RETURN – Monitoring Amazon Rainforest Recovery Capacity with terabytes of RADAR satellite images processed on EOSC infrastructures

Milutin Milenković

with contributions from:
Raymond Oonk, SURF (Samenwerkende Universitaire Reken Faciliteiten), The Netherlands
Wanda De Keersmaecker, VITO (Flemish Institute for Technological Research), Belgium
Johannes Reiche, Wageningen University, The Netherlands
Prof. Martin Herold, GFZ German Research Centre for Geosciences, Germany
Jan Verbesselt, The Belgian Science Policy Office (BELSPO), Belgium

Milutin Milenković (milenkovic@iiasa.ac.at)

Novel Data Ecosystems for Sustainability (NODES), IIASA

Geo-information Science and Remote Sensing Lab, Wageningen University

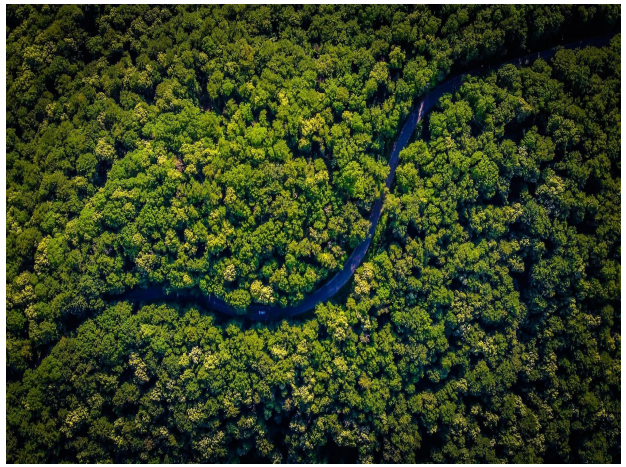
5. December 2022, EOSC Webinar

Pan-European digital assets supporting research communities – Benefits & opportunities

Session: Digital assets supporting SDG 13: Climate action

Resilience

Non-disturbed

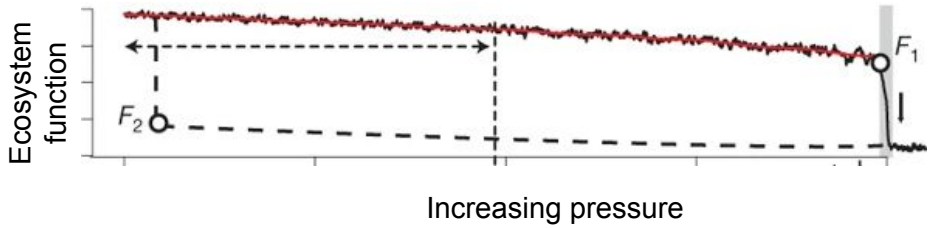


Credit: Unsplash/ Vlad Hilitanu

Disturbed

Credit: www.earthjournalism.net

M. Scheffer *et al.* (2009) *Nature*



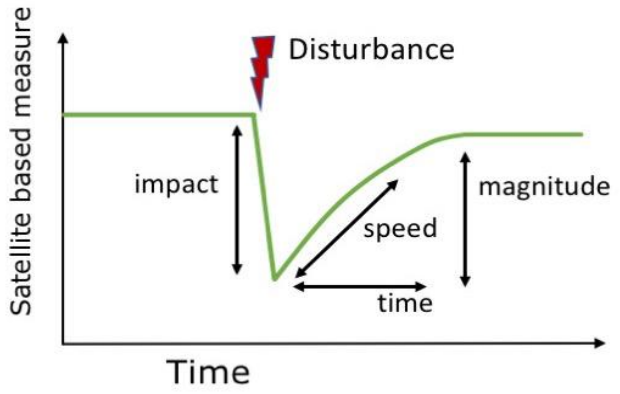
van Belzen, J. *et al.* (2017)

