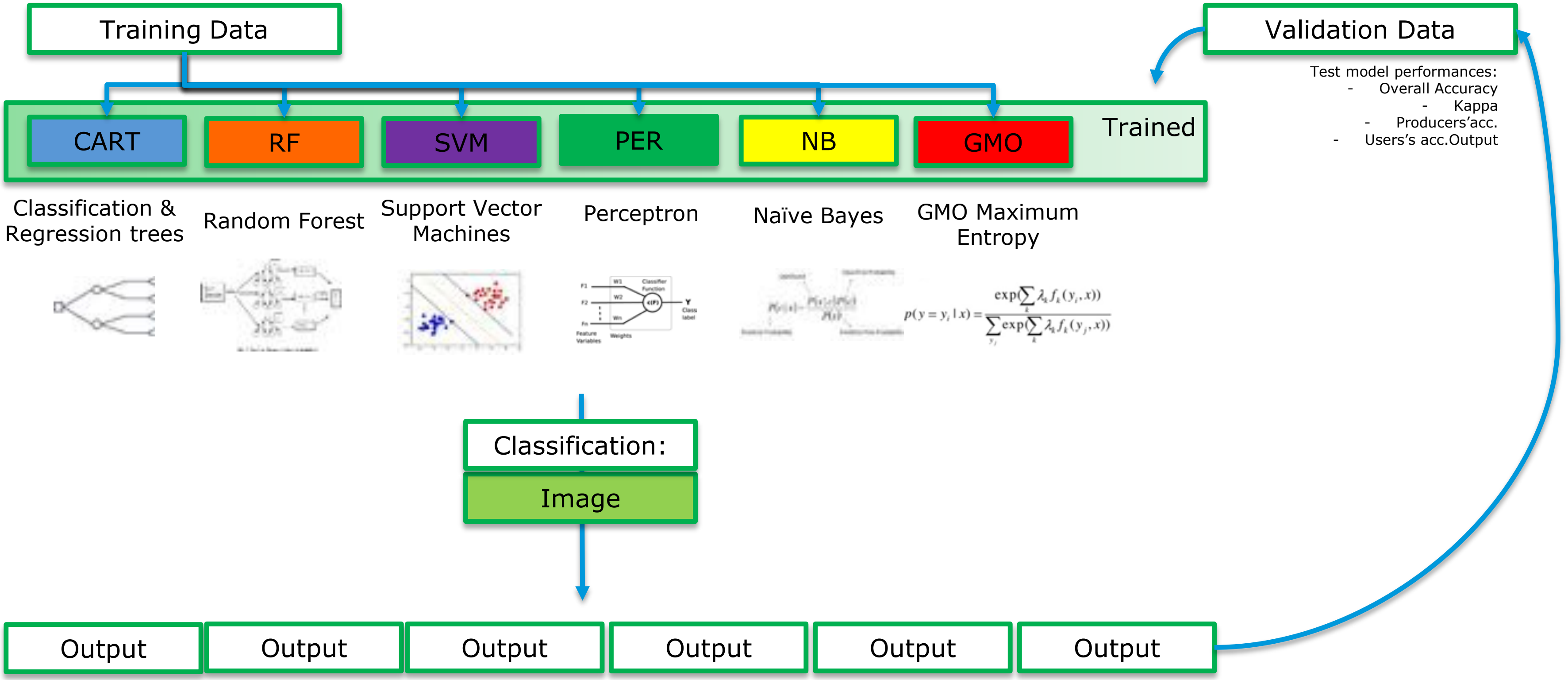
Workflow – data processing & preparation



