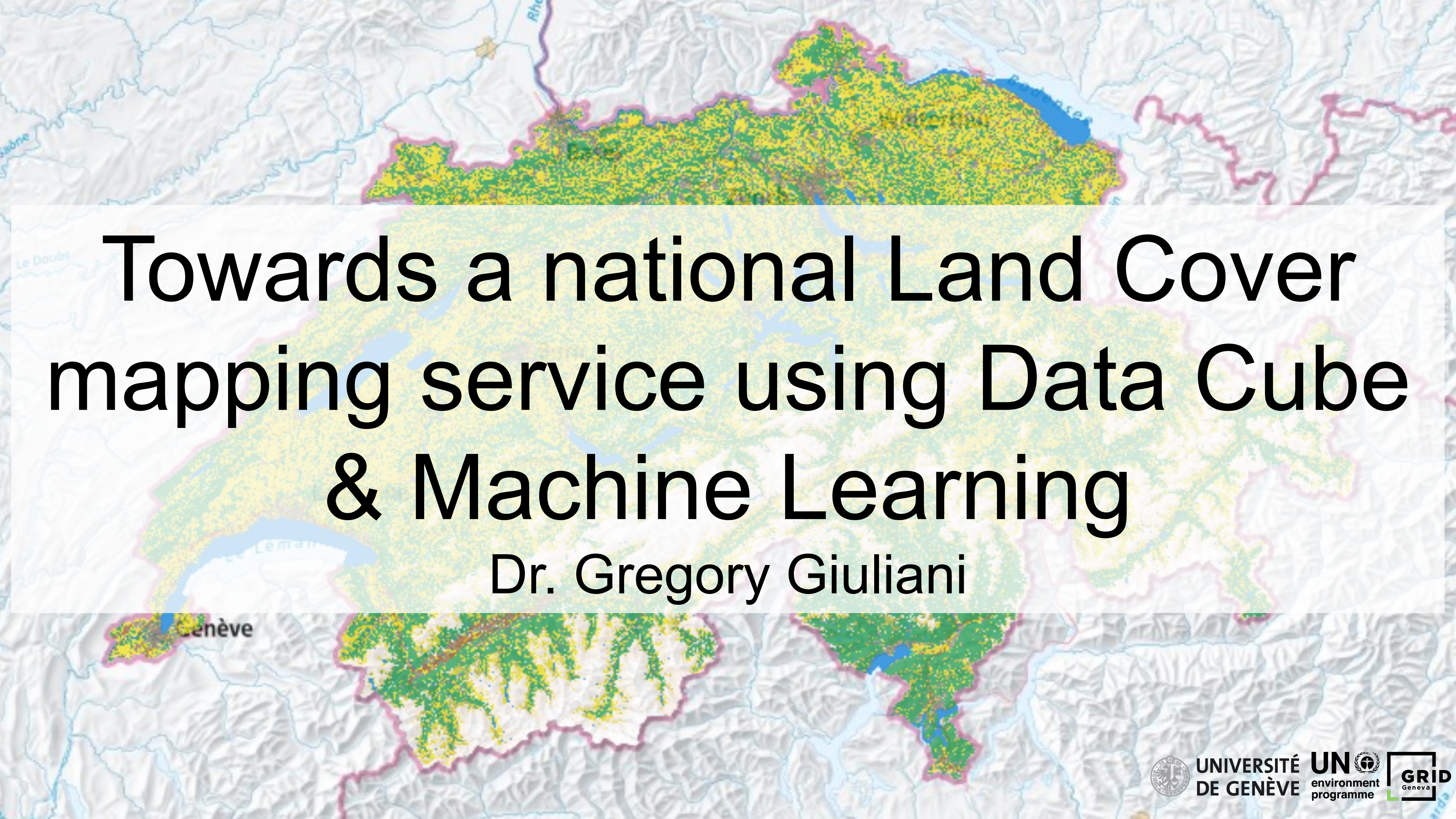


Towards a national Land Cover mapping service using Data Cube & Machine Learning

Dr. Gregory Giuliani

UNIVERSITÉ DE GENÈVE

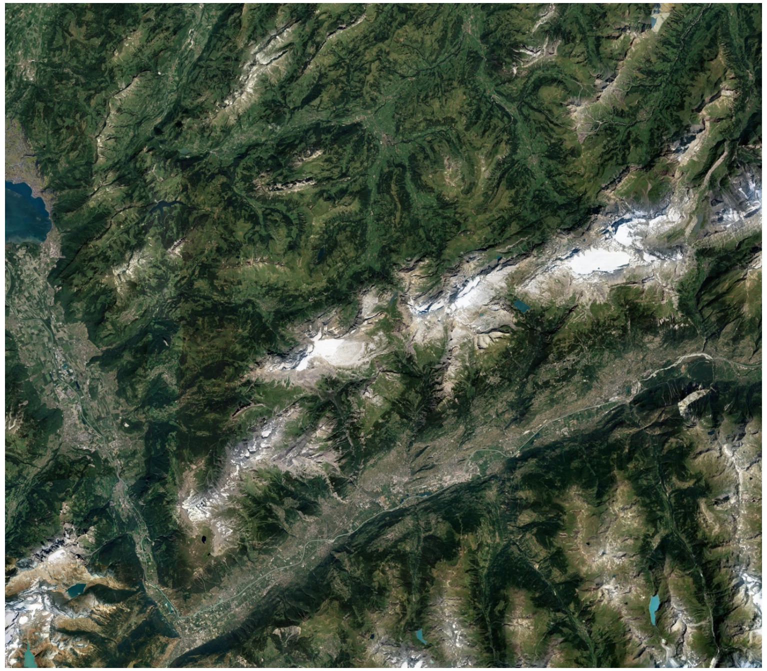
UN environment programme

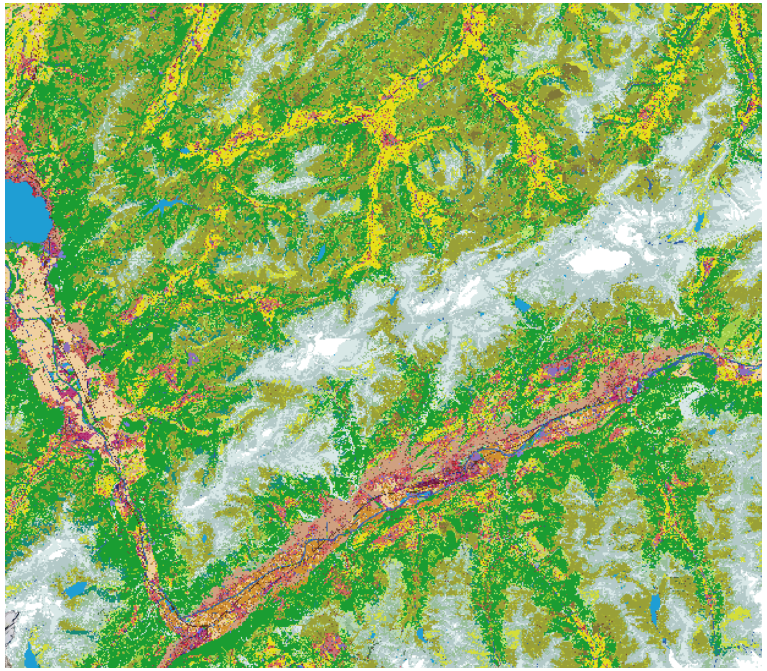
GRID Geneva

Land Cover in Switzerland

- In Switzerland, official LC data (*“Arealstatistik”*) are generated from visual interpretation of aerial photos. These maps are obtained by **visually interpreting and assigning a LC as well as a LU category of the lower-left corner of each sample point from a regular 100m grid cell** corresponding to more than **4 million points over the country**, following three nomenclatures: standard (72 categories); land cover (27), and land use (46) over **four-time periods (1979/85, 1992/97, 2004/09, 2013/18)**.