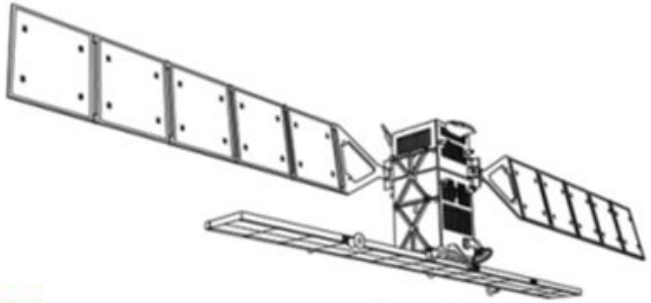
Resilience

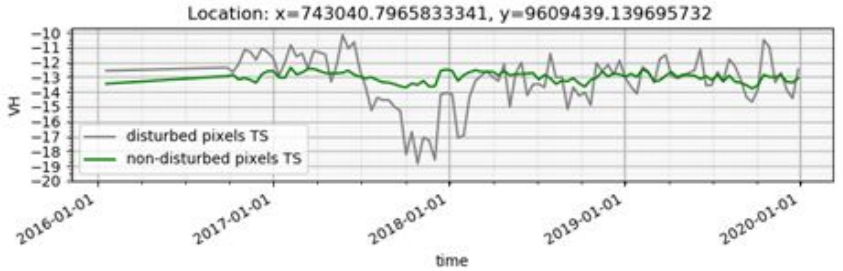
Credit: www.earthjournalism.net

Challenges & research questions:

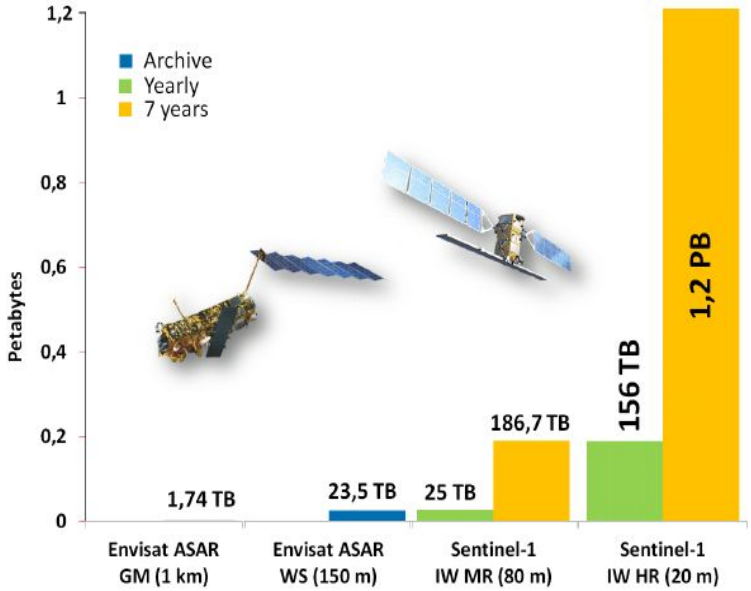
- What are the characteristics of signal recovery indices?
- What EO signal properties are related to forest recovery?
- **What Big EO data approaches are available for upscaling?**