Workflow – input data

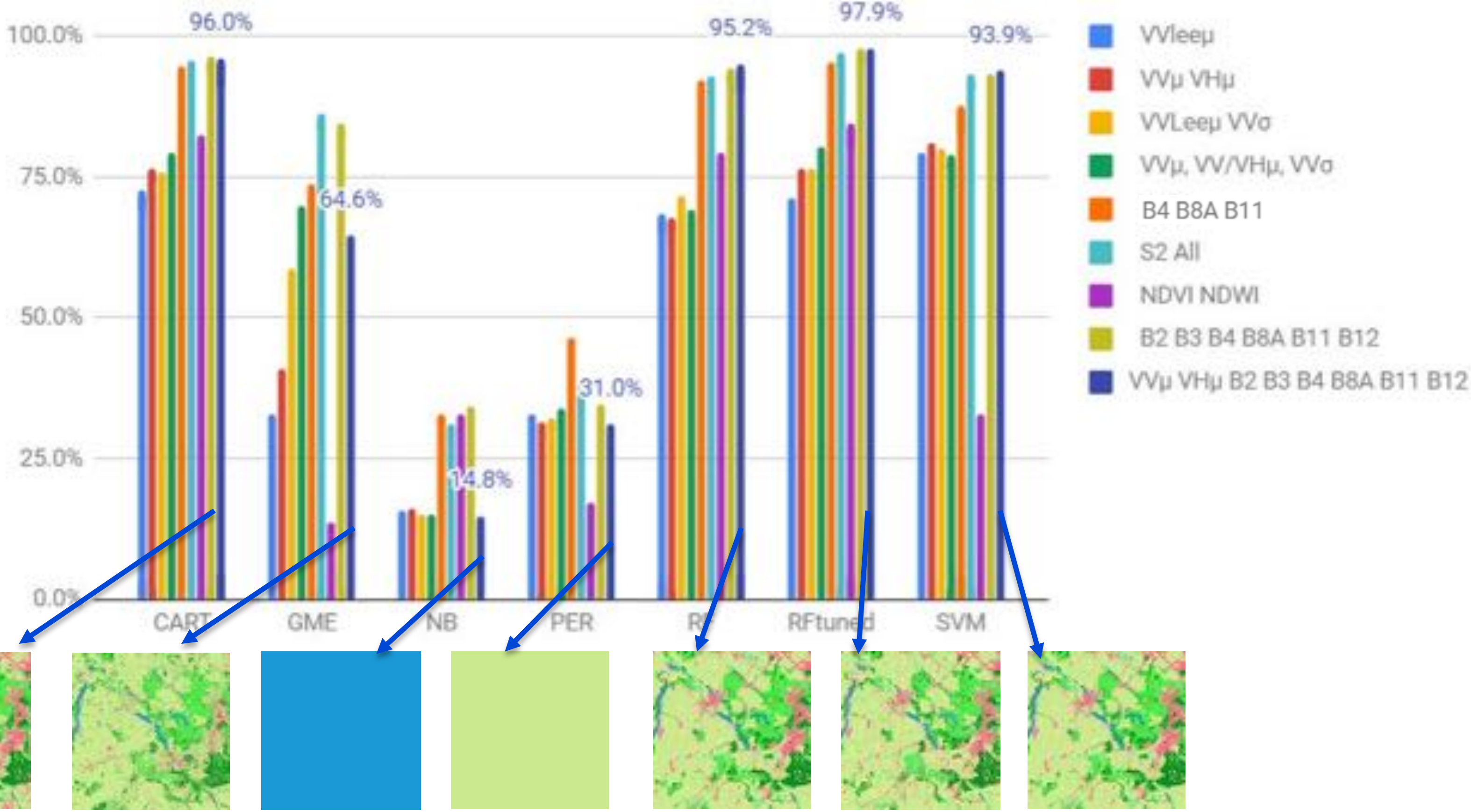


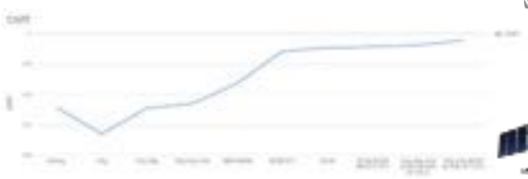
Workflow – classification

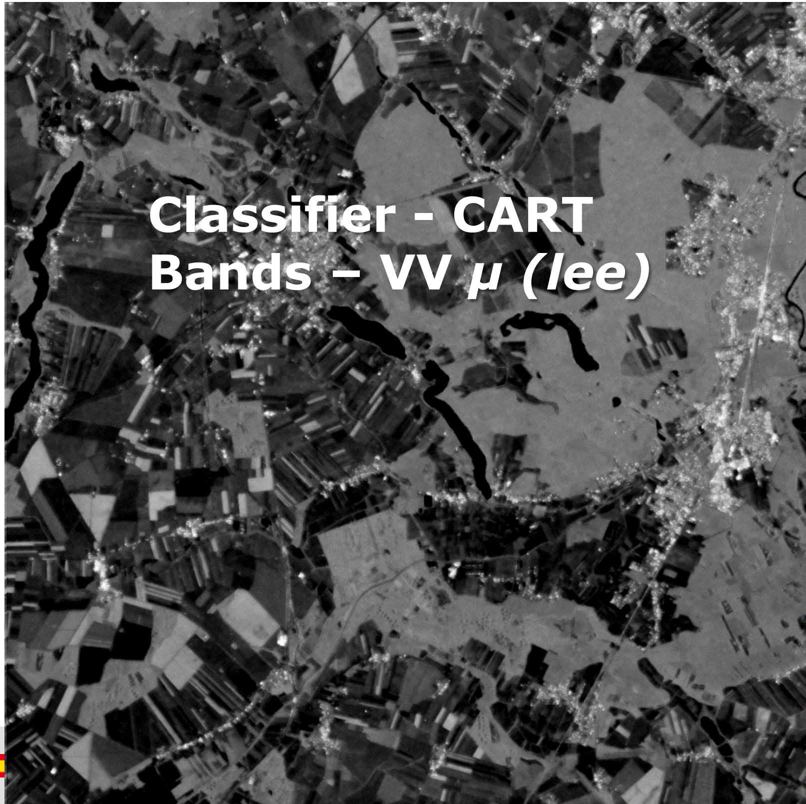


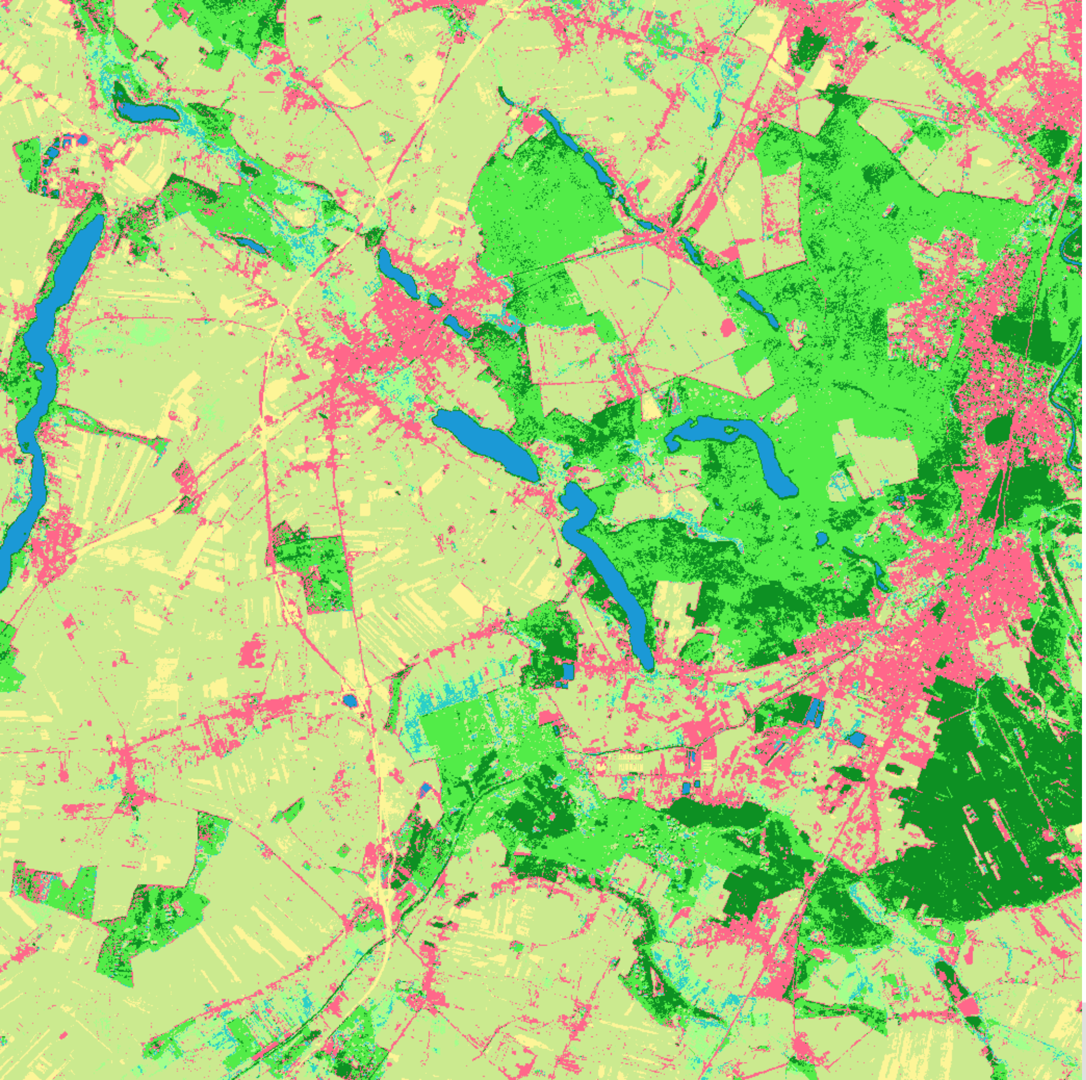
RESULTS OVERVIEW

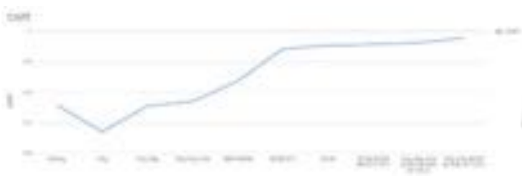
Machine Learning performances – a first overview!