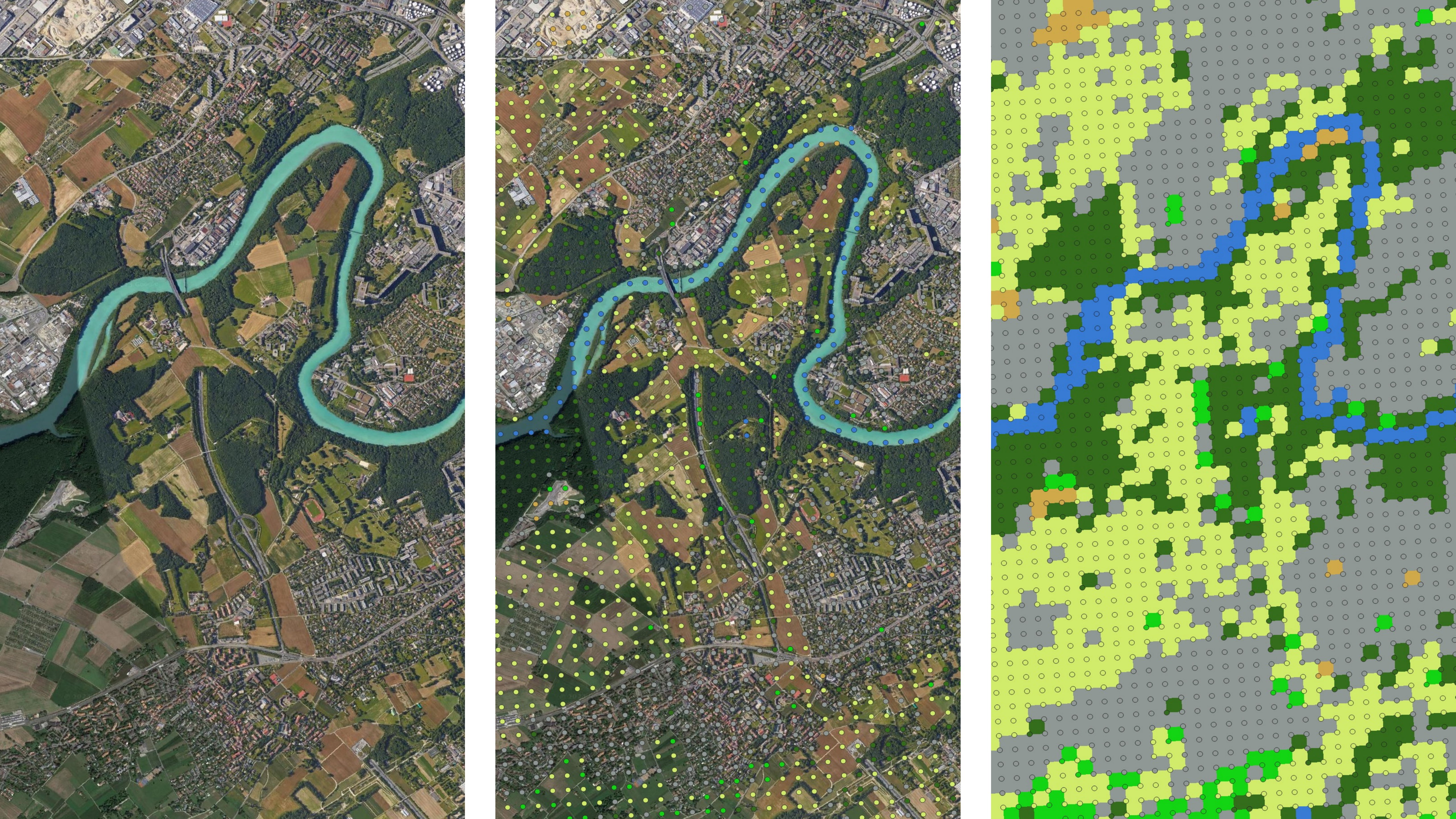
- This dataset is thematically more precise than commonly used classification. However, it suffers from a limited spatial (100m) and temporal resolution (6 years) **impending to correctly capture detailed landscape features, qualities, particularities, configurations as well as rapid changes**.





Arealstatistik Land Cover (NOLC04)

Principal domains	Basic categories
10 – Artificial areas	15 - Lawns 16 - Trees in artificial areas
20 – Grass and herb vegetation	21 – Gras and herb vegetation
30 – Brush vegetation	31 - Shrubs 32 – Brush meadows 33 – Short-stem fruit trees 34 - Vines 35 - Permanent garden plants and brush crops
40 – Tree vegetation	41 – Closed forest 42 – Forest edges 43 – Forest strips 44 – Open forest 45 – Brush forest 46 – Linear woods 47 – Cluster of trees
50 – Bare land	51 – Solid rock 52 – Granular soil 53 – Rocky areas
60 – Watery areas	61- Water 62 – Glacier, perpetual snow 63 – Wetlands 64 – Reedy marshes

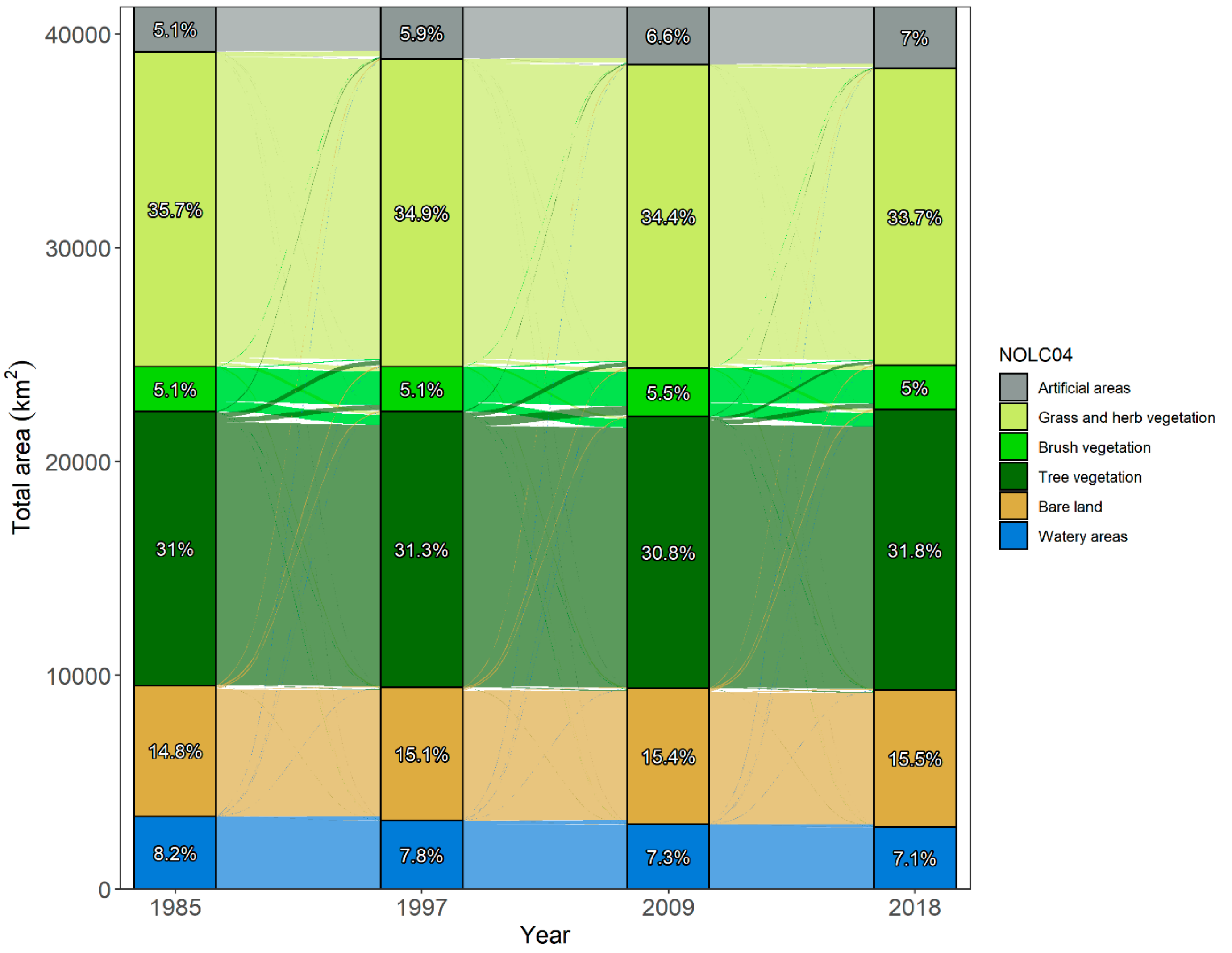


Land Cover Change in Switzerland

- Switzerland has undergone **small, spatially dispersed, dynamic, and gradual change trends**, with high rates of transition between low growing Brush Vegetation and forest LC classes in recent years.

- However, findings also suggest that **identifying drivers and understanding the rate of change are limited by the spatial resolution and temporal update frequency of the ArealStatistik**. The ability to understand these drivers **would benefit from a high-resolution annual LC dataset**.

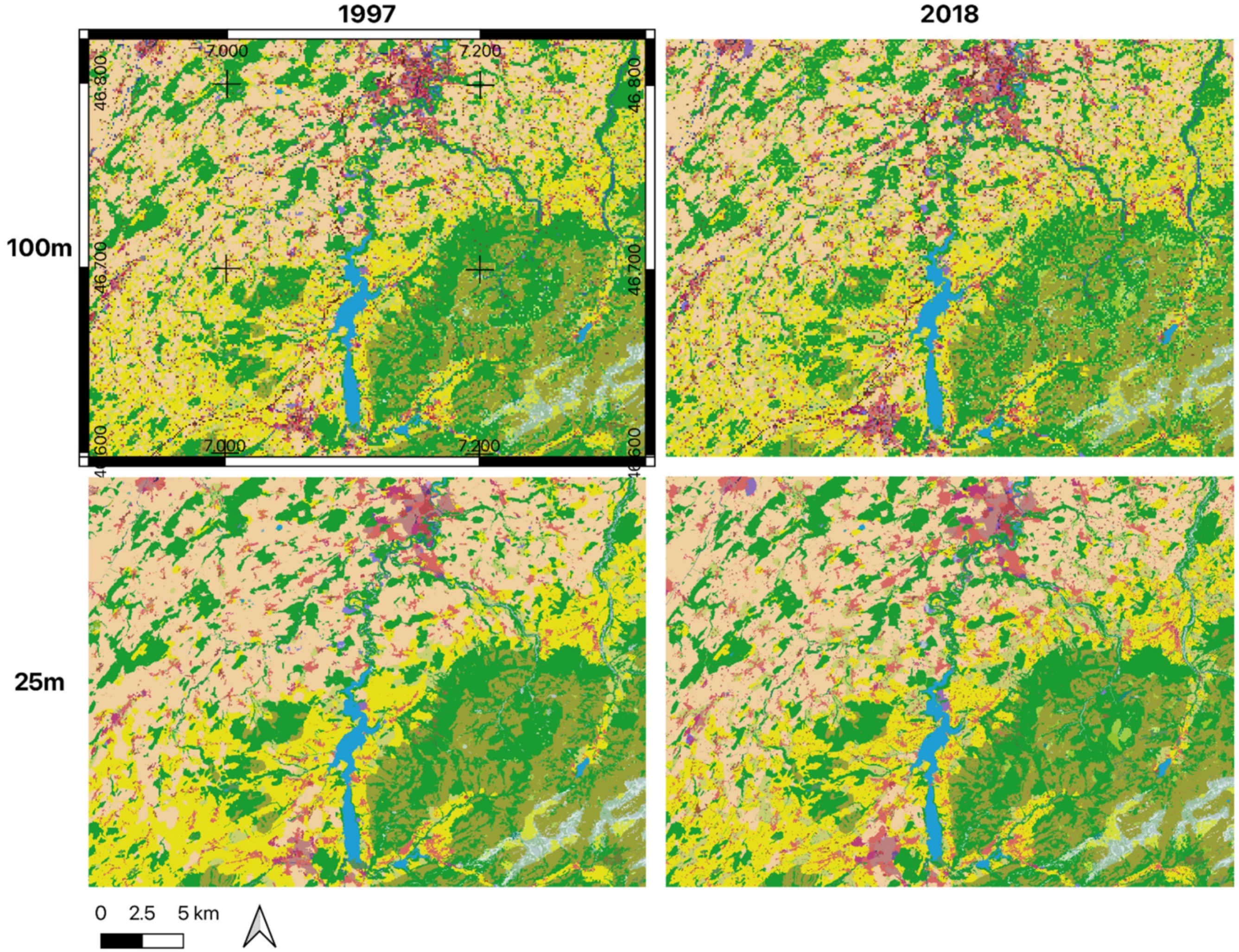
- Such a data product **can be produced using the *ArealStatistik* together with dense satellite data time-series and Machine/Deep Learning techniques**.