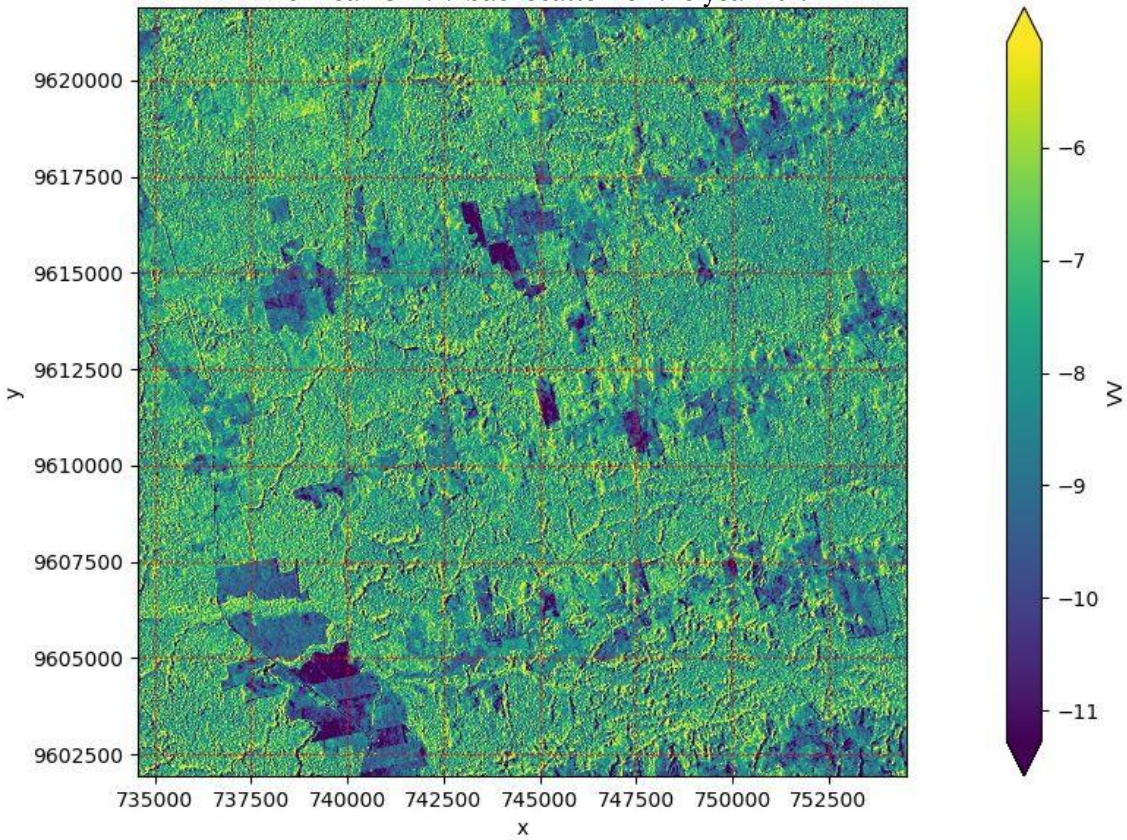




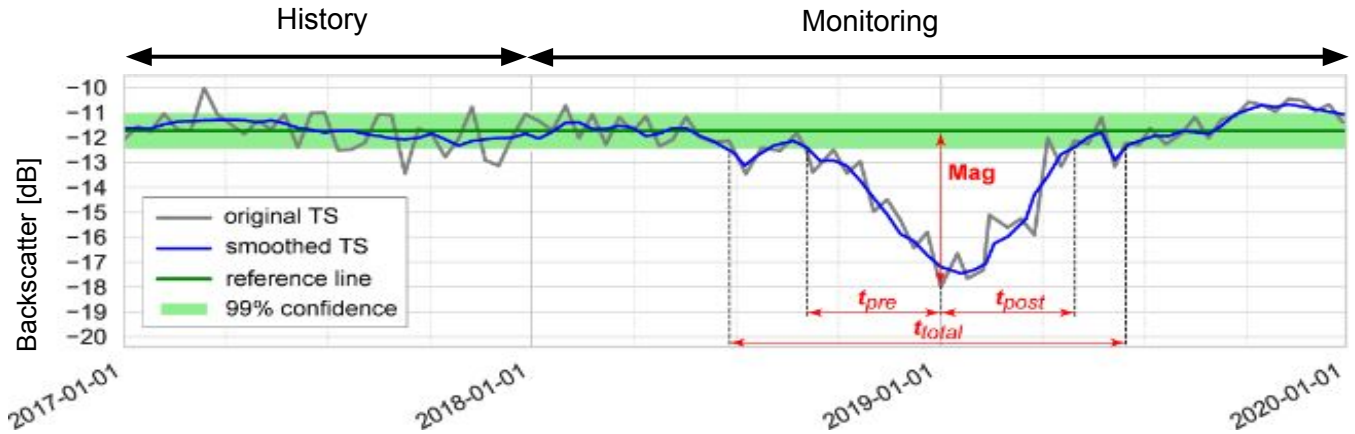
Wolfgang Wagner (2015)