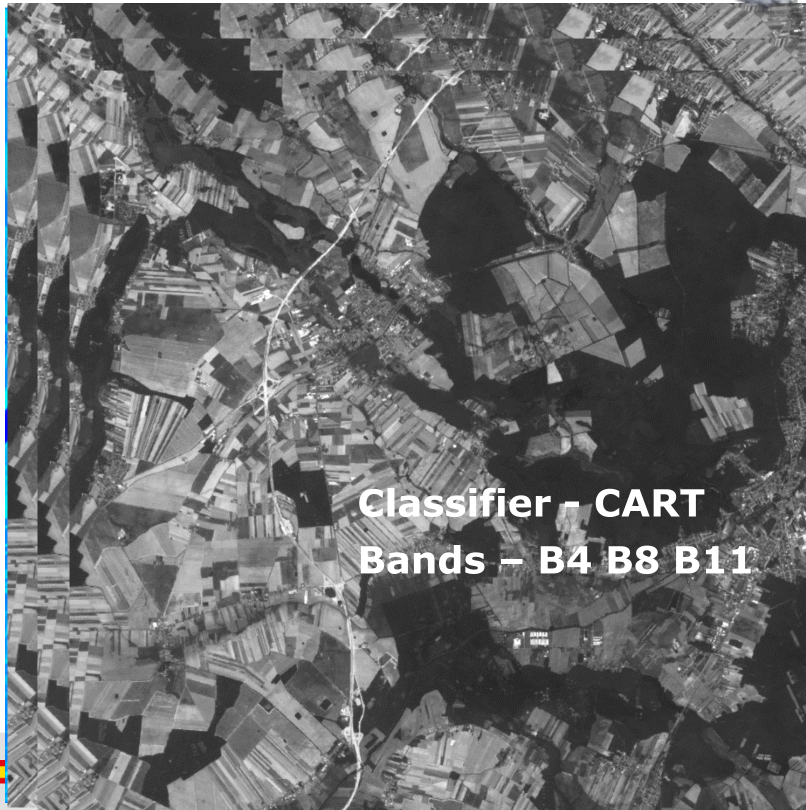


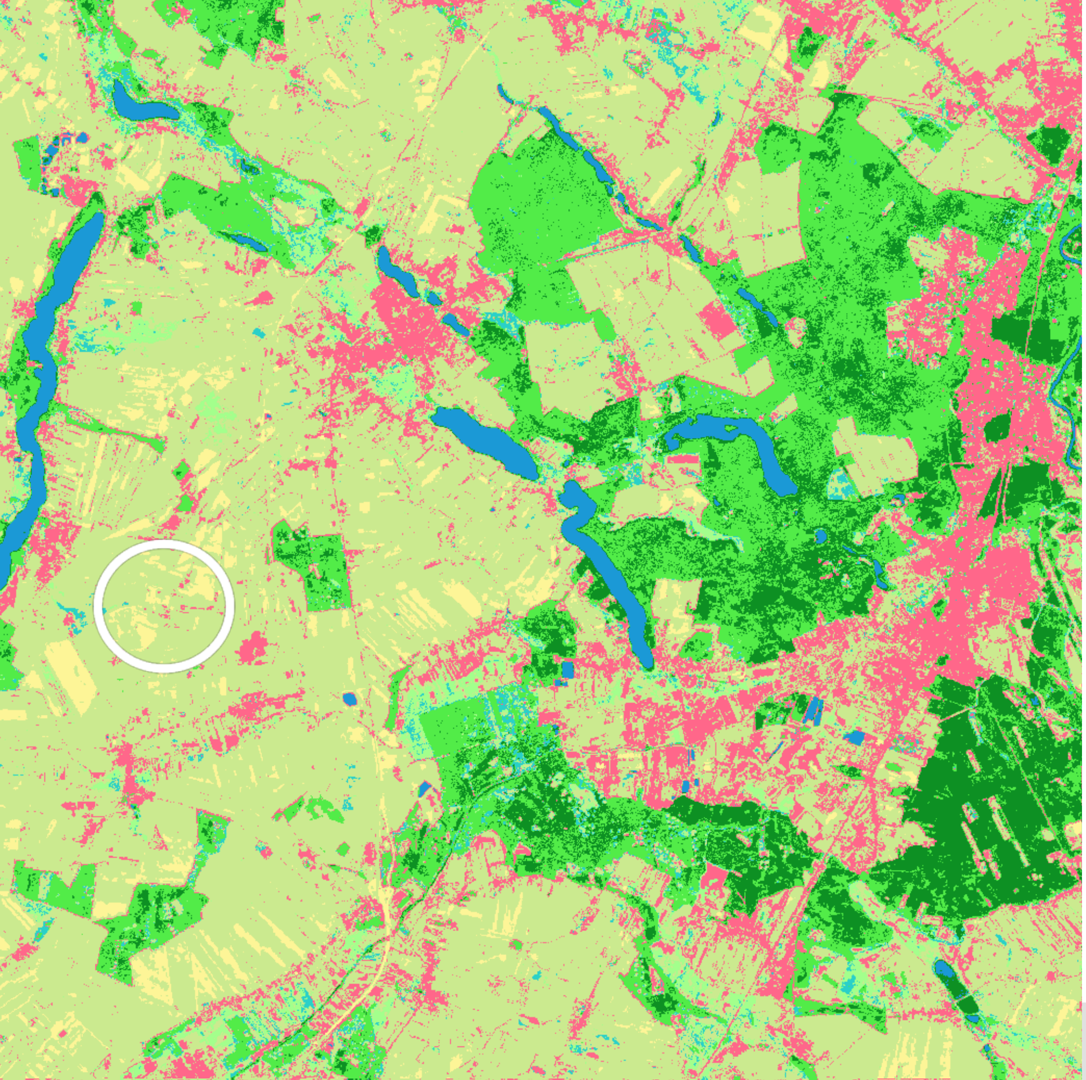


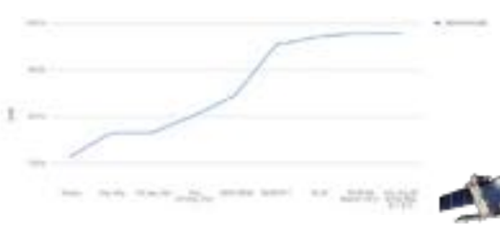

Classifier - CART
Bands – VV μ (lee)