NICHOLSON THOMAS, Isabel Mary, GIULIANI, Gregory. Exploring Switzerland's Land Cover Change Dynamics Using a National Statistical Survey. In: Land, 2023, vol. 12, p. 1–20. doi: 10.3390/land12071386

Objectives

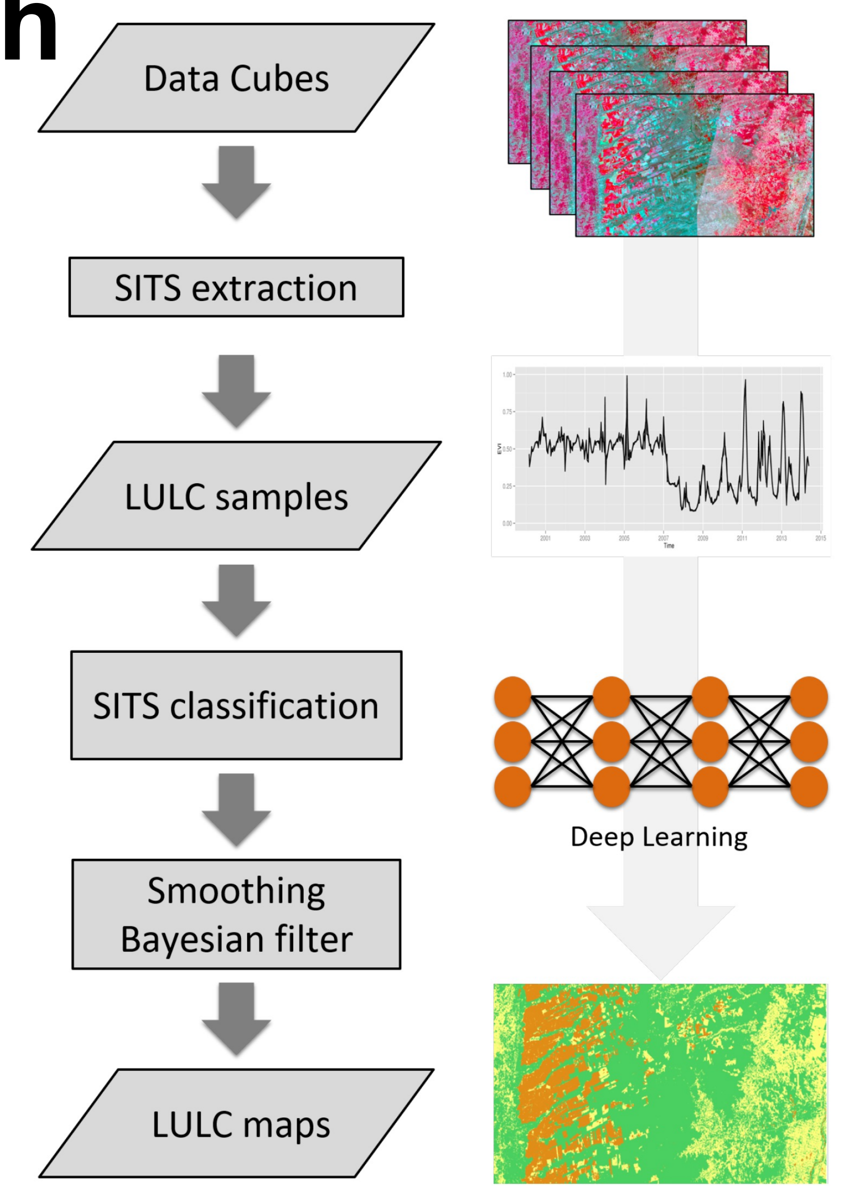
Combine **Data science techniques** (e.g., EO Data Cube, Machine Learning algorithms, and High-Performance Computing) to **develop new methodologies for the production of consistent and reliable yearly, medium-to-high resolution (spatial, temporal, thematic) time-series of LC data** across Switzerland to inform national environmental/territorial policies and planning.



GIULIANI, Gregory et al. Downscaling Switzerland Land Use/Land Cover Data Using Nearest Neighbors and an Expert System. In: Land, 2022, vol. 11, n° 5, p. 615. doi: 10.3390/land11050615

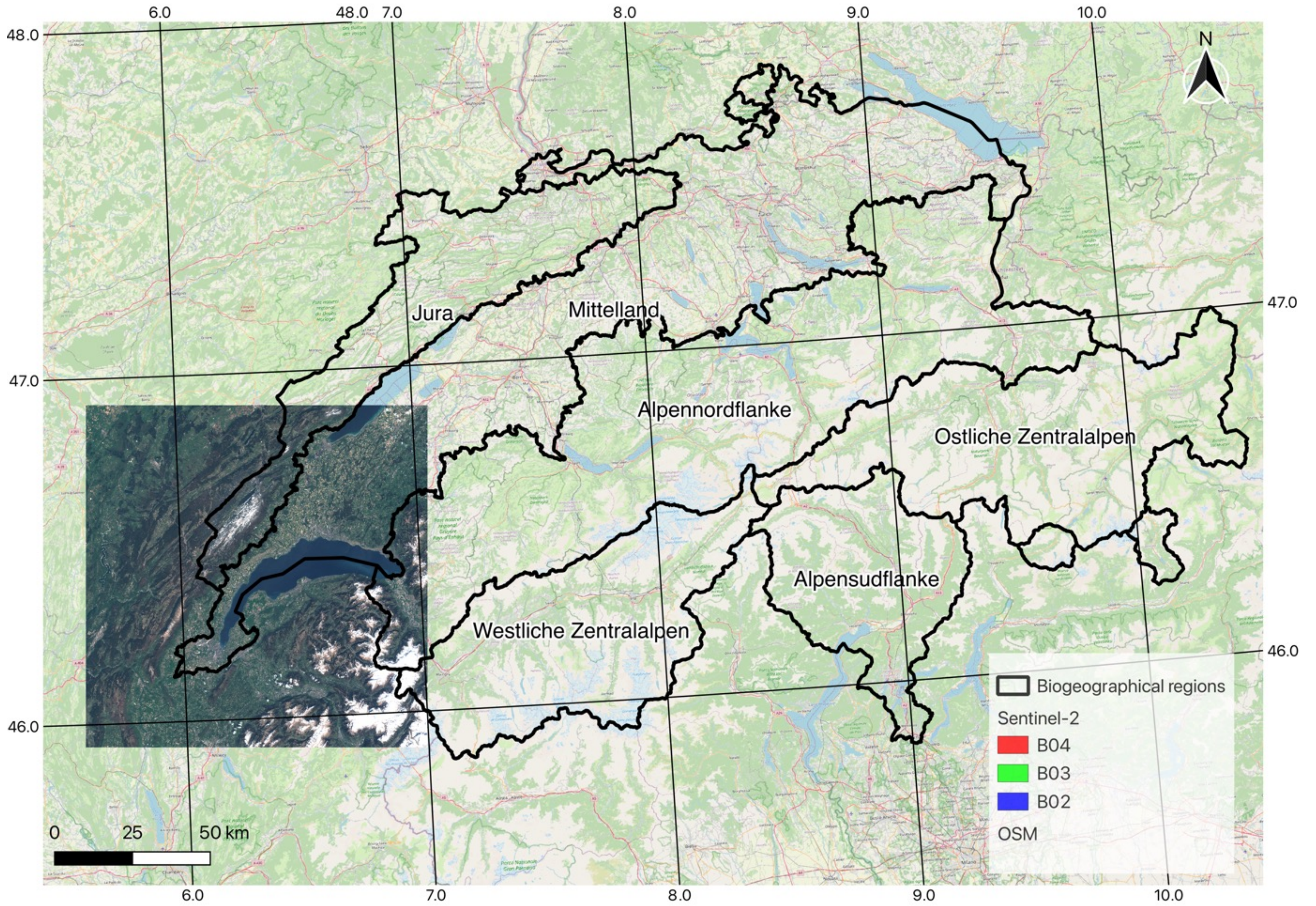
Space-first vs Time-first approach

- Current approaches for classifying images **do not consider intra-annual variability**.
- ***Space-first***: classify images separately. Compare results in time and derive a transition matrix. Uses temporal aggregation (e.g. annual) to reduce the volume of image collections and overcome data gaps (e.g. clouds)
- With dense time-series available in EODC > fully **benefit from the temporal resolution to capture changes**.
- ***Time-first***: classify time series separately. Join results to get maps
- Hypothesis: LC classes of interest are distinguishable partly because of their temporal characteristics
- All values of the time series are inputs for classification methods to label individual pixels.
- Each spatial location is associated with a time series.
- Better suited to track changes continuously.