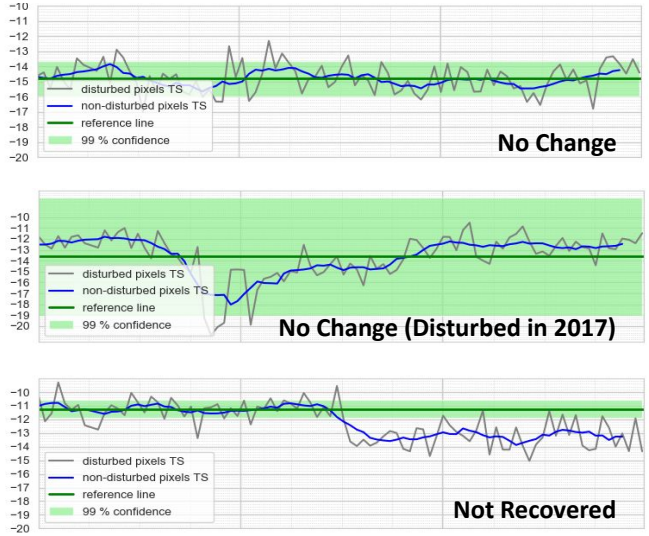


The mean S1-VV backscatter for the year 2019

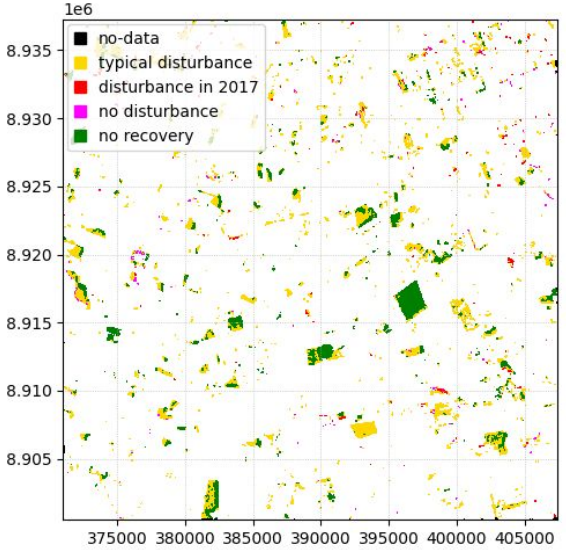


Analysis Framework for Sentinel-1 Time Series

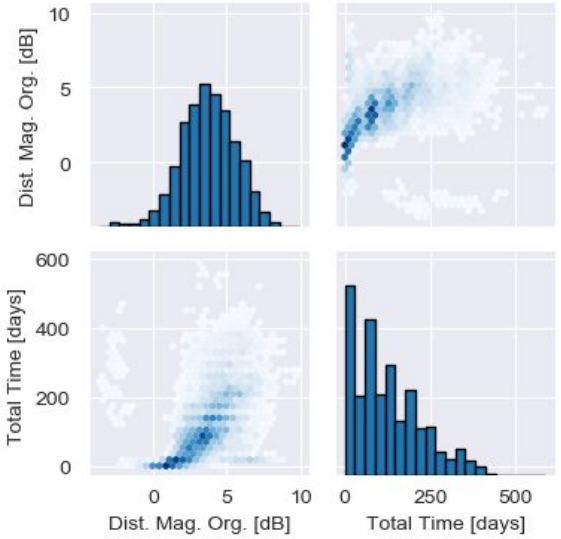




Time series type



Index interrelation



C-SCALE: Implementation Framework

JS

S1 ARD Image Tiles

Georeferenced, BUT NOT Datacube yet:

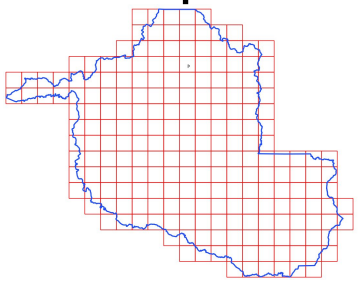
- GRD granules
- irregular time stamps
- along-track geometry and split
- missing dates