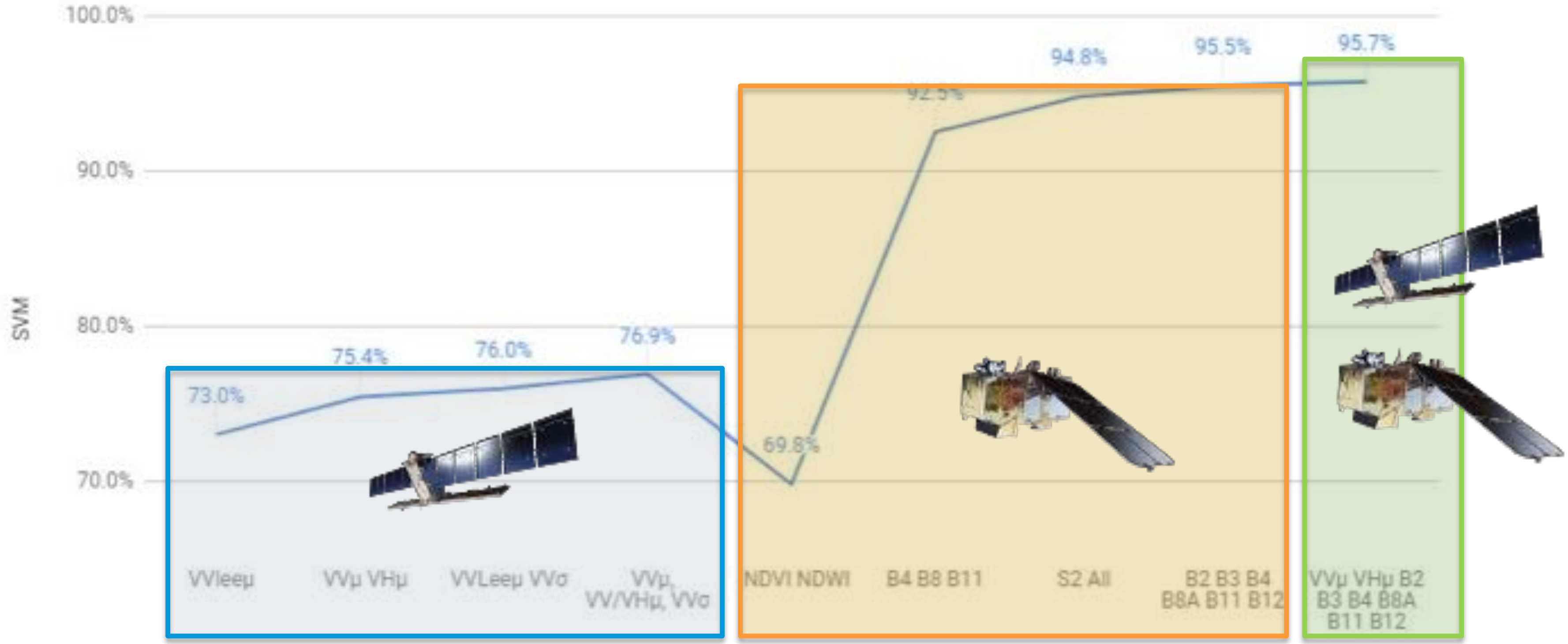

Classifier - CART
Bands – B4 B8 B11





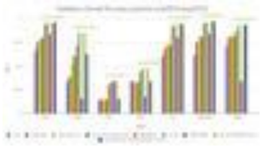
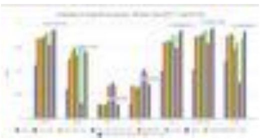
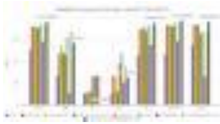

Classifier - RF
Bands – S1 + S2 bands
Classifier - RF
Bands – All except...
Classifier - RF
Bands – All S2 bands
Classifier - RF
Bands – B4 B8 B11
Classifier - RF
Bands - NDVI NDWI

Behaviour to data input
Averaged from best performing models



Improving model performances…

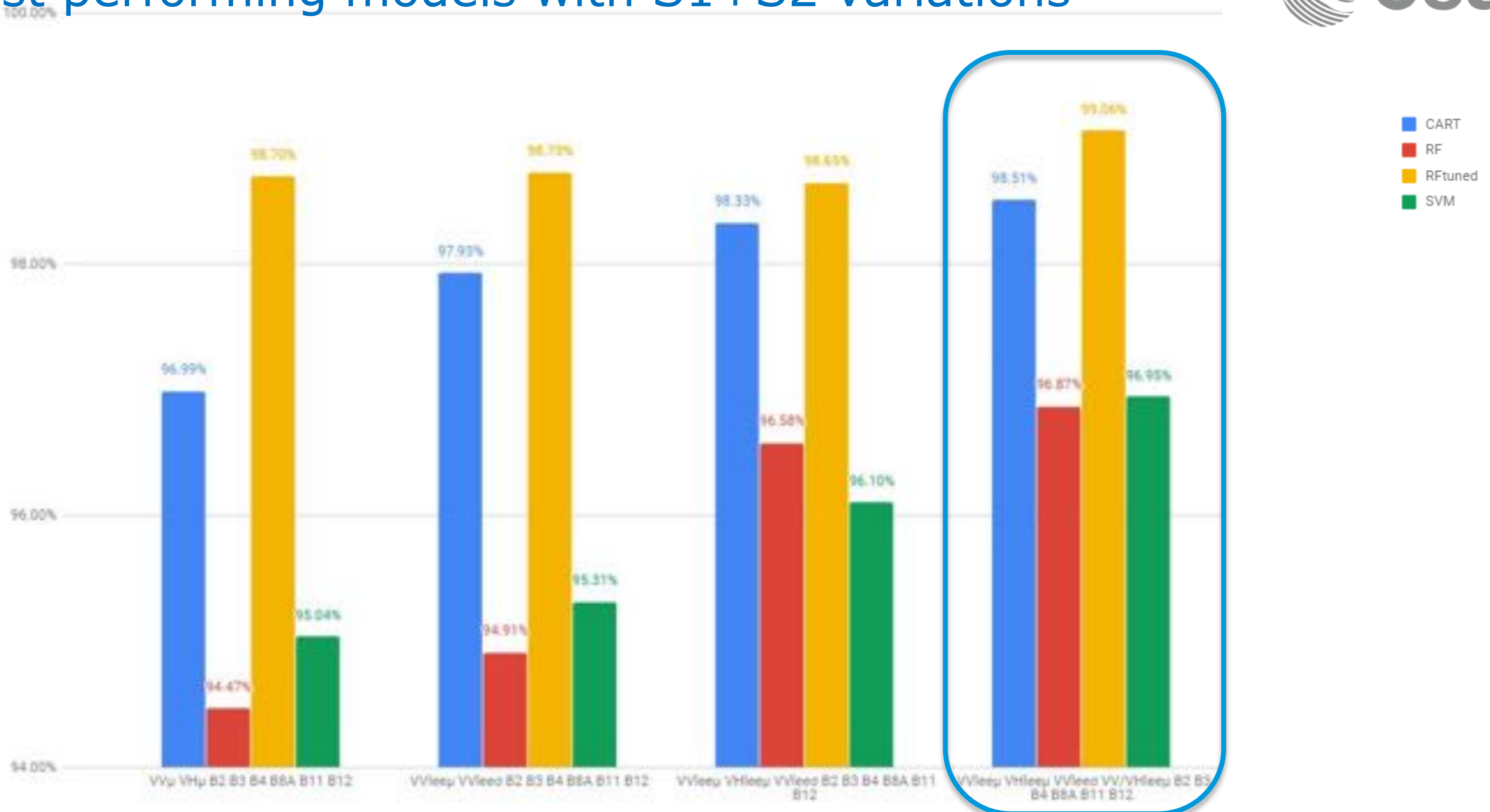
Changing temporal coverage
- updating land-cover maps

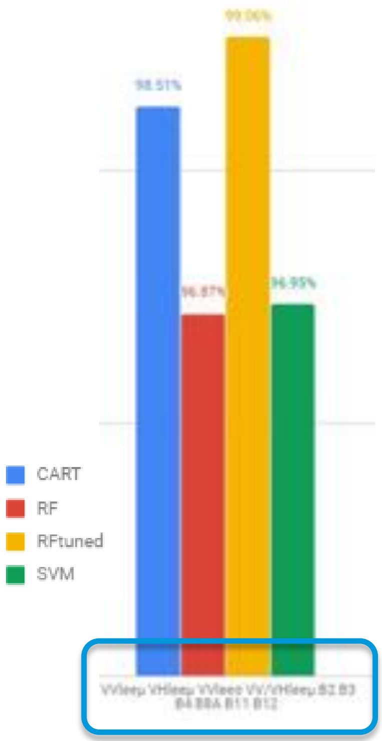
Changing training data
In size or coverage