Camara, Gilberto, et al. "Big earth observation data analytics: Matching requirements to system architectures." Proceedings of the 5th ACM SIGSPATIAL international workshop on analytics for big geospatial data. 2016.

Study area



6.0
48.0
7.0
8.0
9.0
10.0
48.0
47.0
46.0
47.0
46.0
N
Jura
Mittelland
Alpennordflanke
Ostliche Zentralalpen
Alpensudflanke
Westliche Zentralalpen
Biogeographical regions
Sentinel-2
B04
B03
B02
OSM
0
25
50 km
6.0
7.0
8.0
9.0
10.0

SWISS DATA CUBE *in Numbers*

A unique Analysis Ready Data Archive

39 years
FROM 1984 to 2023

10 sensors
LANDSAT 5/7/8/9;
SENTINEL-1AB/2AB/3/5P

Official gov. data
DEM; Climate models; Land Cover,...

EO data products
NDVI, NDWI, EVI, LAI, … time-series

> 450 million
PIXELS

> 3000 billion
OBSERVATIONS

10-30-90m
PIXEL RESOLUTION

~ **80'000** images
INGESTED

~**30** TB
ANALYSIS READY DATA

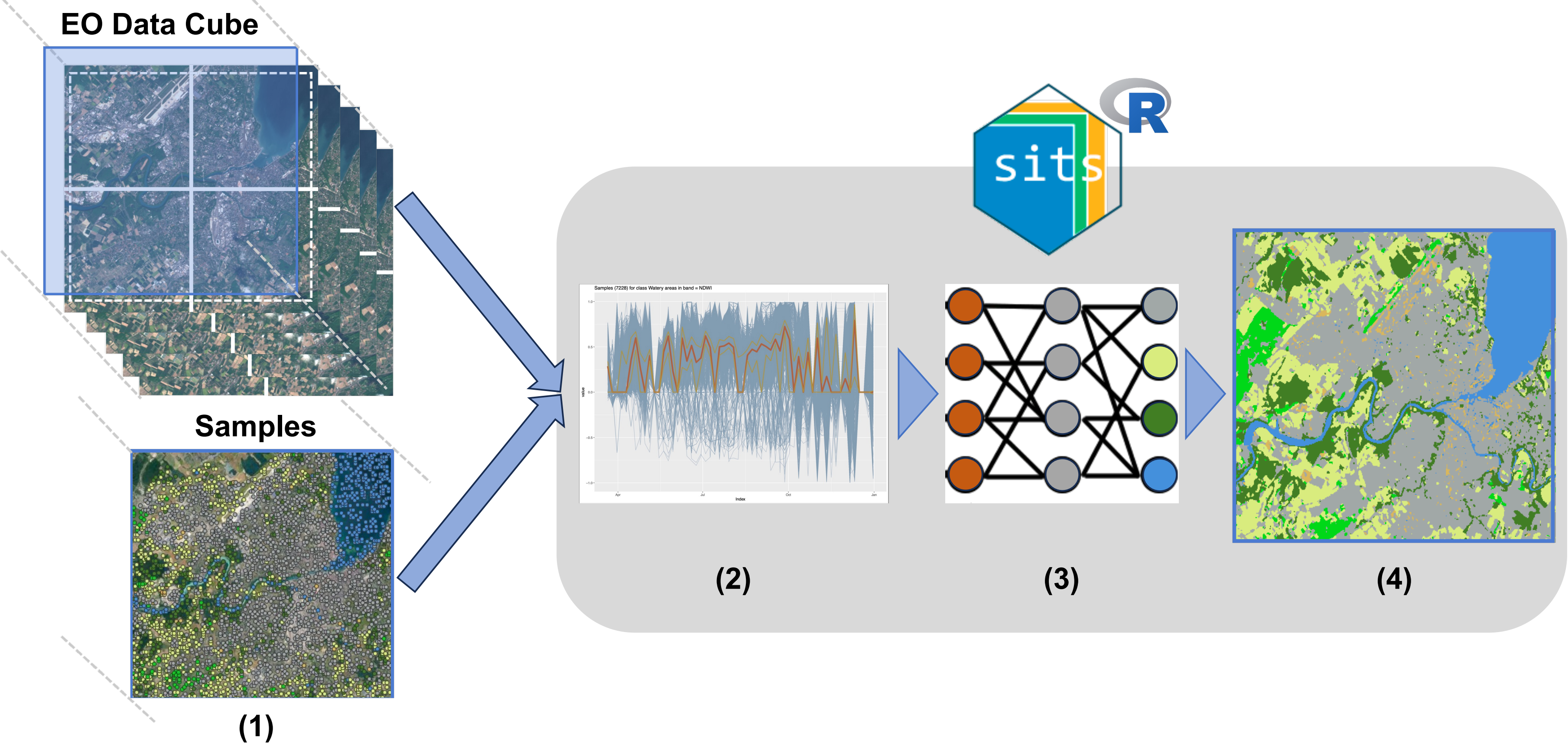
~**40** millions CHF
COST OF DATA WITHOUT OPEN DATA ACCESS POLICY

Chatenoux B., Richard J.-P. Small D., Roeoesli C., Wingate V., Poussin C., Rodila D., Peduzzi P., Steinmeier C., Ginzler C., Psomas A., Schaepman M., Giuliani G. (2021) The Swiss Data Cube: Analysis Ready Data archive using Earth Observations of Switzerland, ***Nature Scientific Data***. *8:295* https://doi.org/10.1038/s41597-021-01076-6

GRID Genève UNIVERSITÉ DE GENÈVE University of Zurich[UZH] WSL

SITS - Satellite Image Time Series Analysis for EODC

The R package **Satellite Image Time Series Analysis for Earth Observation Data Cubes** (SITS - https://github.com/e-sensing/sits) provides the necessary capabilities to **work with big satellite image data** sets and to fully **support all steps of land use and land cover classification workflow**: sampling selection, time series clustering, machine learning model training and validation, classification, and maps post-processing.



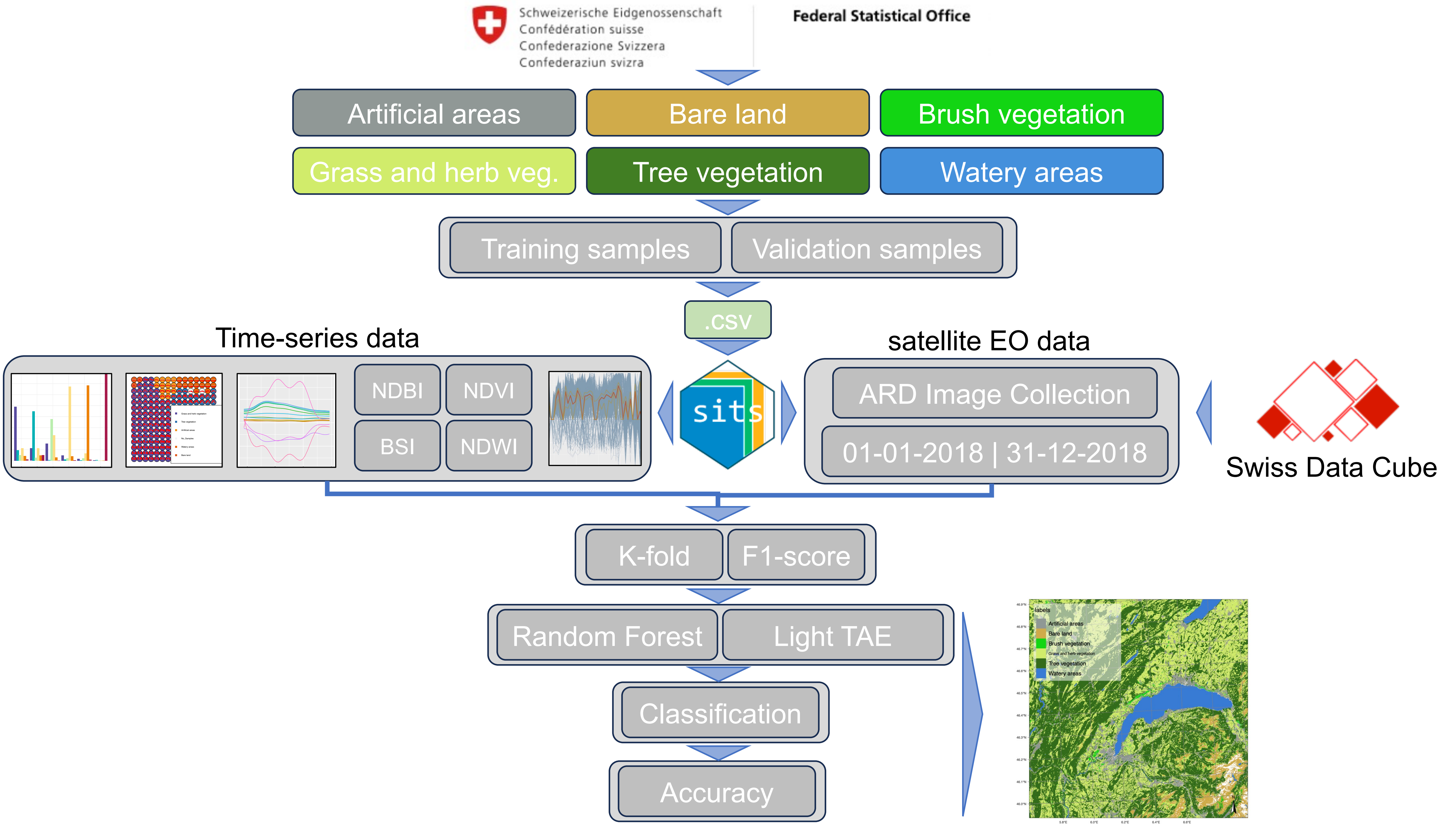
(1) Data preparation
ARD collection, regular cube, indices, samples