collection2xarray

Data Pool

single-date, merged, and tilled Images

xarray



Crop & Transform BASH

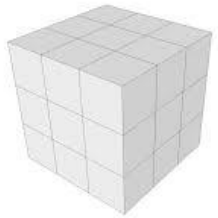
Disturbance Image Tiles

TS Categories & Features

Disturbance Map

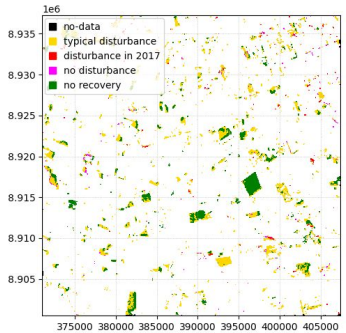
- Georeferenced
- Wall-to-wall coverage

6-day S1 Data Cube with Disturbance-Recovery Features

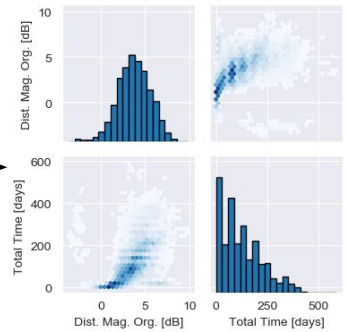


xarray

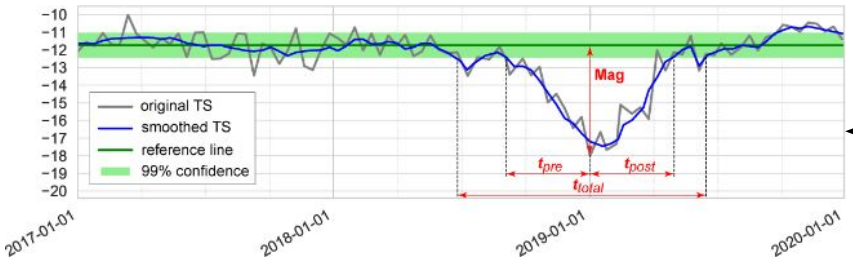
Mapping



Analysis



Visualize TS



C-SCALE: Implementation Framework

eodc

Data transfer

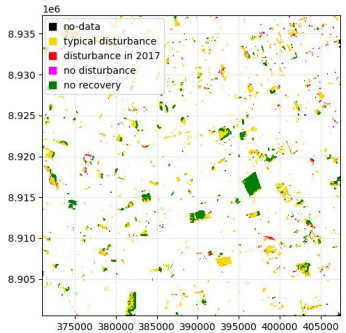
SURF

EUROPEAN OPEN SCIENCE CLOUD

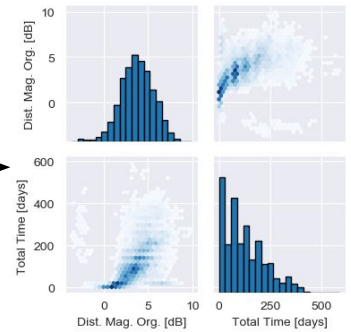
6-day S1 Data Cube with Disturbance-Recovery Features

