SCALABLE AND ENERGY EFFICIENT COMPOSITING OF SENTINEL-2 TIME SERIES

Pablo d‘Angelo, Paul Karlshöfer, Uta Heiden

DLR, Remote Sensing Technology Institute

Bare surface composites
Scalability, Energy efficiency

- Soil is important part of earths ecosystem
- Bare surface composite
  - Average reflectance of bare surface pixels, statistical products and quality layers
  - Applications: Soil Health, Agriculture, Erosion, …
- Long time series needed to catch bare soil occurence.
  - “Clean” reflectance, without vegetation or plant residue
  - Cloud Issues
- What about the environmental impact of (re)processing 431 TB (445806 S2 scenes)?

Sentinel-2 Image archive

Reflectance composite

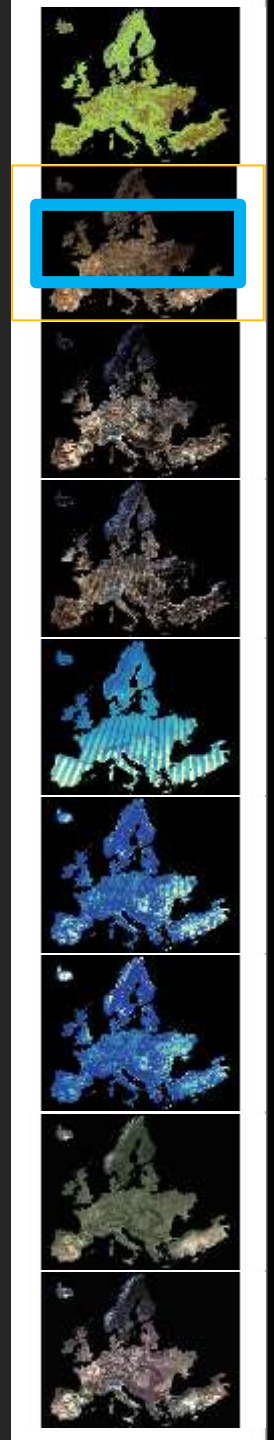
Bare Surface Reflectance composite

SoilSuite 2018 – 2022

- Sentinel-2
- 2018 – 2022
- < 80 % cloud cover
- > 20° sun elevation
- 20 m pixel size
- 10 bands