(2) Time-series extraction
Quality control & filtering

(3) Train ML/DL model
Based on time-series

(4) Land Cover map
Classification; Accuracy

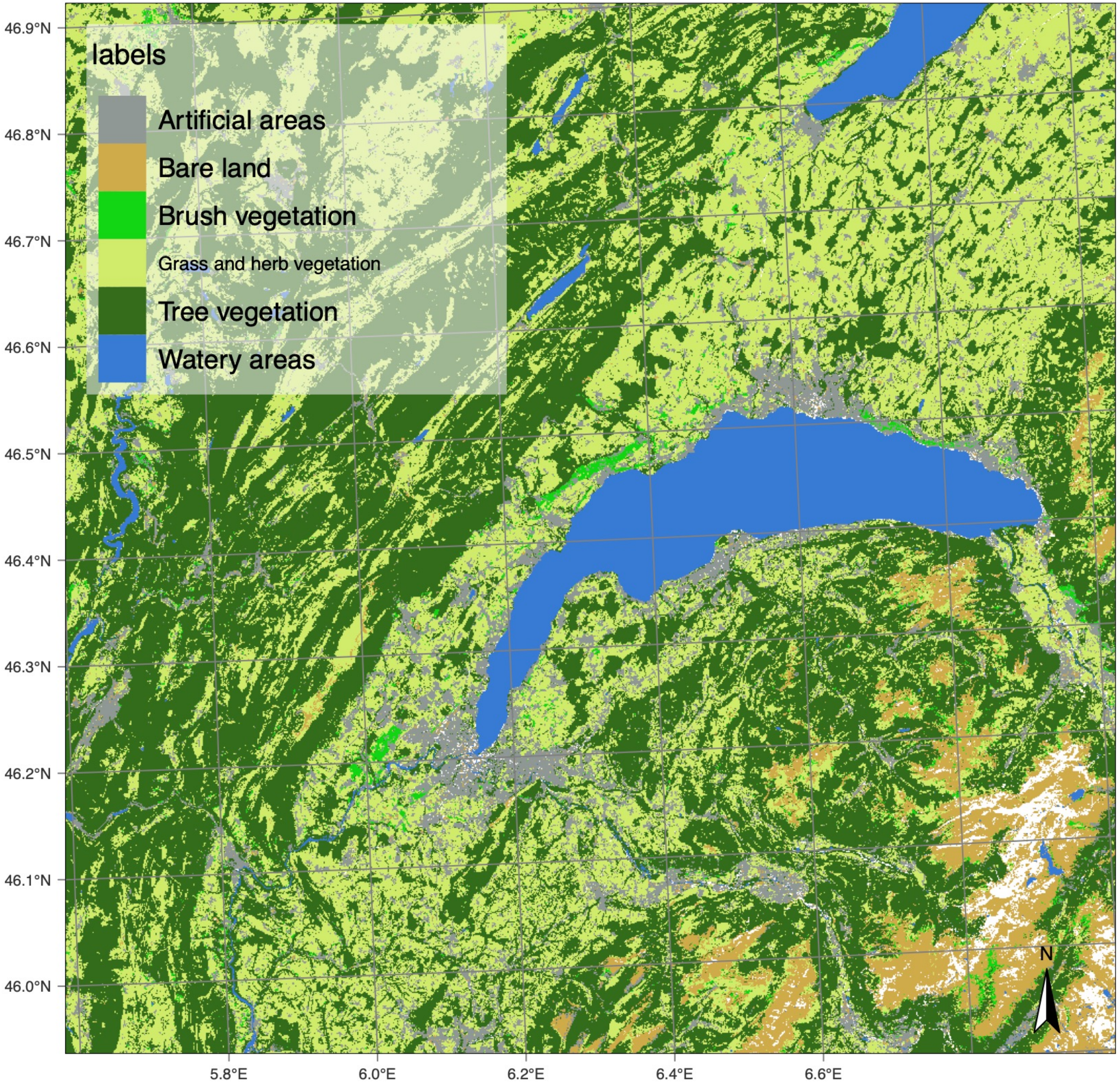
Simoes, R.; Camara, G.; Queiroz, G.; Souza, F.; Andrade, P.R.; Santos, L.; Carvalho, A.; Ferreira, K. Satellite Image Time Series Analysis for Big Earth Observation Data. Remote Sens. 2021, 13, 2428.



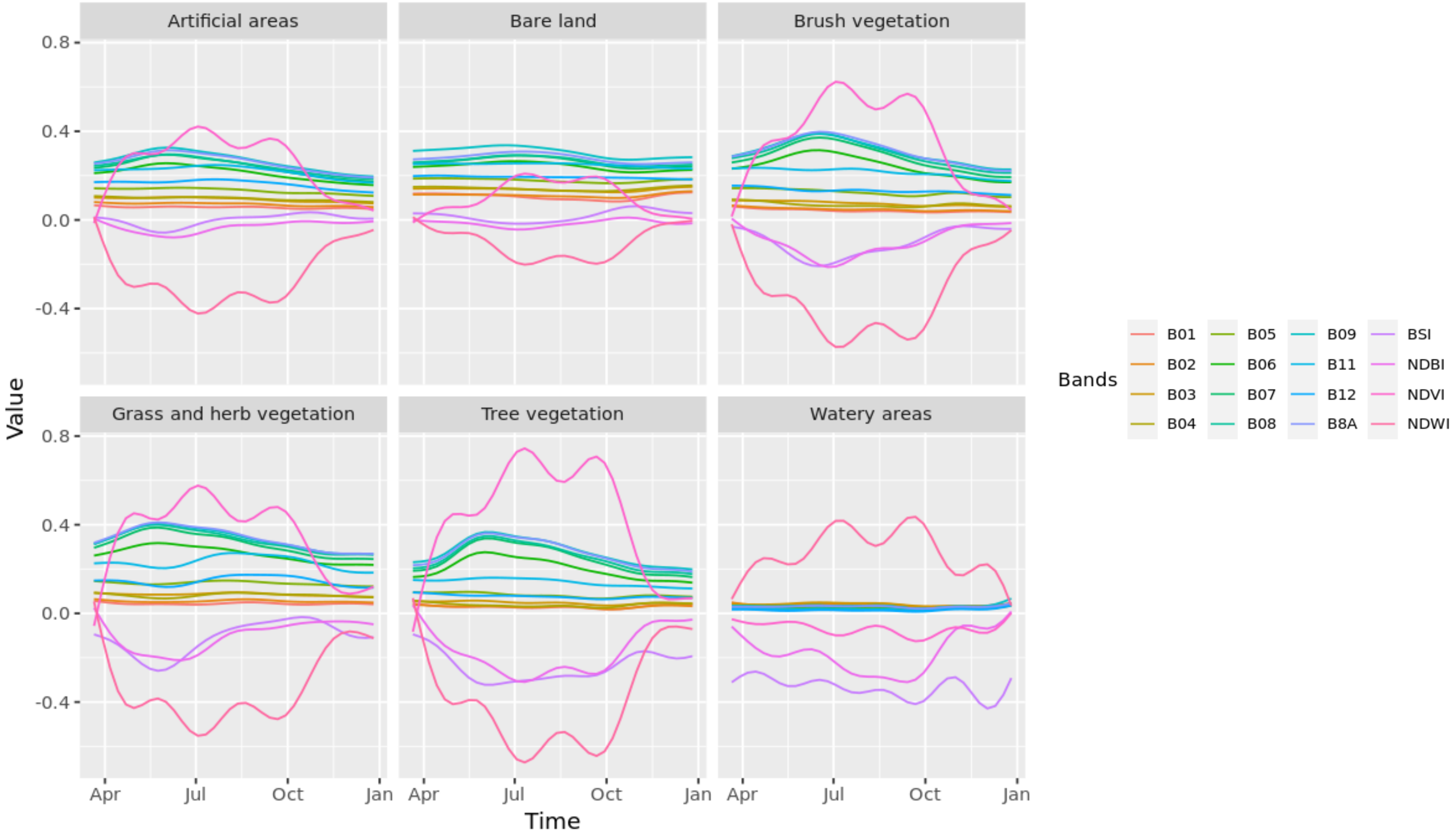
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Federal Statistical Office
Artificial areas
Bare land
Brush vegetation
Grass and herb veg.
Tree vegetation
Watery areas
Training samples
Validation samples
.csv
Time-series data
NDBI
NDVI
BSI
NDWI
sits
satellite EO data
ARD Image Collection
01-01-2018 | 31-12-2018
Swiss Data Cube
K-fold
F1-score
Random Forest
Light TAE
Classification
Accuracy
labels
Artificial areas
Bare land
Brush vegetation
Grass and herb vegetation
Tree vegetation
Watery areas