xarray

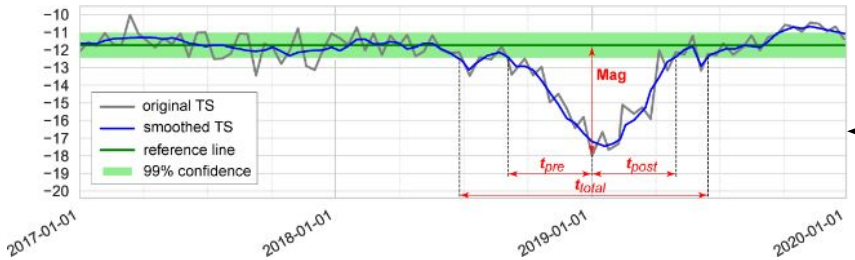
Mapping



Analysis



Visualize TS



C-SCALE: Implementation Framework

eodc

Data transfer

RCLONE

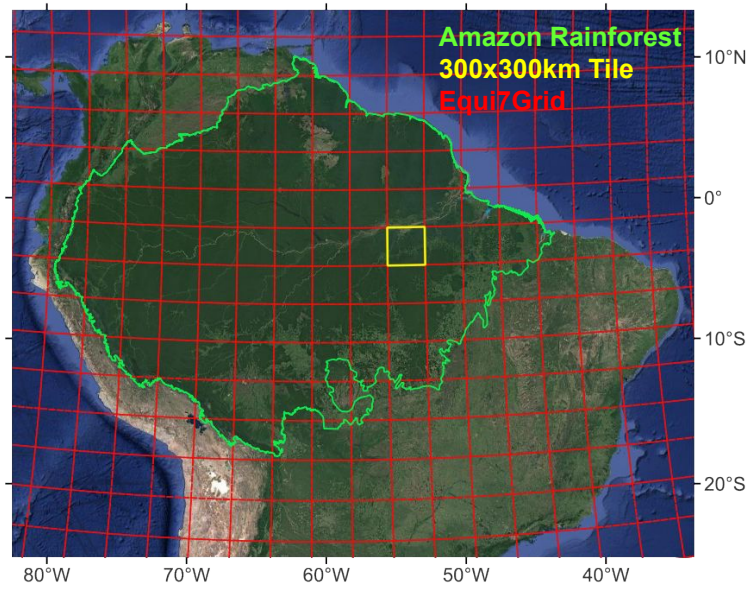
SURF

EUROPEAN OPEN SCIENCE CLOUD

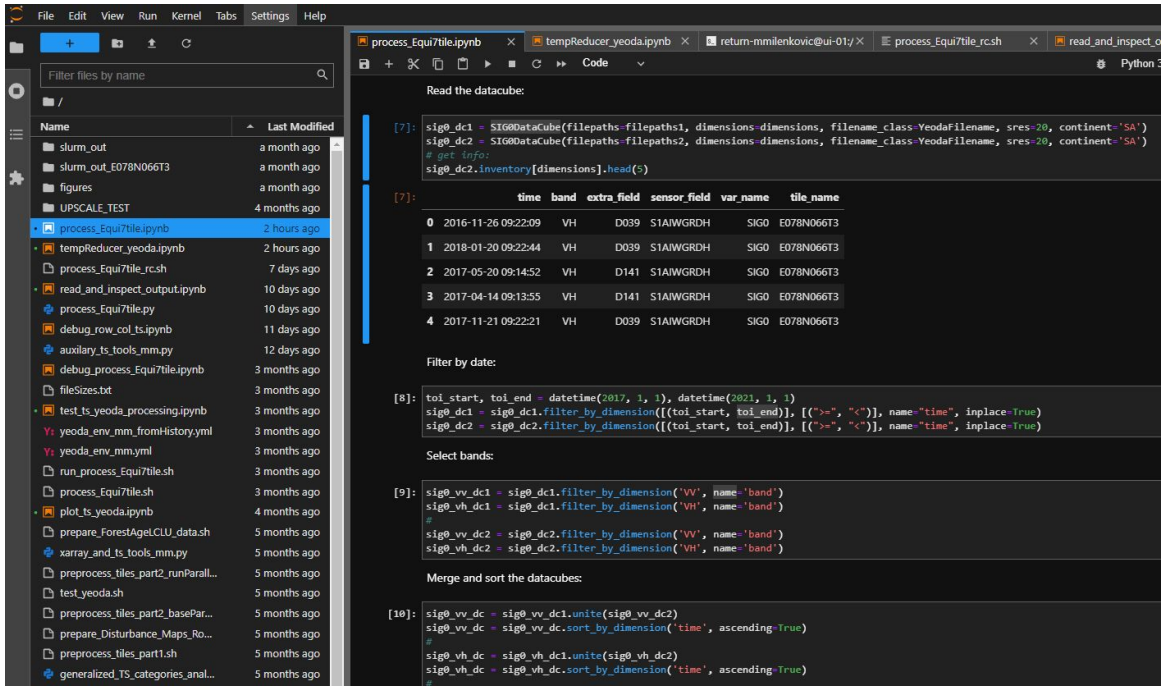
xarray

yeoda

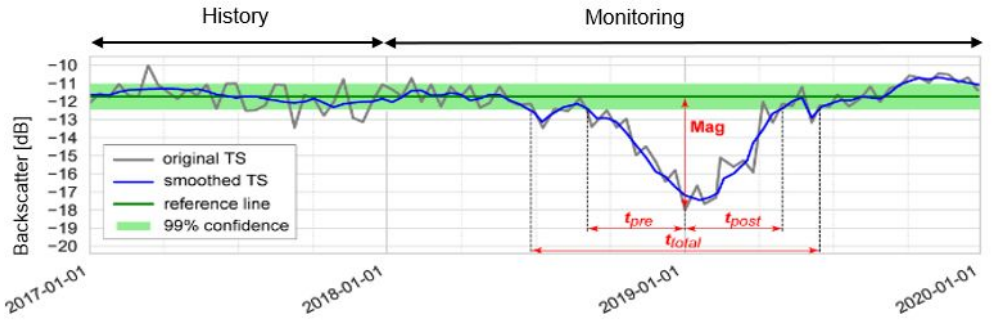
slurm workload manager



Credits: Bernhard Bauer-Marschallinger, TU Wien



C-SCALE: Computational Setup



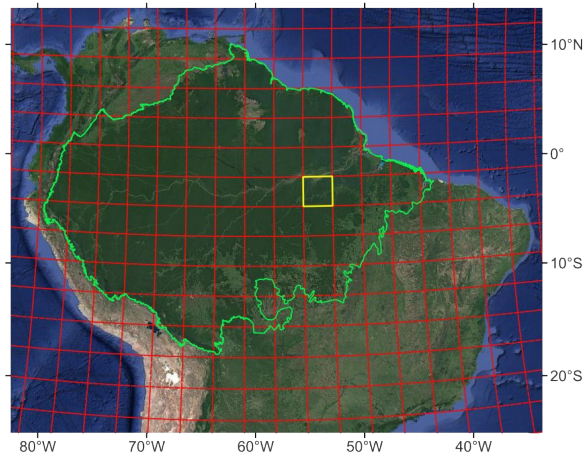
Multi-level Spatial Indices:

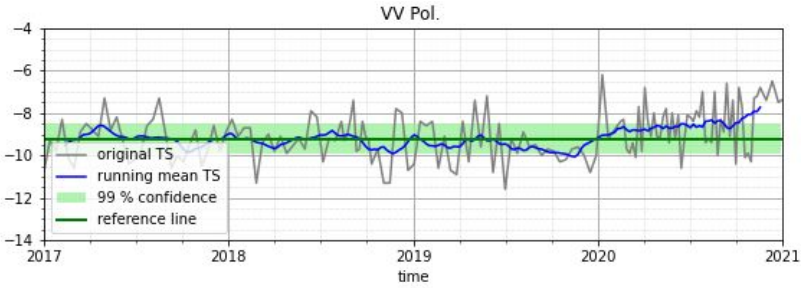
- 106 Equi7Grid Tiles 300 x 300 km^2
- 18170 Processing chunks (1000 x 1000 pixels) x ~1050 images
- **18.1 billion of pixels** - 20 x 20 m in size