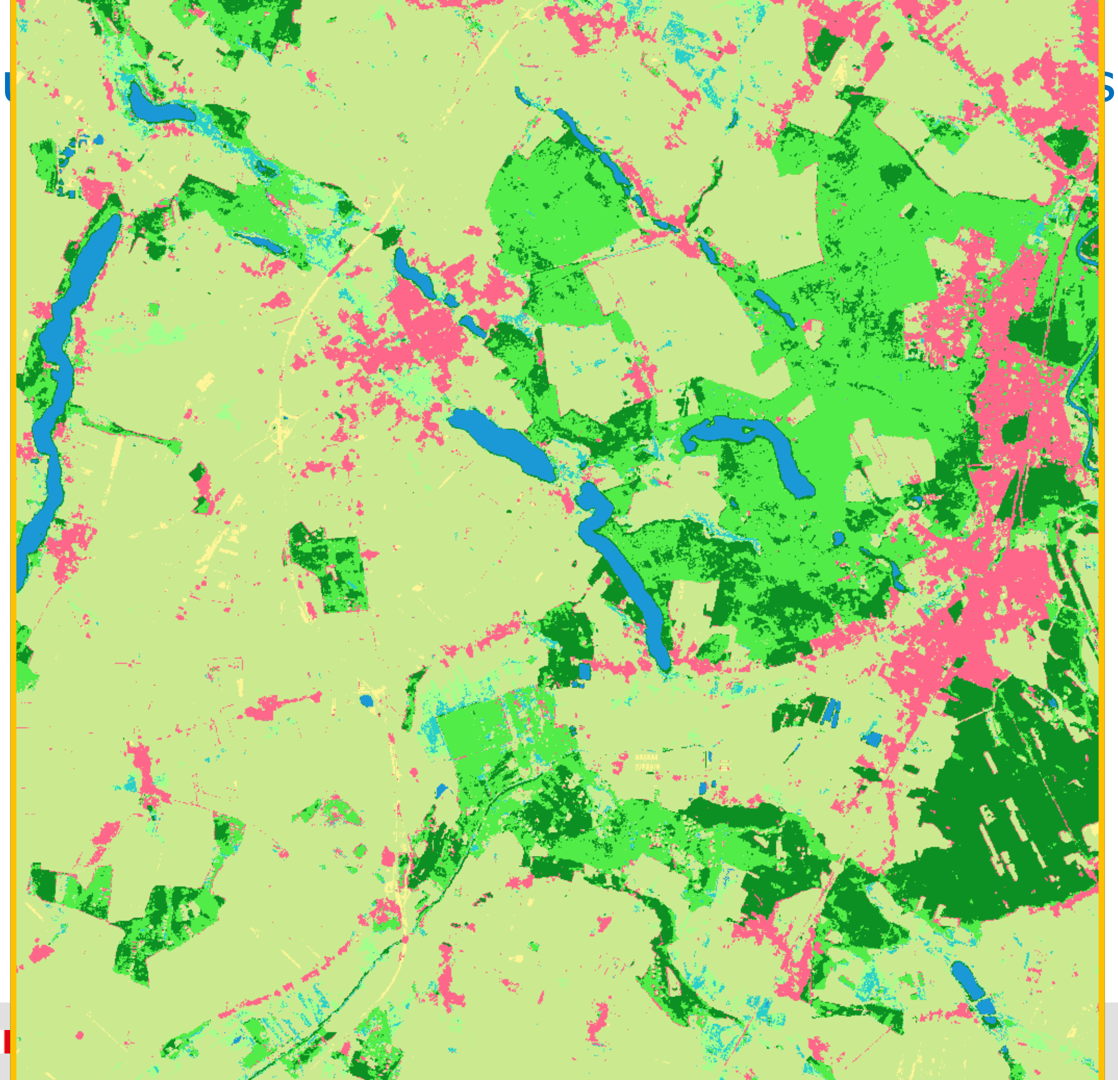
Trying new band selection
Different combinations of S1+S2

Best performing models with S1+S2 variations



Other accu s





Random Forest (Tuned)