Results

- The workflow has been tested over the Lake Léman region to classify one year (2018) of Sentinel-2 images (113 images).
- 410’000 samples from the *Arealstatistik* (287’000 for training (70%); 123’000 for validation (30%)).
- Random Forest (RF); Temporal Convolutional Neural Network (tempCNN); Lightweight Temporal Self-Attention Encoder (LTAE)



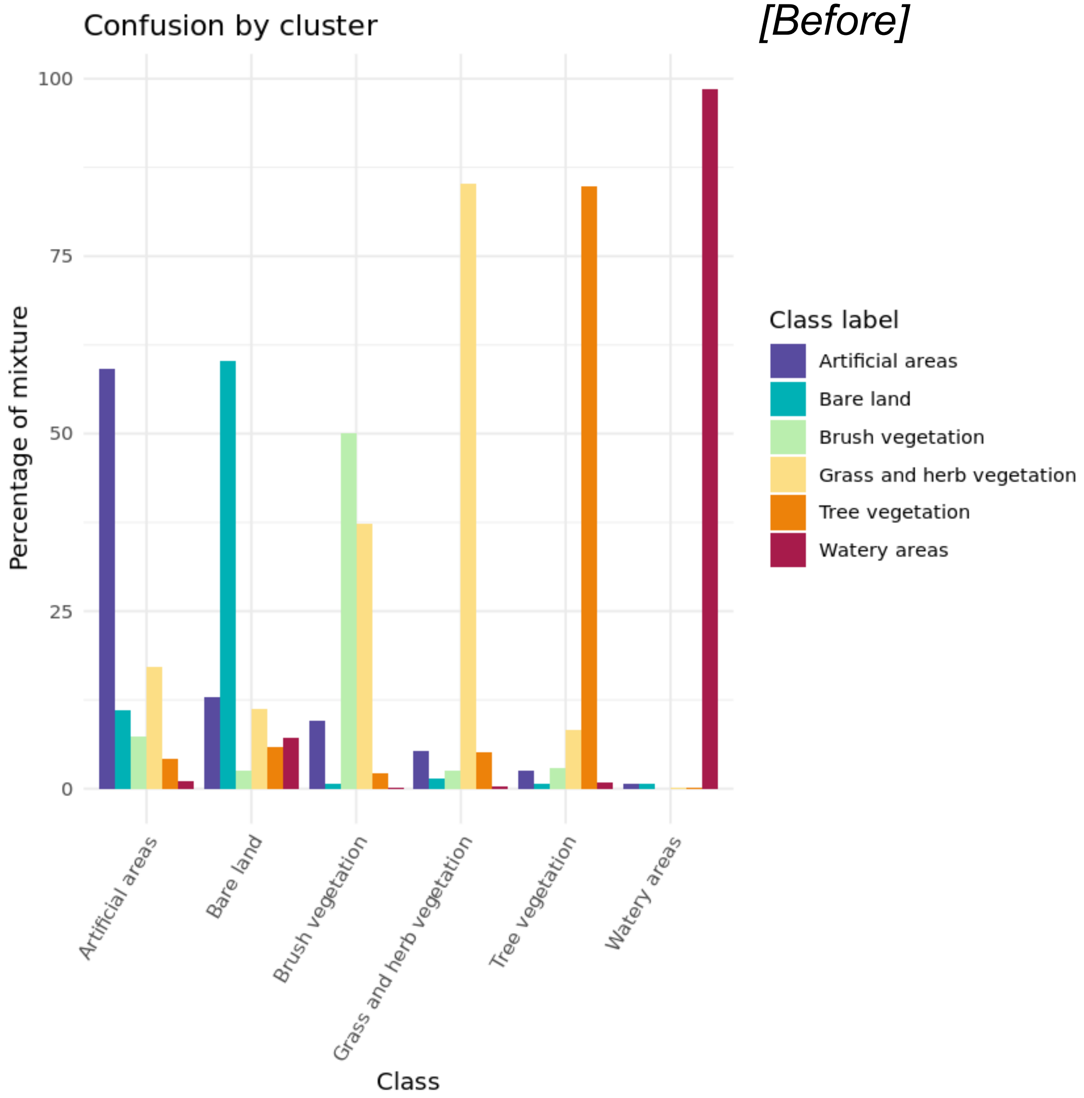
Time-series patterns



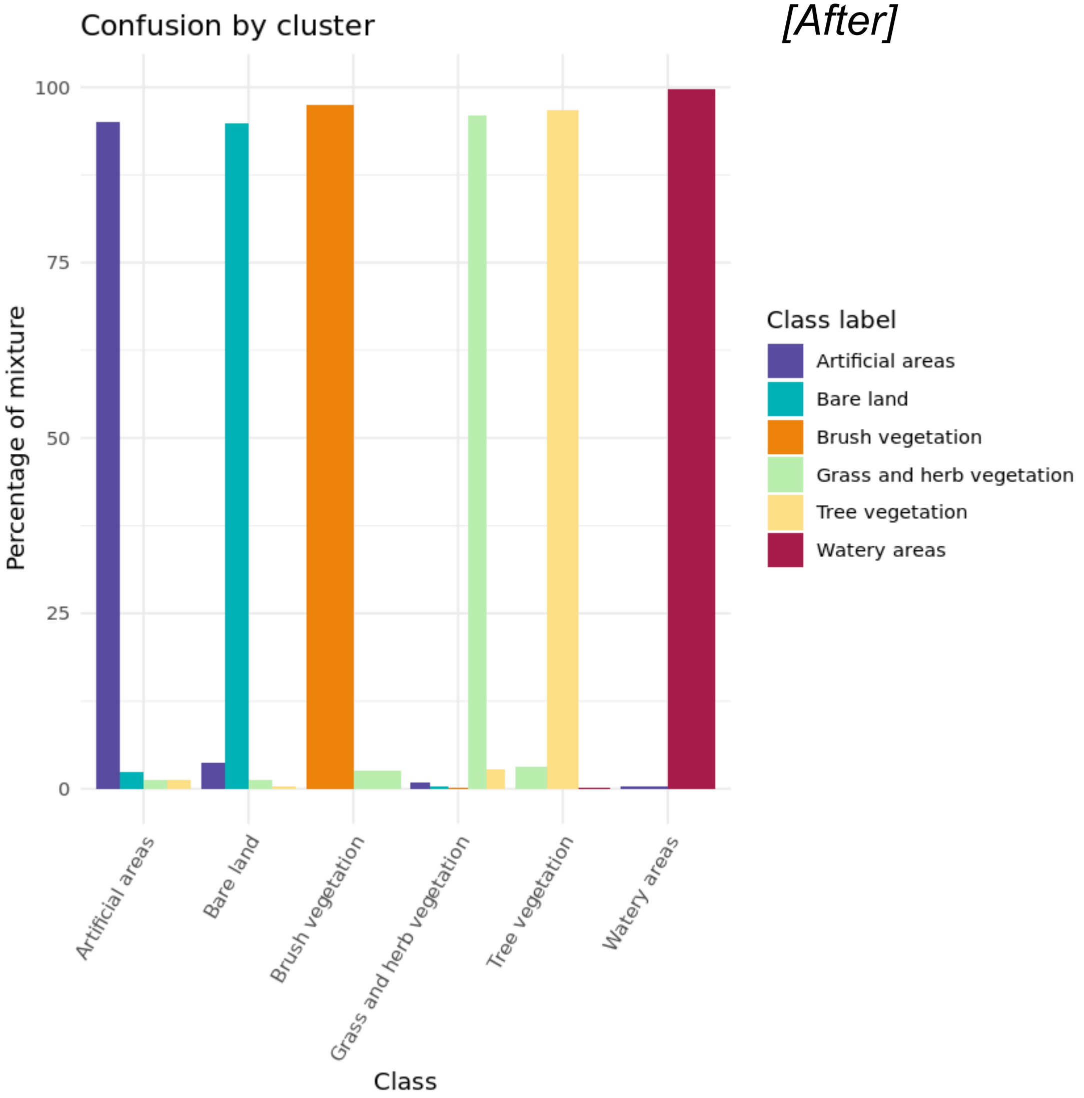
Improving the quality of training samples

(1) Quality assessment with Self Organizing Maps (SOM)
(2) Noisy samples detection
(3) Imbalance reduction