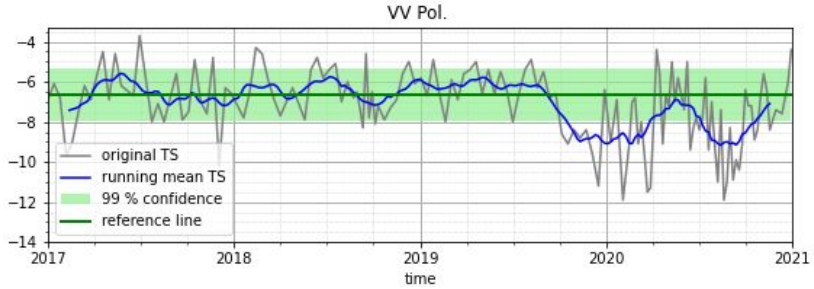
Processing characteristics:

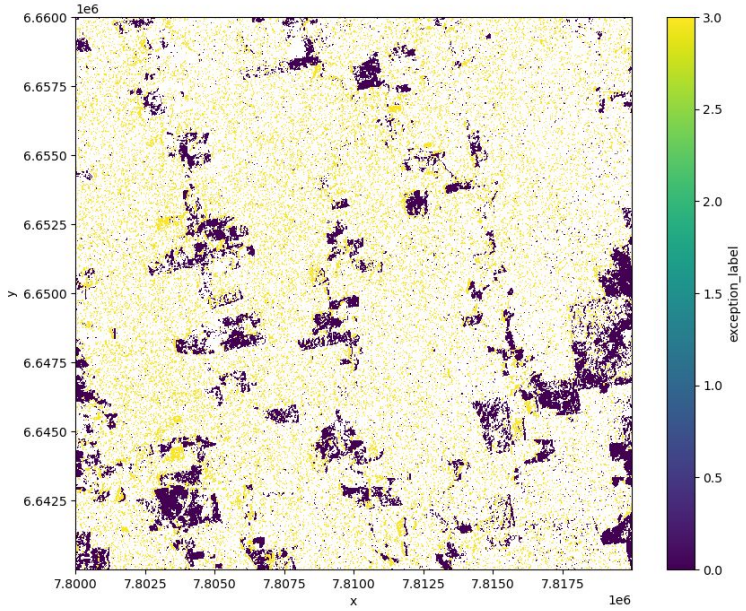
- 200 concurrent jobs (800 max)
- ~11h processing time / single chunk □ 36.4 km^2/h
- 3 cores/job
- **0.6 M CPUh and 41 day in total**