CART

Support Vector Machines

Random Forest

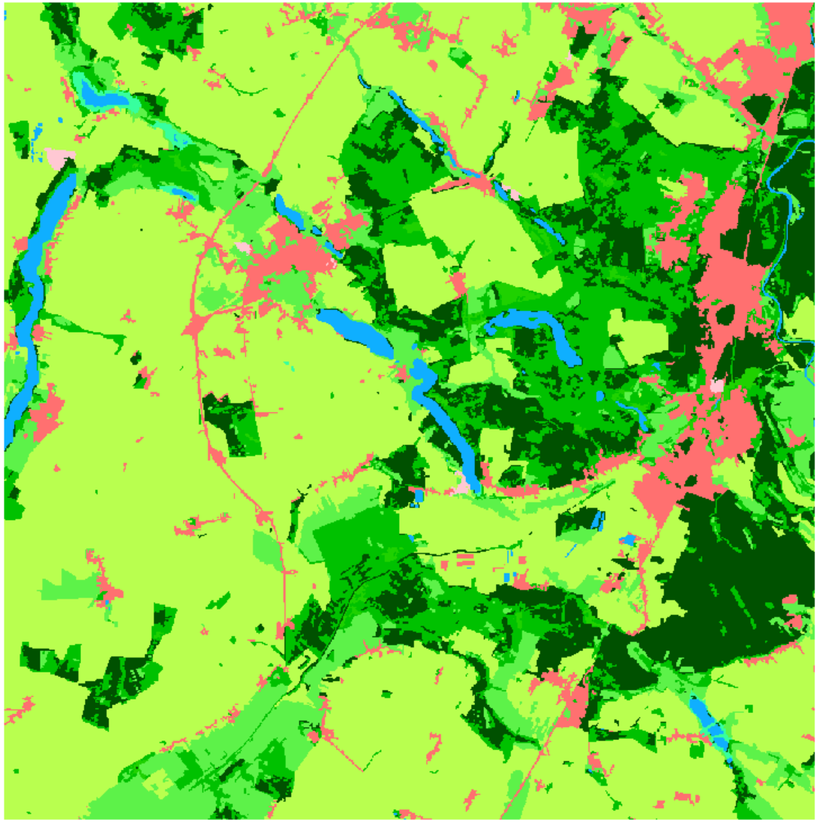
Further accuracy metrics…

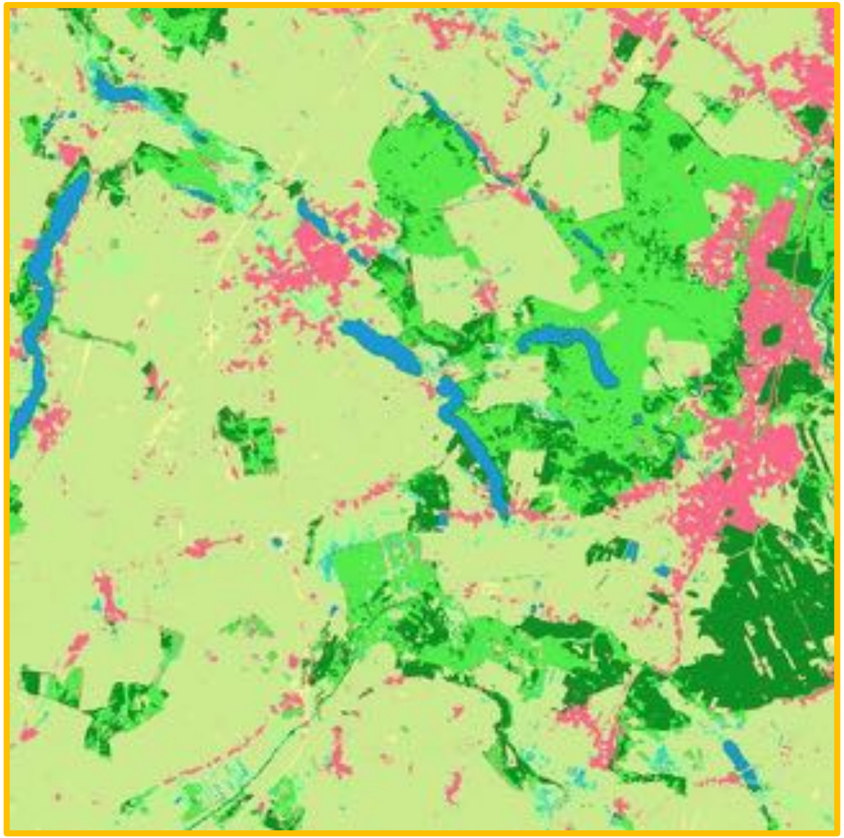
	Reference									
Classified	Water body	Wetland	Urban Fabric	Agriculture	Coniferous Forest	Mixed Forest	Grassland	Bare soil	PA	UA
Water body	193	0	0	0	0	0	0	0	1.000	1.00
Wetland	0	29	0	1	0	0	0	0	0.967	0.94
Urban Fabric	0	0	239	2	0	0	0	0	0.992	1.00
Agriculture	0	1	0	428	0	1	0	0	0.995	0.99
Coniferous Forest	0	0	0	0	154	2	0	0	0.987	0.99
Mixed Forest	0	0	0	0	2	157	0	0	0.987	0.98
Grassland	0	1	0	1	0	1	60	0	0.952	1.00
Bare soil	0	0	0	0	0	0	0	5	1.000	1.00

Kappa
0.9881

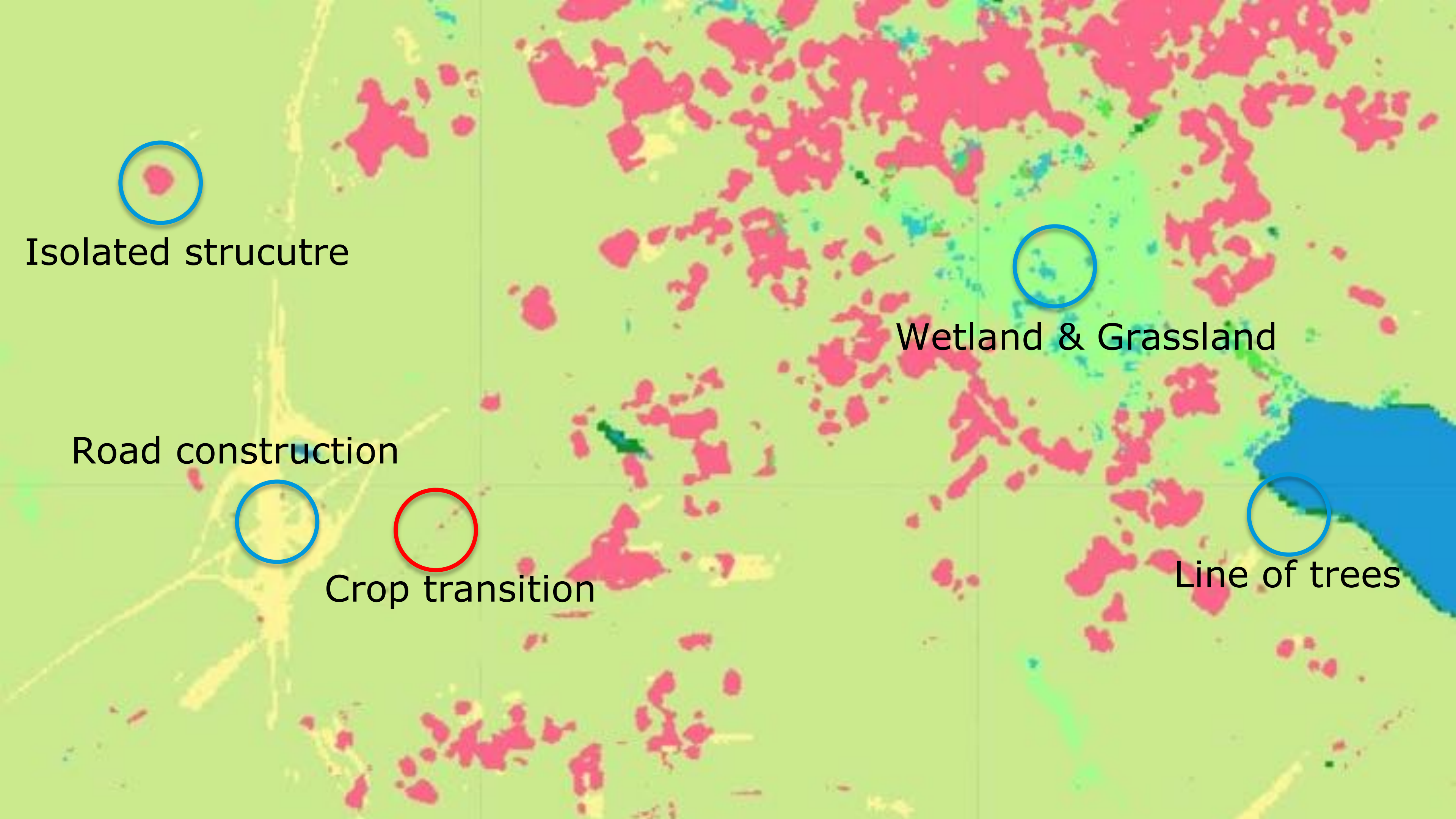
Overall Accuracy
0.9906

Ground Truth/Reference VS Final model





waterbody
wetland
urbanfabric
agriculture
coniferousforest
mixedforest
grassland
baresoil


Isolated strucutre
Wetland & Grassland
Road construction
Crop transition
Line of trees

Limitations…bear in mind that….

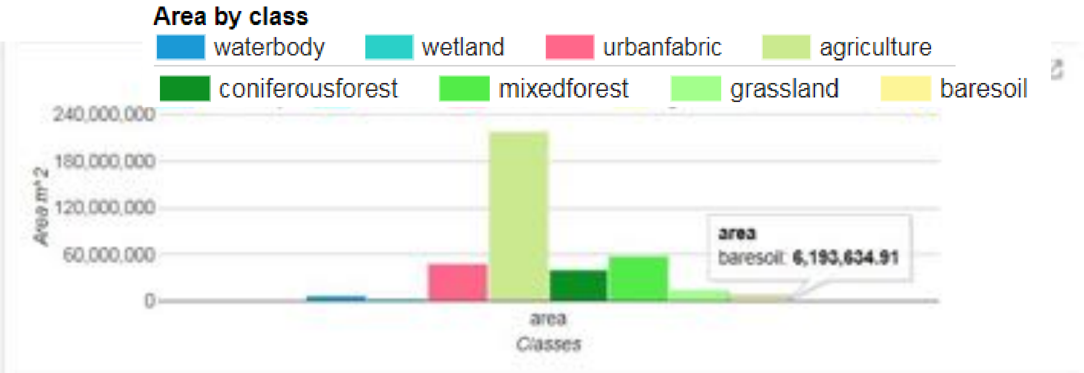
So…. Is this then the best model?

- Training data (plays a **huge** role) affecting:
  - Models performances
  - Smaller size – the more data input…not always is the best
  - Bigger size – more difference in 'additive power' but models behaved + similarly
- Tuning hyperparameters of ML are **done manually**
  - Some models were inserted by default & data was not normalized


What potential...?