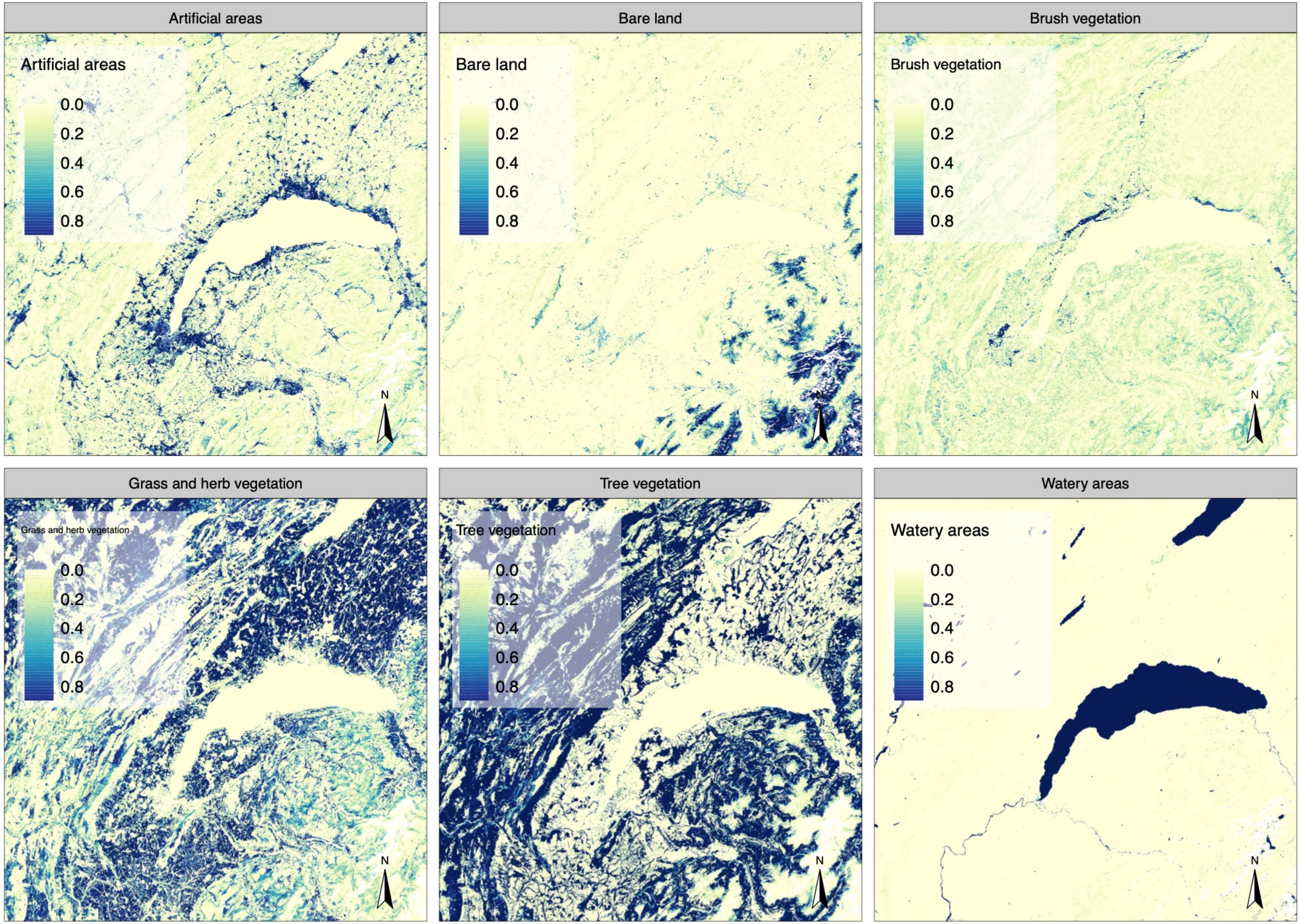
[Before]

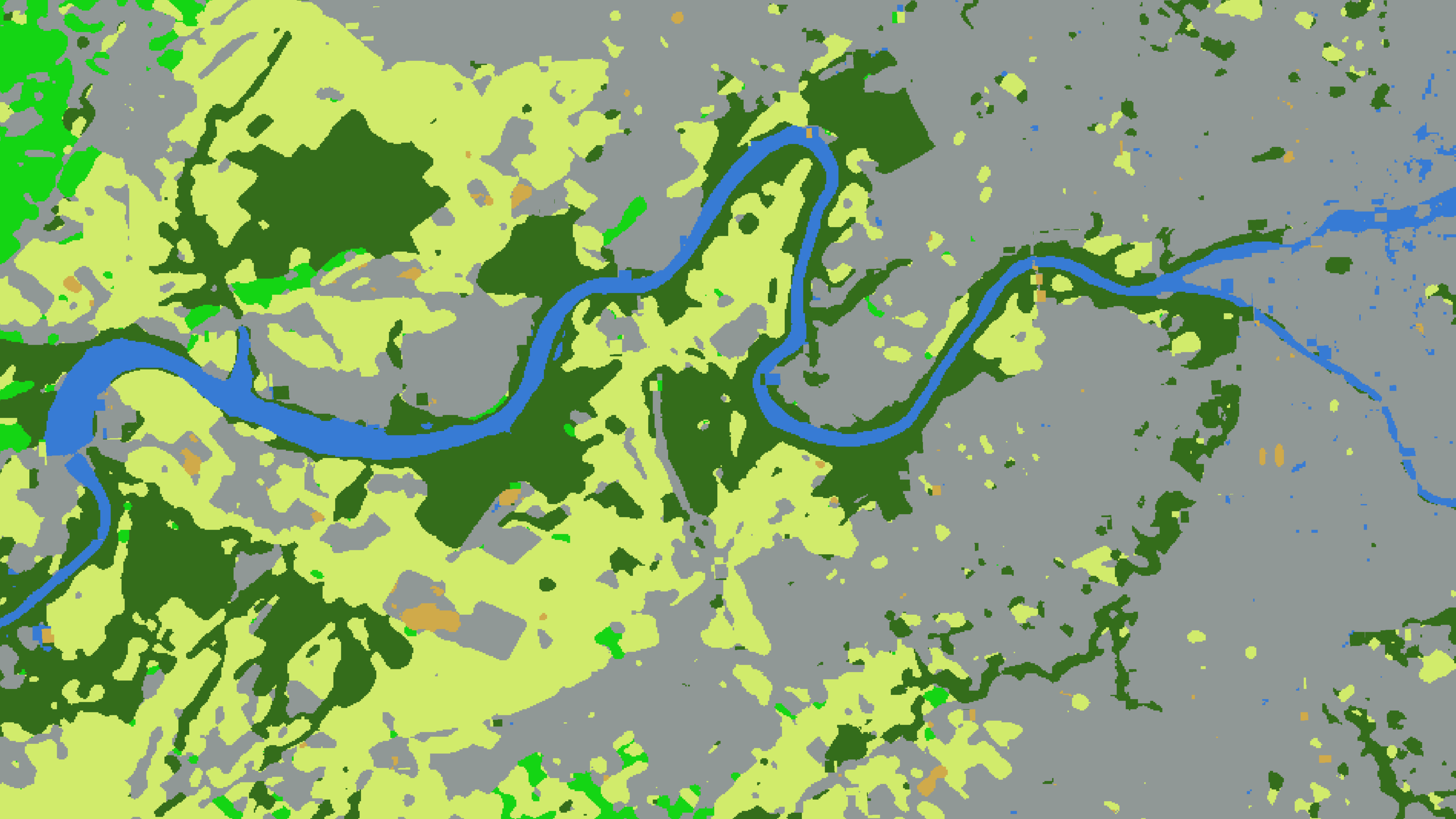


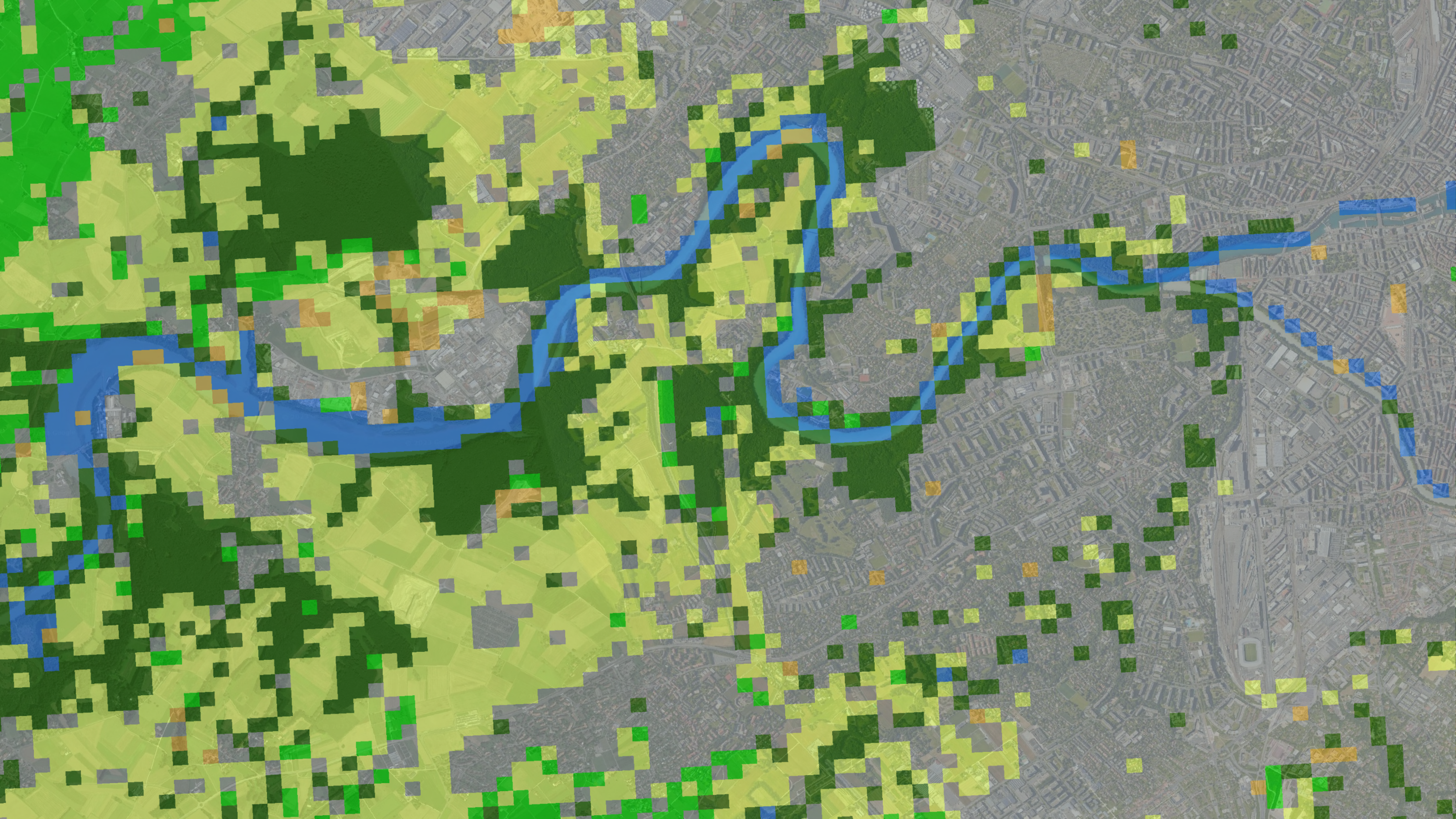
[After]