Bare Surface Reflectance Composite

– Mean



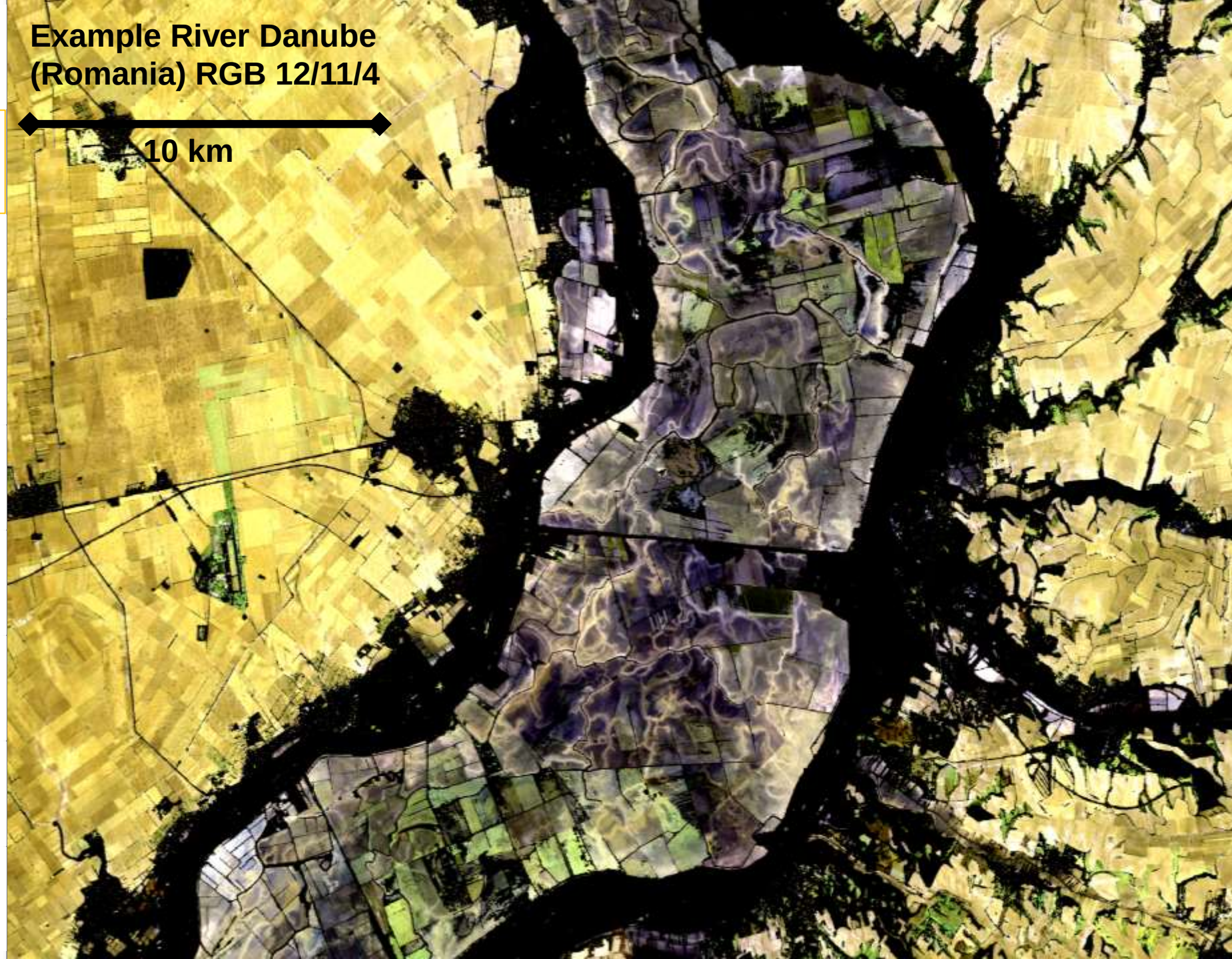
DOI: 10.15489/qkud8cudg596

SoilSuite 2018 – 2022

- Sentinel-2
- 2018 – 2022
- < 80 % cloud cover
- > 20° sun elevation
- 20 m pixel size
- 10 bands

Bare Surface Reflectance Composite

– Mean

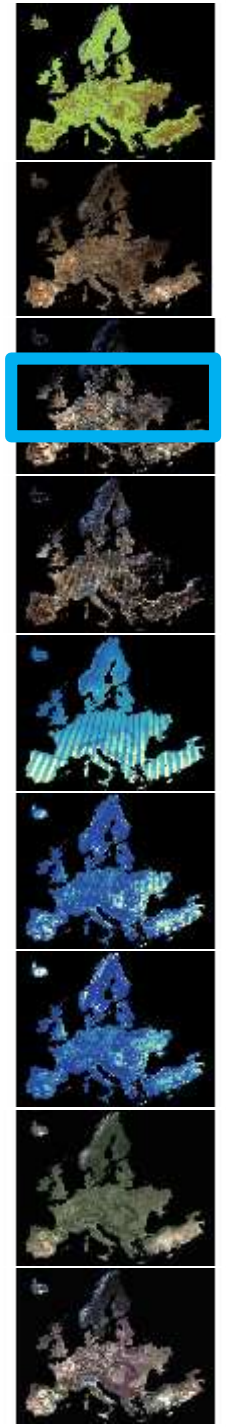


SoilSuite 2018 – 2022

- Sentinel-2
- 2018 – 2022
- < 80 % cloud cover
- > 20° sun elevation
- 20 m pixel size
- 10 bands

