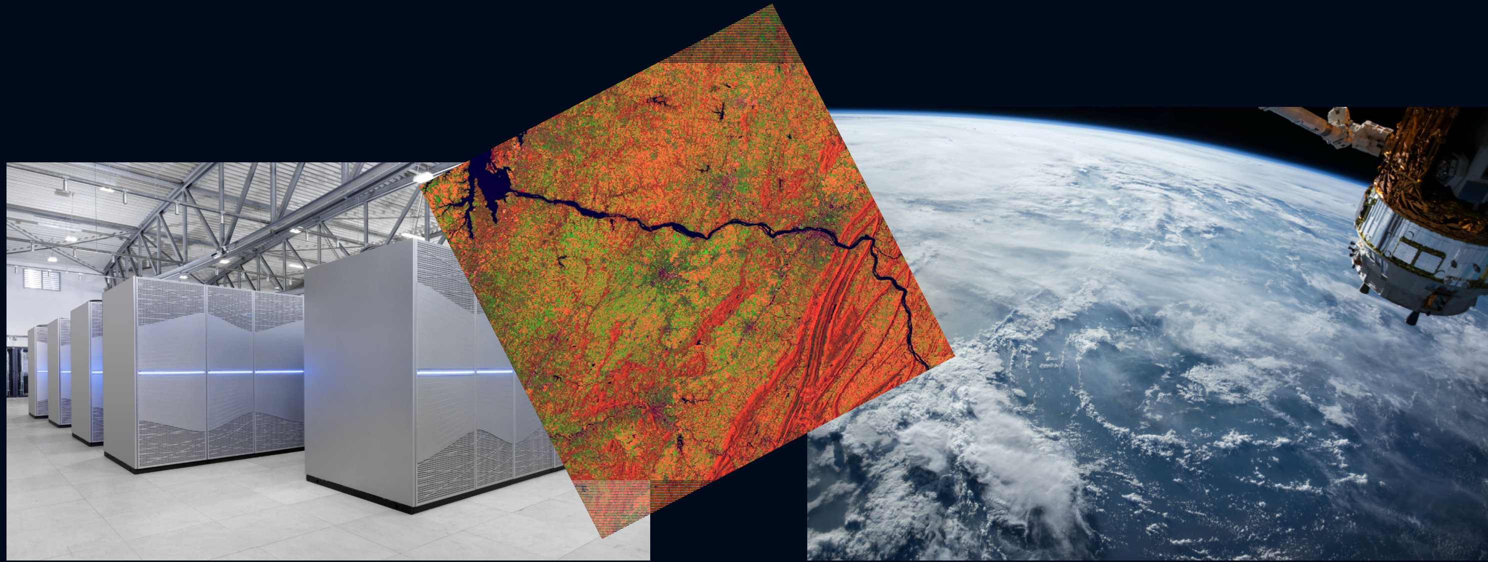


# AN AUTOMATIC APPROACH FOR THE PRODUCTION OF A TIME SERIES OF CONSISTENT LAND-COVER MAPS BASED ON LONG-SHORT TERM MEMORY

WE4.O2: Classification and Clustering I


ROCCO SEDONA[1,2], CLAUDIA PARIS[3], LIANG TIAN[2], MORRIS RIEDEL[1,2], GABRIELE CAVALLARO[1,2]

1. UNIVERSITY OF ICELAND, ICELAND
2. FORSCHUNGSZENTRUM JÜLICH, GERMANY
3. UNIVERSITY OF TWENTE, THE NETHERLANDS

# OUTLINE

- Introduction and motivation
- Framework for generation of multi-year training set
- A case study
- Scaling on modular supercomputing architectures
- Conclusions

# INTRODUCTION

- Problem:
  - Need for frequent update of land cover maps
  - Changes on the ground (i.e., due to new buildings or infrastructures, extreme meteorological events, deforestation, …)
  - Consistency of the output products
  - Large scale (national or even continental level)
  - t large scale

# MOTIVATION

- Aim: Update Land Cover maps more frequently

- How?
    - Using time series of S2 data
    - Extract reliable multi-year dataset for training
    - Use LSTM to exploit temporal information

C. Paris, L. Bruzzone, D. Fernández-Prieto, "A Novel Approach to the Unsupervised Update of Land-Cover Maps
by Classification of Time Series of Multispectral Images," IEEE Transactions on Geoscience and Remote Sensing, Vol. 57, No. 7, pp. 4259-4277, 2019