Probabilities







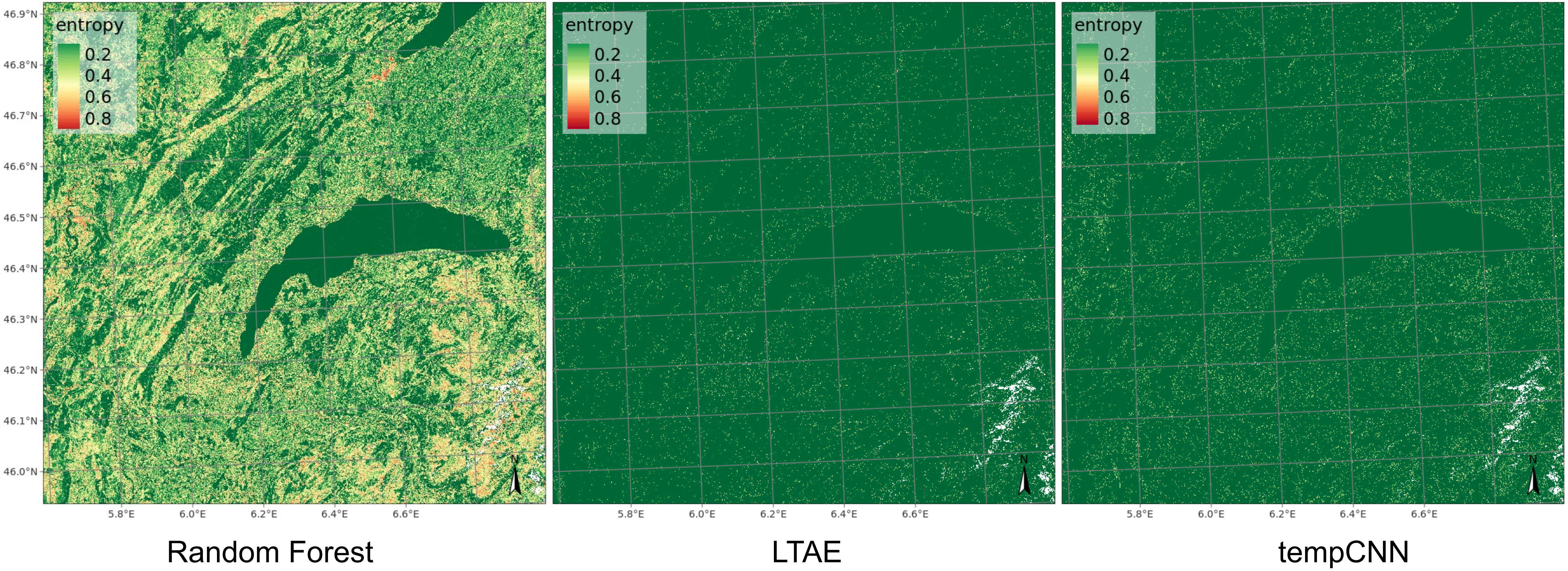
Cross-validation of training dataset

> estimate the inherent prediction error of a model.
> measure of model performance on the training data and not an estimate of overall map accuracy
> k-fold validation method

	RF	LTAE	tempCNN
Accuracy	0.9849	0.9703	0.9794
95% CI	(0.982, 0.9874)	(0.9664, 0.9739)	(0.9761, 0.9823)
Kappa	0.9817	0.964	0.975

Models have similar performance on training data

Uncertainty



DL models have less uncertainty

Accuracy

	RF		LTAE		LTAE (tuned)		tempCNN		tempCNN (tuned)	
Overall	0.85		0.83		0.86		0.86		0.90	
	UA	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**
Artificial areas	0.67	0.63	0.75	0.54	0.69	0.65	0.69	0.65	0.75	0.67
Bare land	0.65	0.38	0.57	0.40	0.64	0.41	0.64	0.41	0.72	0.49
Brush vegetation	0.68	0.11	0.38	0.23	0.68	0.15	0.68	0.15	0.71	0.19
Grass and herb vegetation	0.87	0.87	0.86	0.83	0.88	0.89	0.88	0.89	0.91	0.89
Tree vegetation	0.85	0.95	0.82	0.95	0.86	0.95	0.86	0.95	0.87	0.97
Watery areas	0.99	0.90	0.99	0.89	0.99	0.90	0.99	0.90	0.99	0.90

RF (space-first; median): 0.82

Initial conclusions

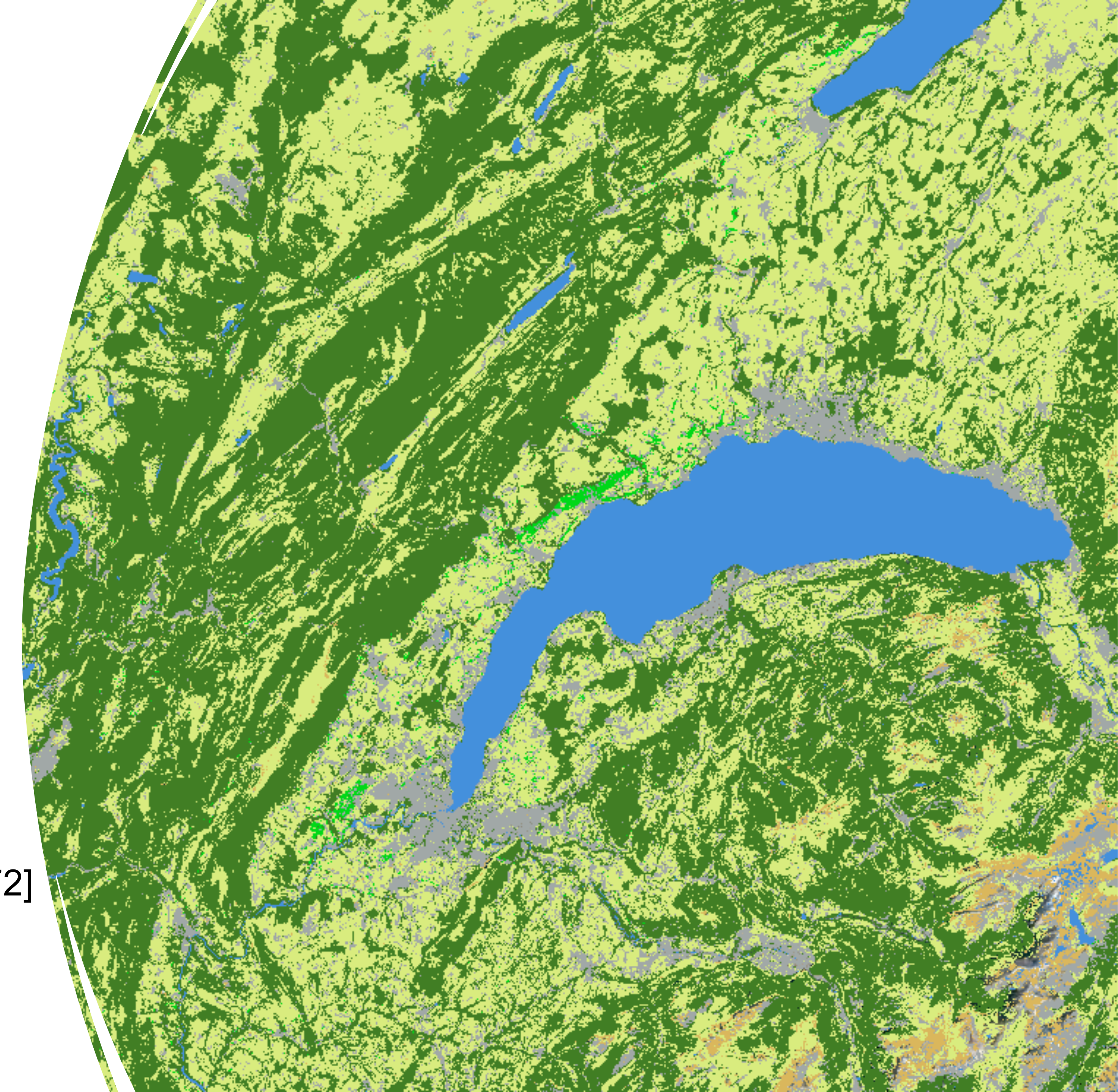
- time-first approach is performing better than space-first approach!
- tempCNN (tuned) appears to perform better than RF and LTAE
- DL methods (tempCNN; LTAE) have lower uncertainties than ML methods (RF)
- Good samples are essential!
- Hyperparameters tunning is essential too!

- Higher spatial resolution; detect more subtle details
- High accuracy (both overall and by class)
- Improved georeferencing

- Such approach is complementary to the official national statistics

Next steps

- Sampling strategy/representativity of samples
- Samples filtering and clustering;
- Add contextual data: DEM
- Add new indices: ARI (grassland/shrubland)
- Compute surface comparison between ArealStatistik & LC product
- 27 classes
- Land Use [NOLU04 – 4/10/46 categories]
- Standard Nomenclature [NOAS04 – 4/17/27/72]
- Landsat
- Produce the national map time-series



gregory.giuliani@unige.ch
gregory.giuliani@unepgrid.ch
http://www.unige.ch/envirospace/people/giuliani/