Bare Surface Reflectance Composite

– Standard deviation



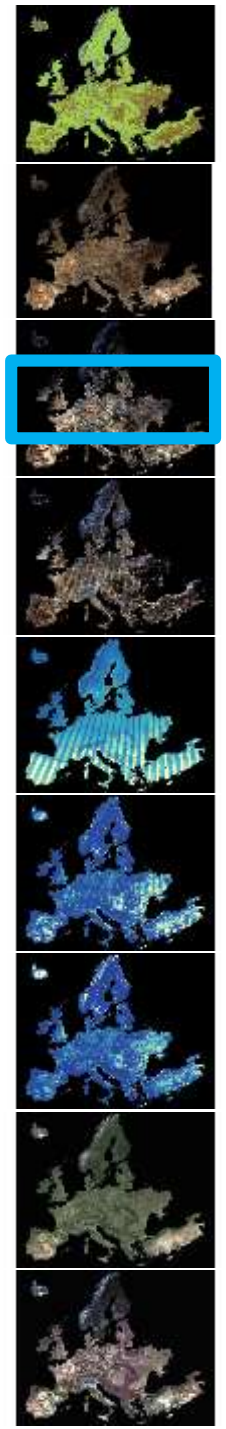
DOI: 10.15489/qkud8cudg596

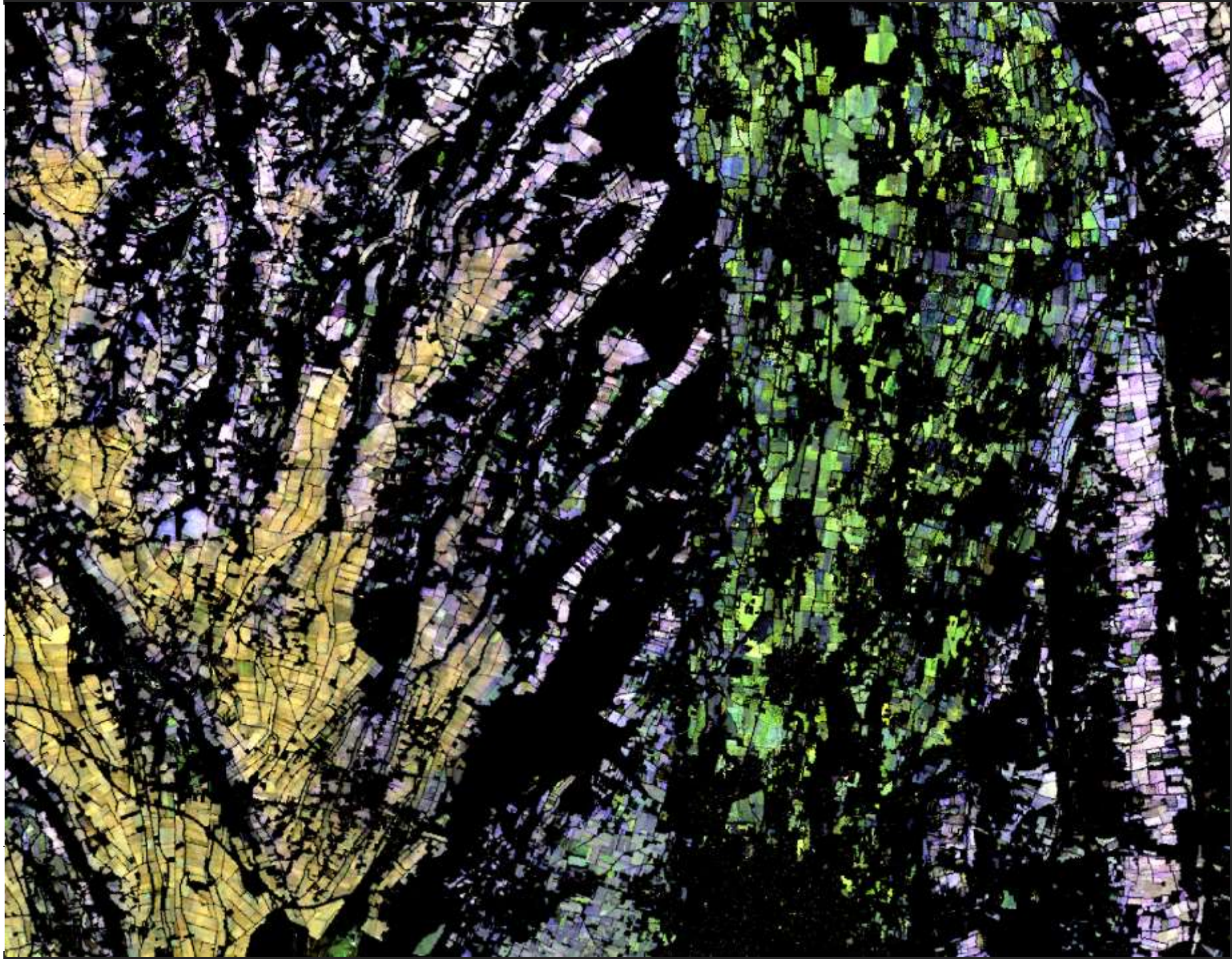
SoilSuite 2018 – 2022

- Sentinel-2
- 2018 – 2022
- < 80 % cloud cover
- > 20° sun elevation
- 20 m pixel size
- 10 bands