# MULTIYEAR TRAINING SET

Step 1: Data retrieval

- Input Sentinel-2 data downloaded with Sentinelsat API
  - L2A (atmospherically corrected)
  - Excluding clouds and partial orbits
  - Scaling data
- Extraction of corresponding Corine LC (CLC) of 2018

Map $\mathbf{M}^{y_1}$

$\mathbf{TS}^{y_1}$

https://sentinelsat.readthedocs.io/en/stable/

# MULTIYEAR TRAINING SET

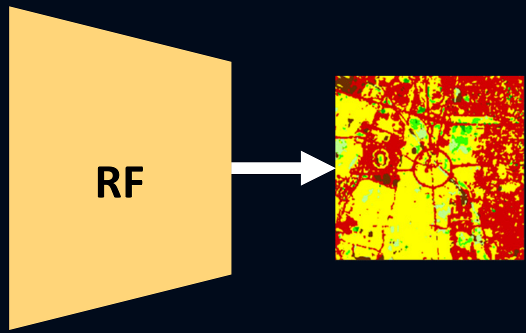
## Step 2: Clustering

- K-means clustering to retrieve most reliable samples
- Associate with labels from CLC
- Feature space made up of a set of spectral indices (i.e., vegetation, water snow and soil indices)

# MULTIYEAR TRAINING SET

Step 3: RF

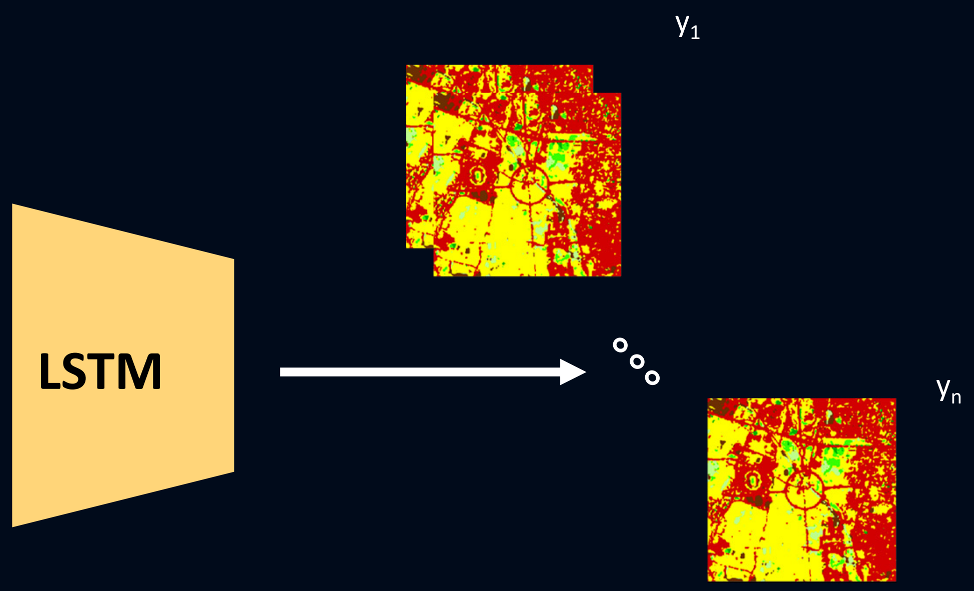
- RF trained on acquisitions of one year
- Produce maps with pseudo-labels for three years
- Extract a multi-year set of samples only where the labels agree
- Stratified random sampling