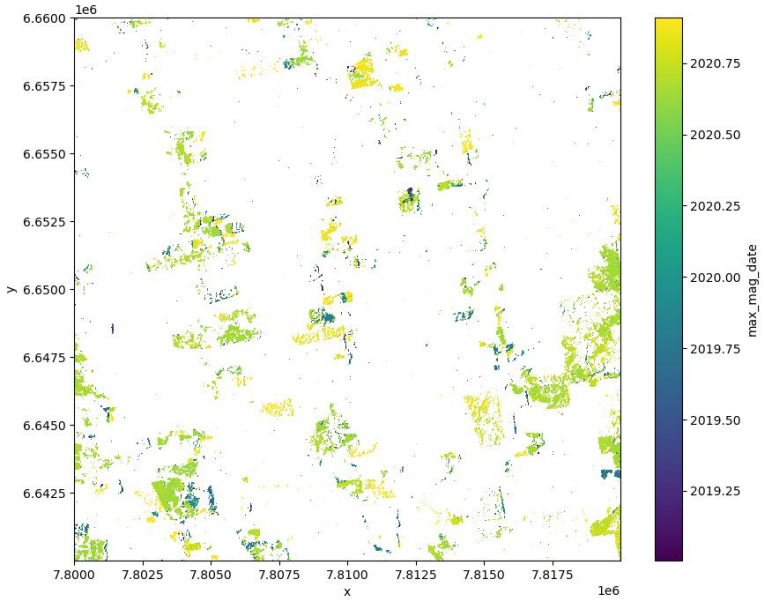
C-SCALE: Preliminary Results

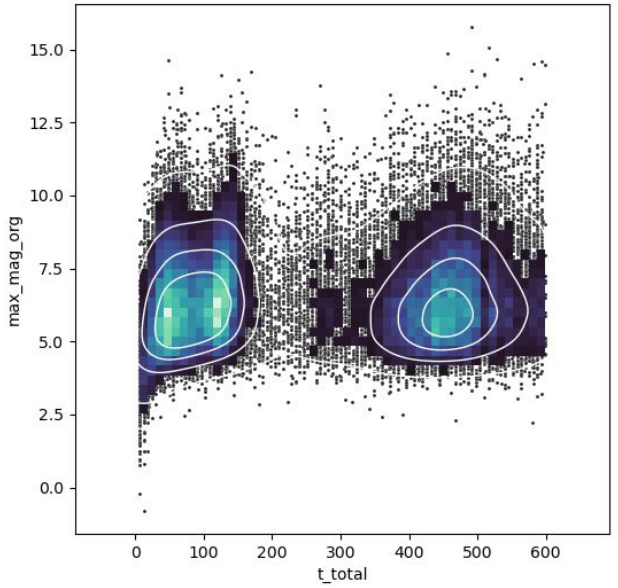










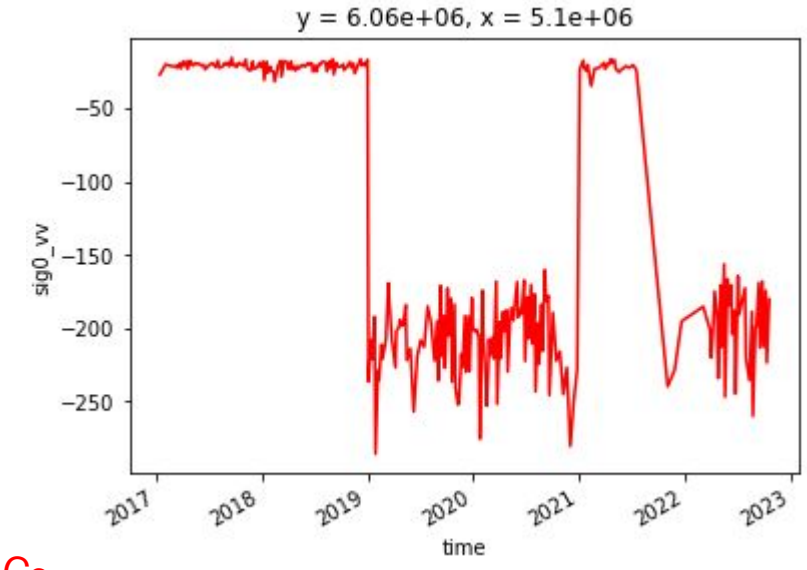


Lessons (hopefully) learned

Concurrent jobs and potential python errors

```
[return-mmilenkovic@ui-01 use-case-return]$ squeue -u $USER
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
3343832_[3-4%2]    normal process_ return-m PD       0:00      1 (JobArrayTaskLimit)
      3343832_1    normal process_ return-m  R    5:27:32      1 wn-hb-03
      3343832_2    normal process_ return-m  R    5:27:32      1 wn-hb-03
```

Differently scaled data



Corrupted tiff files

```
[return-mmilenkovic@ui-01 S1_SA_output_new]$ cat slurm_out/2999280_9.out
Mon Sep 19 09:56:01 CEST 2022
wn-hb-04.spider.surfsara.nl
return-mmilenkovic
START PROCESSING process_Equi7tile_rc.sh
srun python /home/return-mmilenkovic/mm_tools/use-case-return/process_Equi7tile.py -t 'E078N060T3' -r 1 -c 10 -s 1000 -o '/project/return/Share/mm/S1_SA_ou
tput_new/E078N060T3/'
ERROR 1: TIFFFetchDirectory:/project/return/Share/EODC_SA020M/V1M0R1/E078N060T3/SIG0_20180807T091643__VV_D141_E078N060T3_SA020M_V1M0R1_S1AIWGRDH.tif: Can n
ot read TIFF directory count
ERROR 1: TIFFReadDirectory:Failed to read directory at offset 38
```

Running archive at the data provider side

```
rclone copy --dry-run -i -c eodc_surf_sftp:S1_CSAR_IWGRDH/SIG0/V1M0R1/EQUI7_SA020M/V1M0R1/ /project/return/Share/EODC_SA020M/V1M0R1/

Transferred:          2.405 TiB / 2.405 TiB, 100%, 64.246 GiB/s, ETA 0s
Checks:         163743 / 163743, 100%
Transferred:     66769 / 66769, 100%
Elapsed time:      40.1s
```

RCLONE

Conclusions

Interface – built by the user

- Time-consuming to build
- Potential barrier for a general user

Data – transfer is required

- Data transfer is error prone

Code – flexible & manual upscale

- Debugging friendly
- Taking care of upscaling

	openEO Platform	C-SCALE	GEE
User Interface	In-place	Build by user	In-place
Data	In-place	Data transfer required	In-place
Code	Predefined + UDFs	Flexible	Predefined
Upscaling	Automatic	Manual	Automatic
Reproducibility	Fully	Fully	?

OPEN EARTH MONITOR

Thank you!

Milutin Milenković (milenkovic@iiasa.ac.at)
Novel Data Ecosystems for Sustainability, IIASA