<u>Bare Surface Reflectance Composite</u>

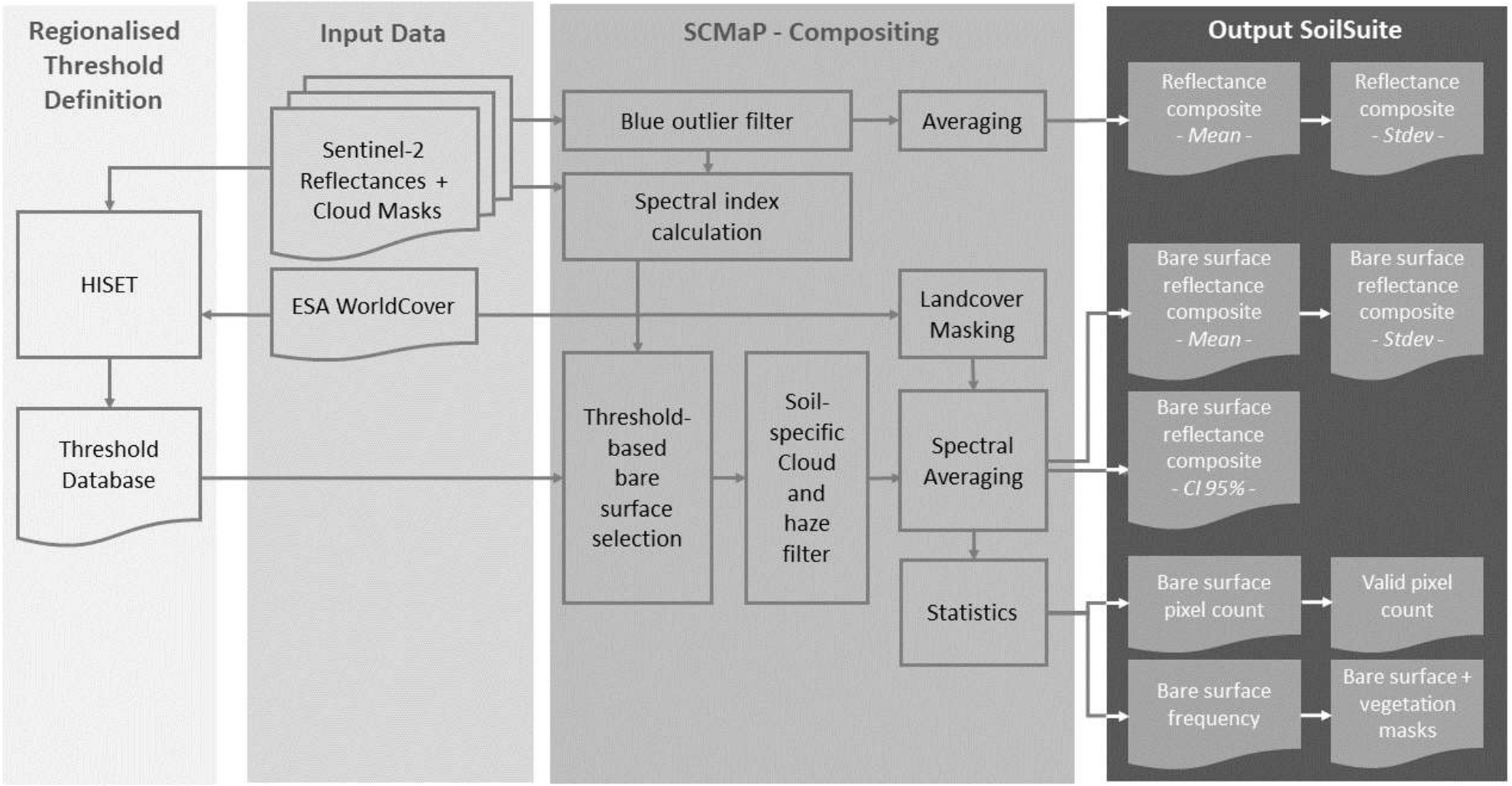
– Standard deviation





SoilSuite for Europe

Soil Composite Mapping Processor (SCMAP)



Karlshöfer et al. 2024 https://doi.org/10.1016/j.geoderma.2025.117340

Heiden et al., 2025: Tech report. https://doi.org/10.15489/qkud8cudg596

Heiden, U. et al.,2022: https://doi.org/10.3390/rs14184526

Rogge et al., 2018: https://doi.org/10.1016/j.rse.2017.11.004

Computation platform terrabyte

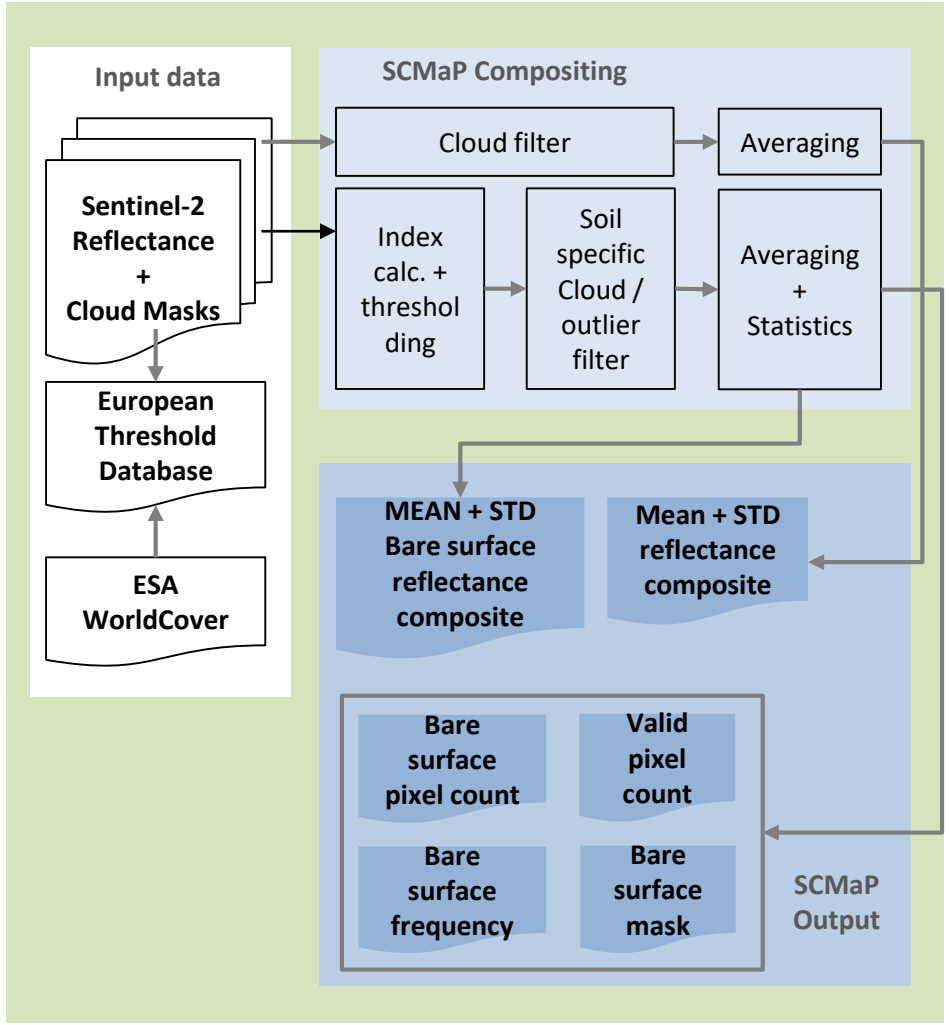
- Joint Leibnitz Rechenzentrum (LRZ), DLR
  High performance data analytics platform
- Compute Cluster
  - 272 CPU-Nodes with ~44.000 Intel Xeon Platinum CPU vCores und 273 TB RAM
  - 48 Nvidia A 100 GPU-Nodes with 188 GPUs and 47 TB RAM + 14 TB GPU RAM
- Online-Storage (DSS), EO data archive
  - 53 PB net, 6 PB user space
  - GPFS Filesystem with 360 Gbit/s Infiniband
- Highly efficient hot water cooling (~3% overhead)
- User friendly cloud services
  - STAC API, Jupyter, remote desktop etc.
- More info: Talk by Peter Friedl, 14:00 Hall L3