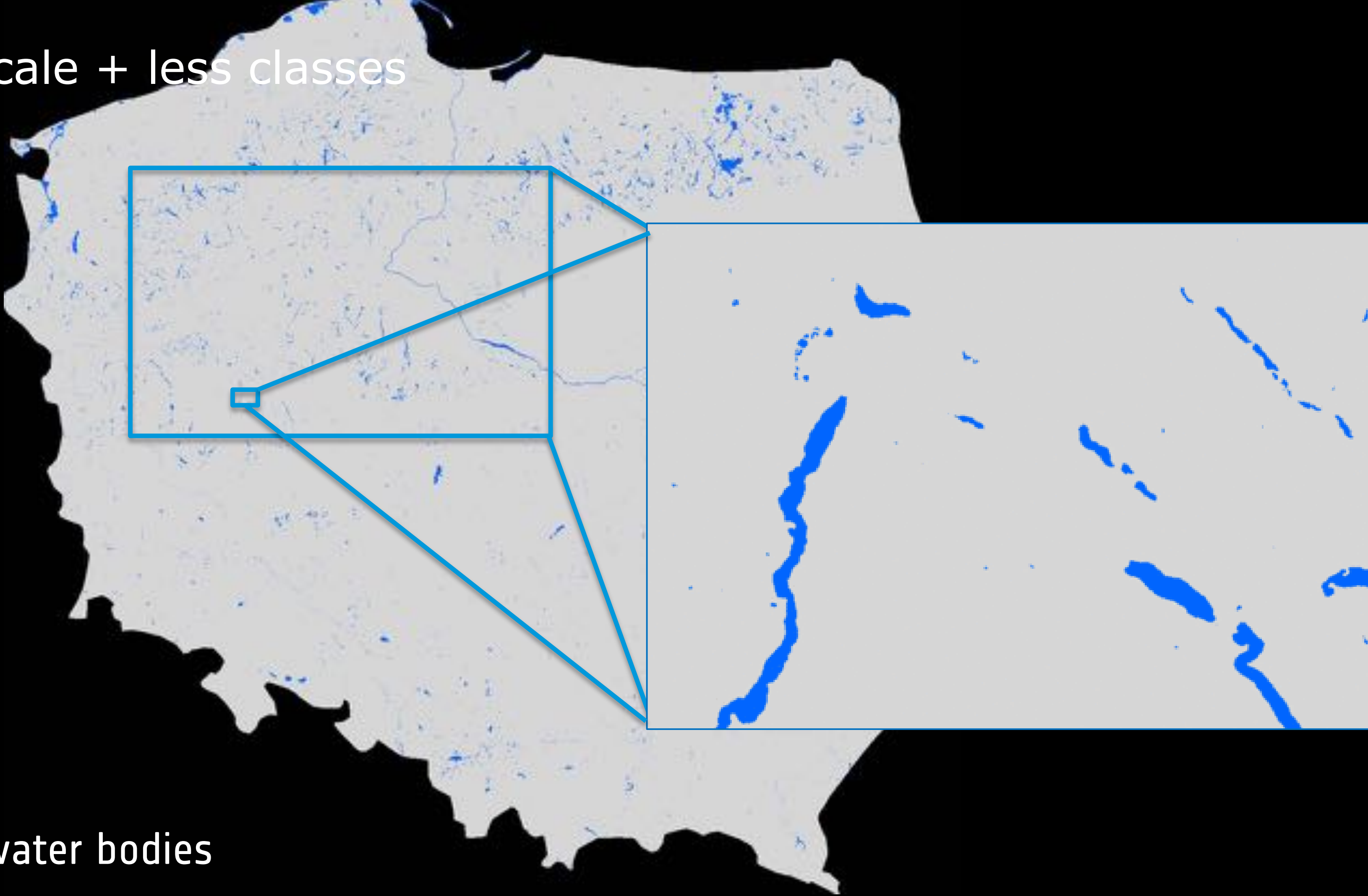
Example of straightforward applications…