# MULTIYEAR TRAINING SET

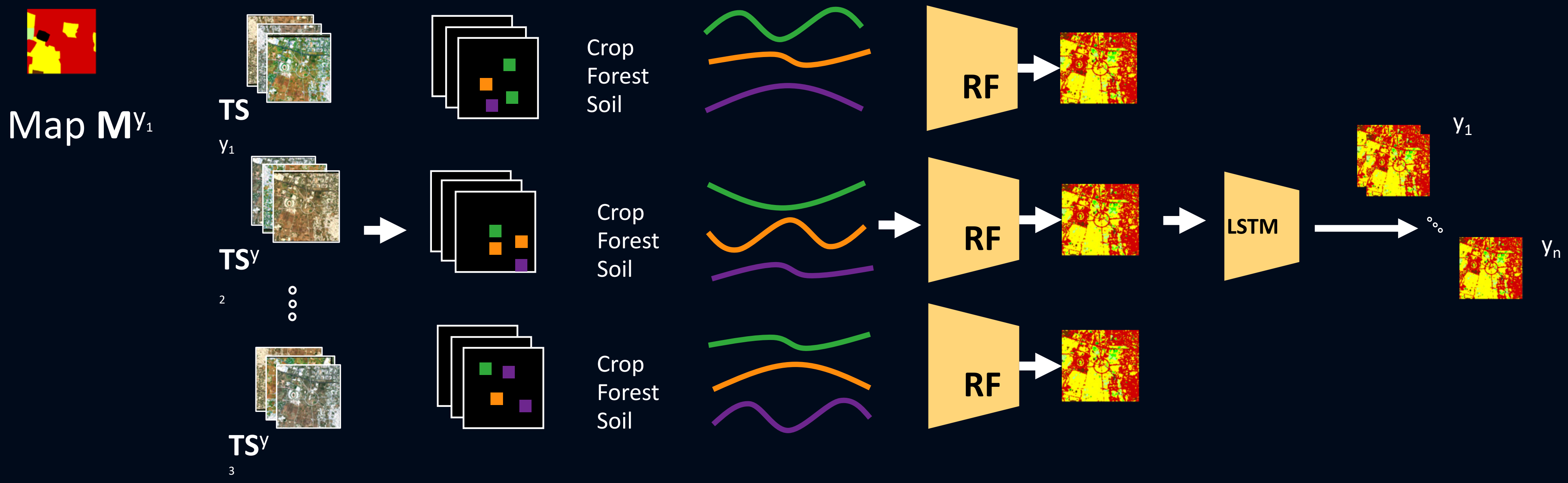
## Step 4: LSTM

- LSTM trained on multi-year TS



# MULTIYEAR TRAINING SET

## Creation of dataset encompassing multiple years




Step 1:
Data retrieval
and preprocessing

Step 2: Clustering

Step 3: RF for pseudo-labels

Step 4:Production of
multi-year LC map

# A CASE STUDY

## Area

- Sentinel-2 data of Trentino, Italy

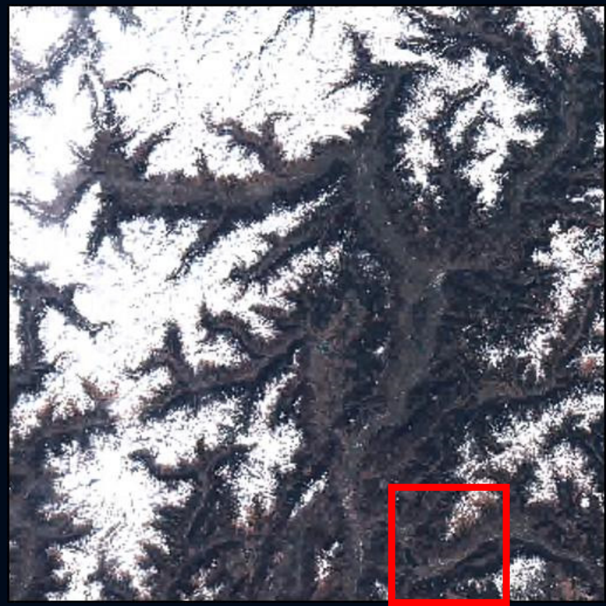
September 2018

# A CASE STUDY

## Why

- Area where manual assessment con be performed
- Vaja Storm in late October 2018

https://efi.int/news/aftermath-vaia-storm-italy-2018-12-17



March 2019

# A CASE STUDY
## Setup

- From 2018 to 2020
    - 20 acquisitions per year
    - 24.000 samples for training, 12.000 for validation
    - Prediction on disjoint set from training
    - 10 classes from CLC i.e., ”Artificial”, ”Grass”, ”Crops”,”Mineral”, ”Rocks”, ”Sand”, ”Broadleaves”, ”Conifers”,”Water”, ”Snow”

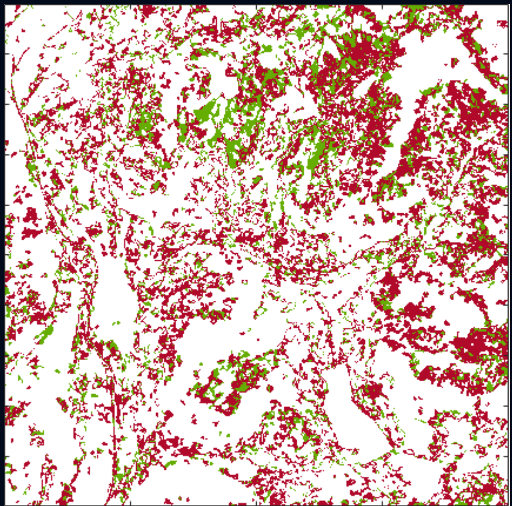
https://land.copernicus.eu/pan-european/corine-land-cover