slido.com #3944109

Guess SoilSuite energy requirements

SCMaP Implementation

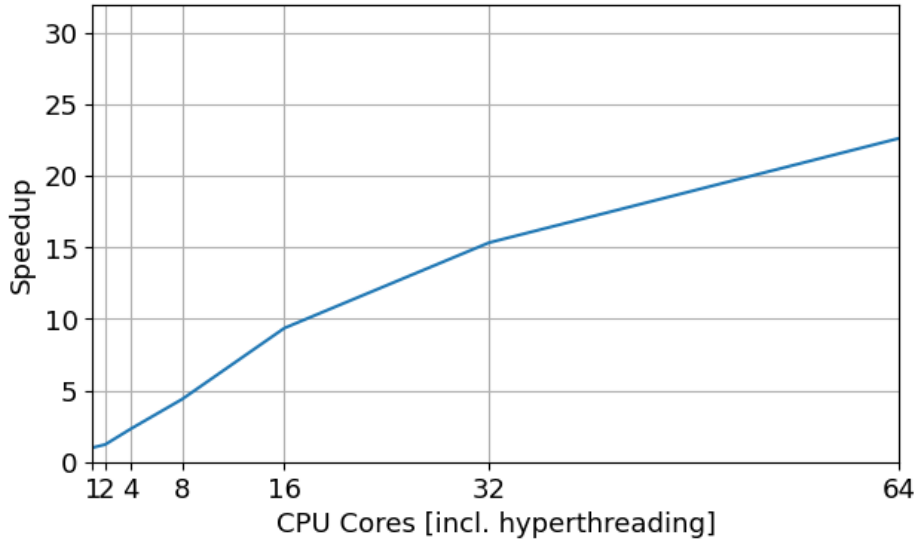
- Python frontend/orchestration
  - Multi Platform support: local workstations, cloud platforms (OTC), HPC
  - Input from: file system, STAC collections, Google cloud storage, any source supported by EODAG
- Core algorithm: C++ Implementation
  - Data I/O via GDAL
  - Local (per-pixel) computations, processing in chunks
  - Parallel processing using OpenMP both at data loading and compositing stage.
- Python product formatting
  - Applying cutlines, COG generation, reprojection etc.