- Analysis of temporal land use and land cover change…



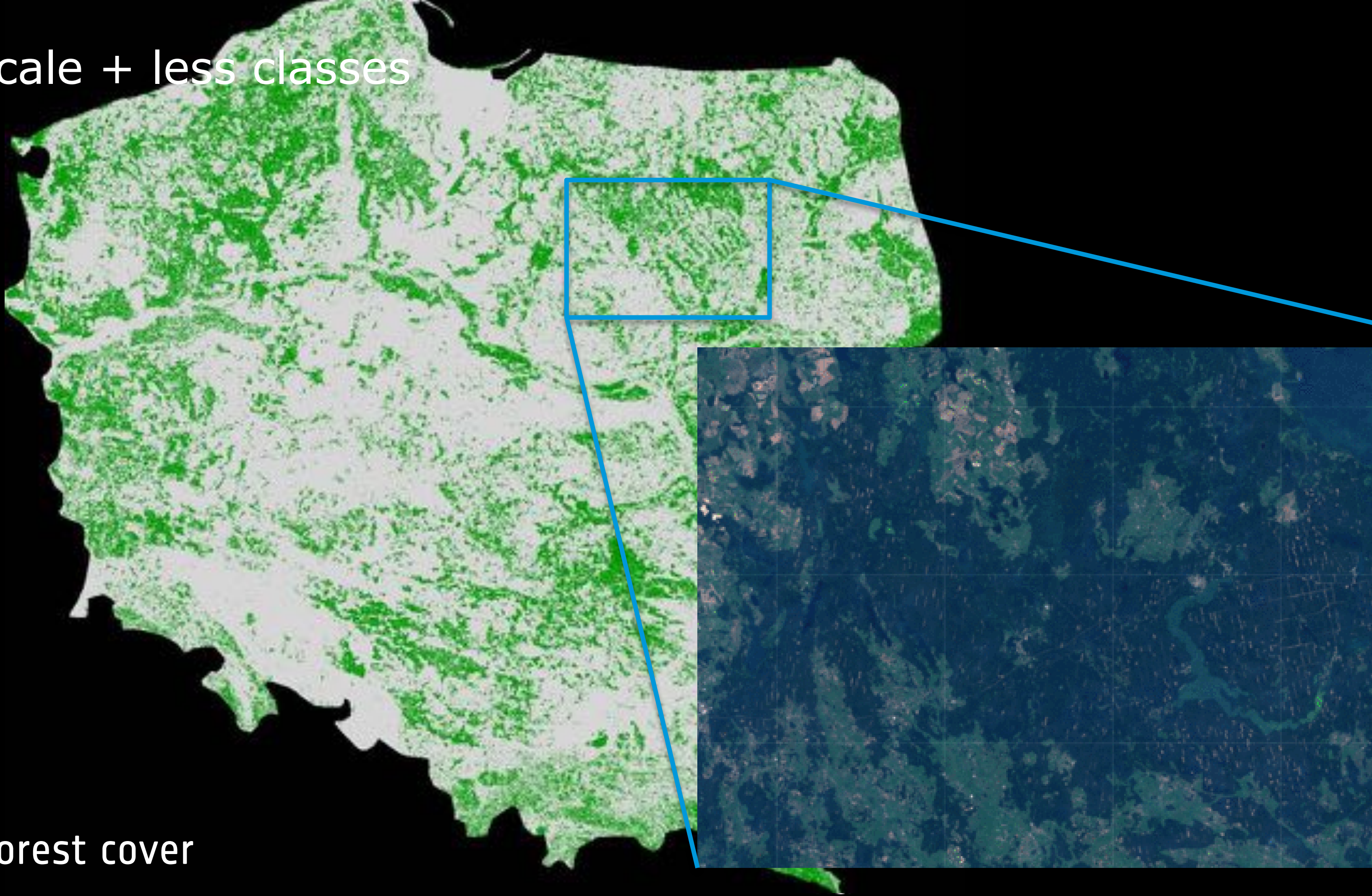
Attention!
This graph moves ☺

Larger scale + less classes



Mapping water bodies

Larger scale + less classes



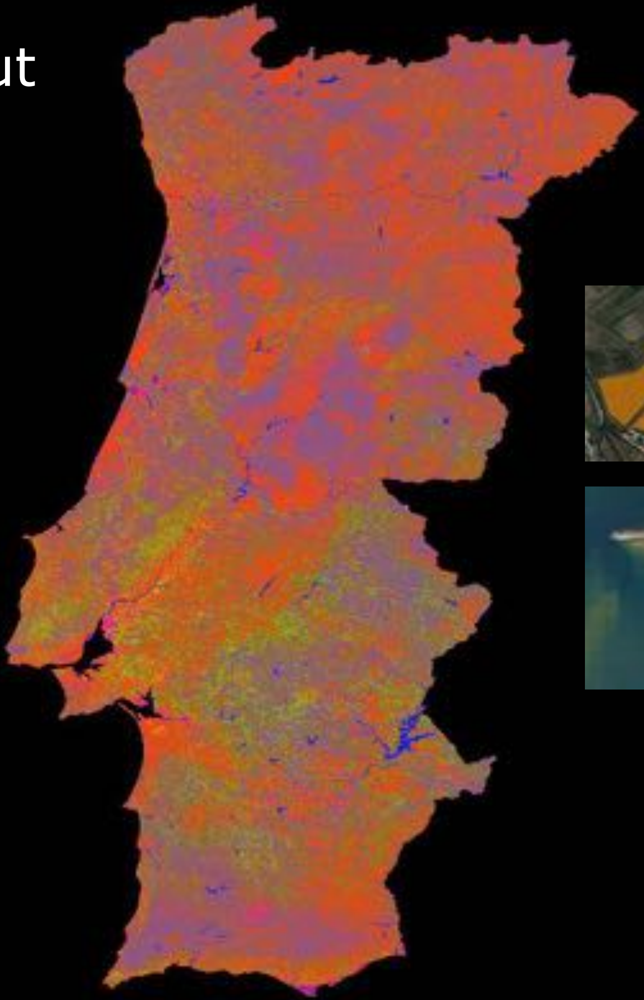
Mapping forest cover

Large scale + data input

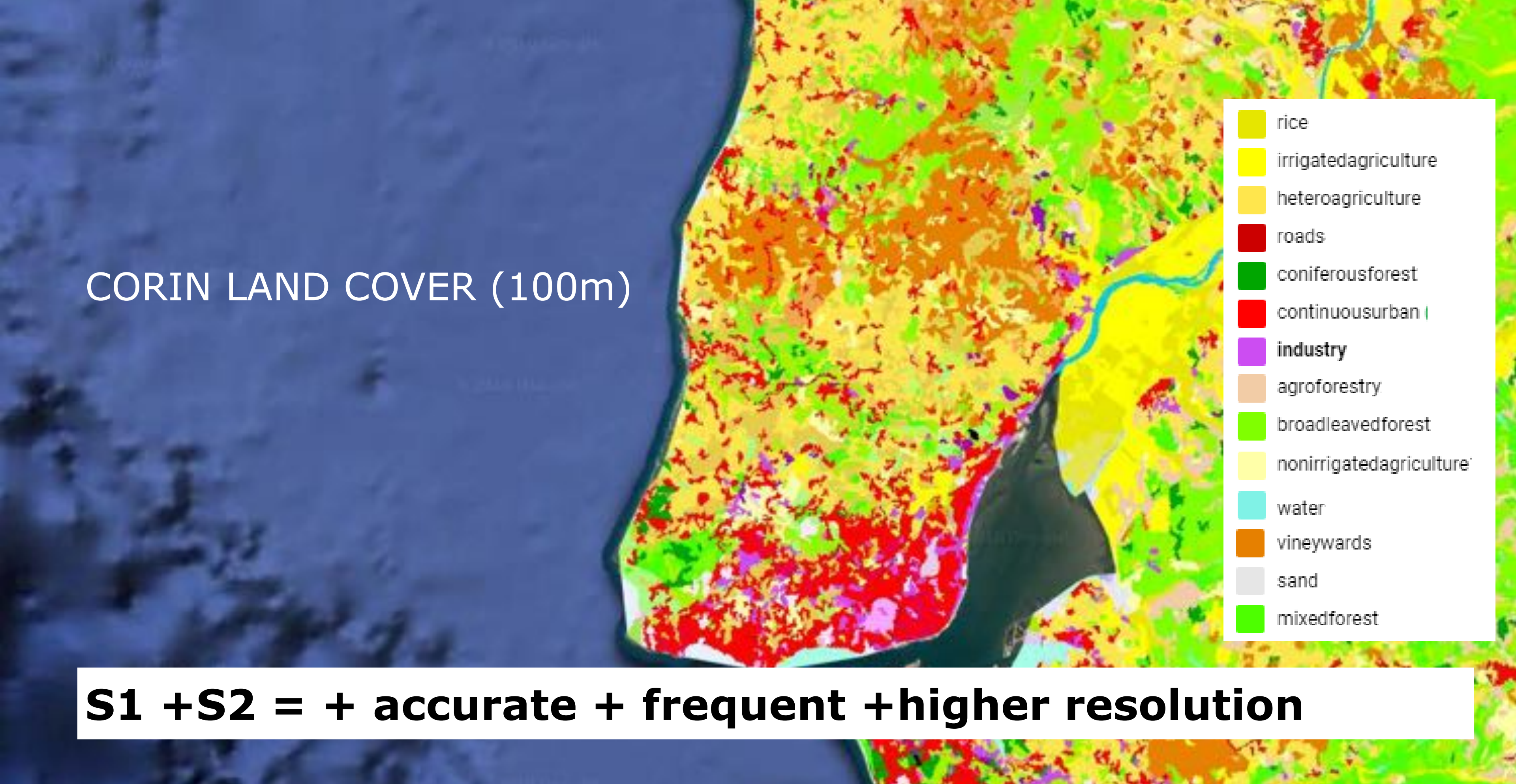
Elevation
(DEM STRM)

+

+



Large scale + data input + classes



S1 +S2 = + accurate + frequent +higher resolution

Overall Conclusions

Google Earth Engine

- GEE offers **powerful** capabilities in **handling large volumes** of remote sensing imagery
- GEE contains state-of-art machine learning algorithms achieving high accuracies and excellent tool for rapidly prototype AI applications.
- A big limitation is the need of manually tune the machine learning algorithms

Added value of fusing Sentinel data

- The integration of texture and spectral information for pixel-based classification improves classification accuracy (S2 outperforms S1 alone but together detect finer structures).
- Data: the more, ~~the merrier!~~

Large scale mapping – further work?

- Results can be used to calculate/estimate use cover and land change dynamics
- Normalize data and test more combinations!

Thank you for your attention!