# A CASE STUDY

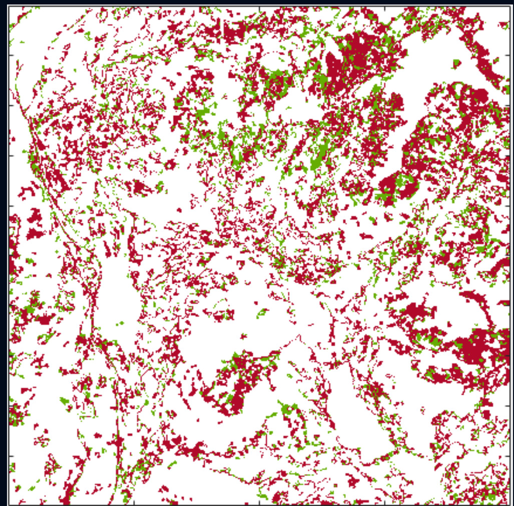
## Preliminary results

LC output map difference

2018-2020 single-year

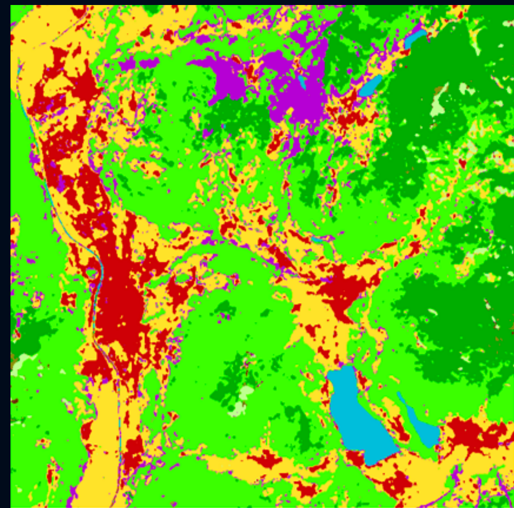


2018-2020 multi-year



True color S2 – 2019

Output LC map 2019



1 change

2 changes

Artificial

Sand

Grass

Broadleaves

Conifers

Crops

Mineral

Water

Rocks

Snow

# A CASE STUDY

Preliminary results

- Analysis of agreement of output LC maps produced by LSTM
- Increased consistency of the multi-year model compared to single-year model

| Class | Multi-Year LSTM | | Single-Year LSTM | |
|---|---|---|---|---|
| | 2018-2019 | 2019-2020 | 2018-2019 | 2019-2020 |
| Artificial | 0.89 | 0.68 | 0.93 | 0.62 |
| Grass | 0.69 | 0.60 | 0.61 | 0.42 |
| Crops | 0.76 | 0.81 | 0.65 | 0.85 |
| Mineral | 0.33 | 0.51 | 0.34 | 0.43 |
| Rocks | 0.43 | 0.46 | 0.45 | 0.43 |
| Sand | 0.14 | 0.31 | 0.05 | 0.16 |
| Broadleaves | 0.85 | 0.87 | 0.79 | 0.84 |
| Conifers | 0.81 | 0.78 | 0.79 | 0.77 |
| Shrubland | 0.64 | 0.66 | 0.63 | 0.62 |
| Water | 0.97 | 0.98 | 0.98 | 0.95 |
| Overall | 0.78 | 0.78 | 0.74 | 0.74 |

# SCALING ON HPC

- The framework can scale on modular supercomputing architectures
- Enhanced processing steps:
  - Pre-processing (i.e., data retrieval and training set extraction)
  - Distributed deep learning with multiple GPUs (i.e., training, inference)
- Enable to work at larger scale


https://www.fz-juelich.de/en/ias/jsc/systems/supercomputers/juwels

# CONCLUSIONS

- Scalable framework to generate multi-year training set
- Preliminary results with  longer TS
- Open questions:
  - How to include samples with changing pseudo-labels into the training set?
  - Increase consistency of the model while preserving ability to detect change
  - Validation (LUCAS, street level information?)
- Modular framework -> add new functionalities

Thank you for your attention

Terima kasih atas perhatian anda