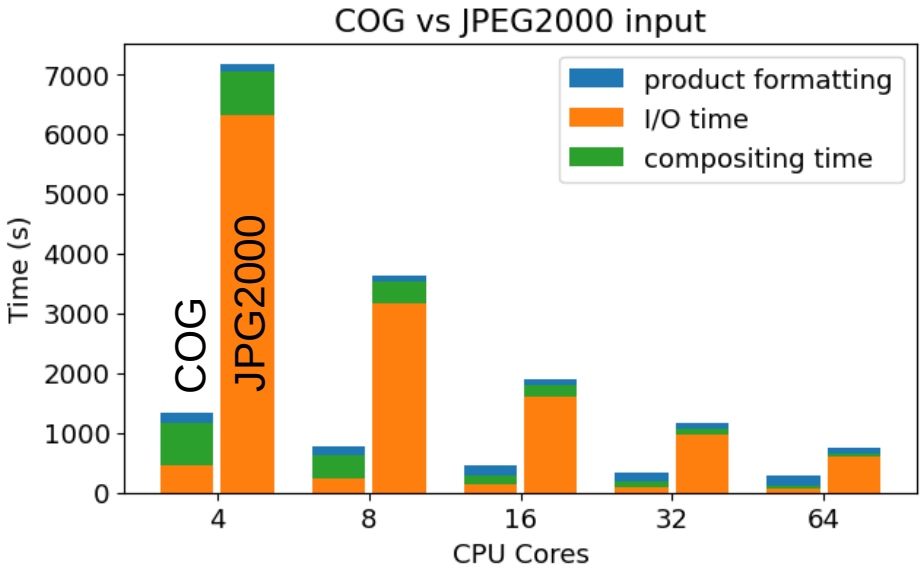


SCMaP Implementation Scalability

- Parallelization:
  - Linear speedup up to 32 threads
  - Balanced IO vs compositing times
  - Product formatting step does not scale well.

- Impact of input file formats
  - cloud optimized GeoTIFF vs JPEG 2000
  - JPEG 2000 6 times slower





Continent wide processing
Format and IO limits.

- Naïve parallel processing I/O limited
  - 80 parallel jobs
  - 2000 to 10000 COG files per job
  - GPFS file system optimized for large chunks
    - Parallel reading of tiny chunks from ~400000 simultaneously open files overloads file system
    - Solution: Parallel loading/decompression of full time series into RAM.
- Scalable processing, EU in 4:08 h
  - single SLURM array job for 1150 S2 Tiles
  - Input data throughput: ~200 Gbit/s
- Power consumption: **26 kWh**

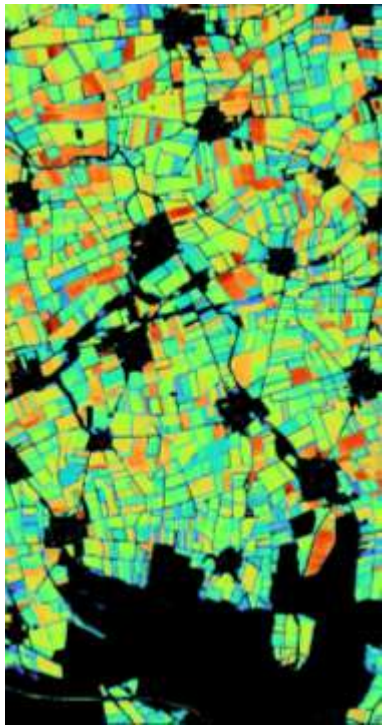
S2 tile grid and bare surface composite

Conclusions

- SoilSuite energy consumption equivalent to driving 150 km with an electric car.
  - Consider algorithms, data formats and platform.
  - Importance of data storage formats.
    - Don’t ignore time series access patterns.
    - Is multiple files per band storage really the best way?
  - “High performance” in HPC is true.
- Future
  - Data storage vs compute costs?
  - Release SCMaP as OpenEO custom function.
  - Comparison to resource consumption of other continent wide products?

Open Data access

Bare surface



Bare frequency

[Poll results]

Poll results