Global-scale analysis of Sentinel-1 satellite data at the Earth Observation Data Centre

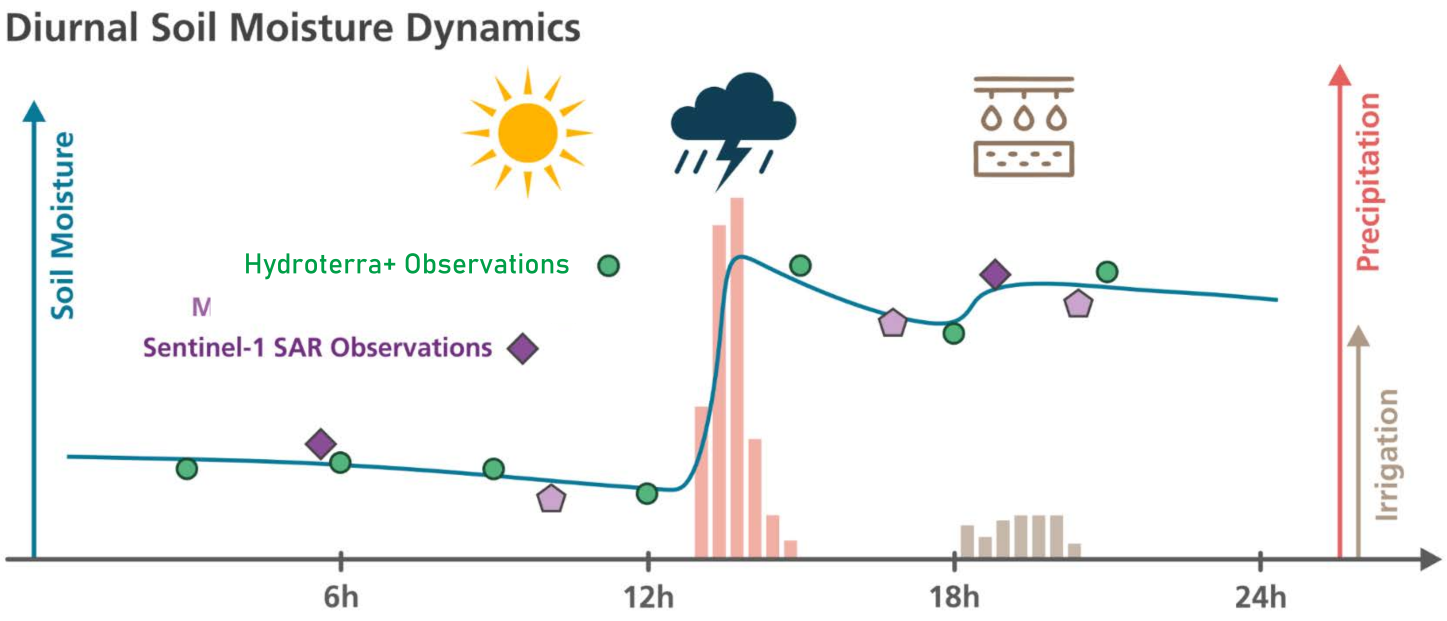
Wolfgang Wagner

Department of Geodesy and Geoinformation
Technische Universität Wien

Earth Observation Data Centre for
Water Resources Monitoring

Our Research Focus: Hydrologic Monitoring

- There is an urgent need for improving hydrological monitoring capabilities
  - Climate change pushes the water cycle into unknown territory
  - The terrestrial water cycle is not natural anymore
- Hydrological monitoring is challenging because
  - Hydrological processes are very dynamic → a short revisit time is essential
  - Anthropogenic activities take place at small space scales → a high spatial resolution is needed
  - Extremes need to be put in context → long and consistent time series are needed



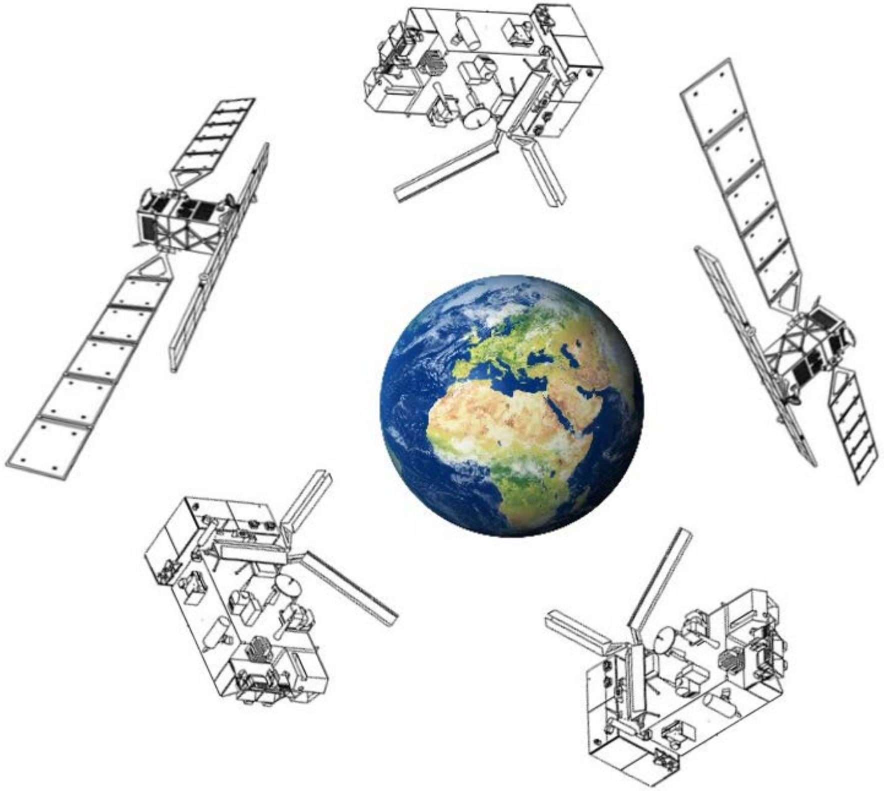
Scatterometer and SAR Satellite Constellations

Metop ASCAT

Frequency: 5.255 GHz
Polarisation: VV

Sampling: 6.25 km
Daily coverage: 82%

Satellites
Metop-A: 2006-2021
Metop-B: 2012 ongoing
Metop-C: 2018 ongoing
Metop-SG B1: foreseen 2026



Sentinel-1 SAR IW

Frequency: 5.405 GHz
Polarisation: VV+VH

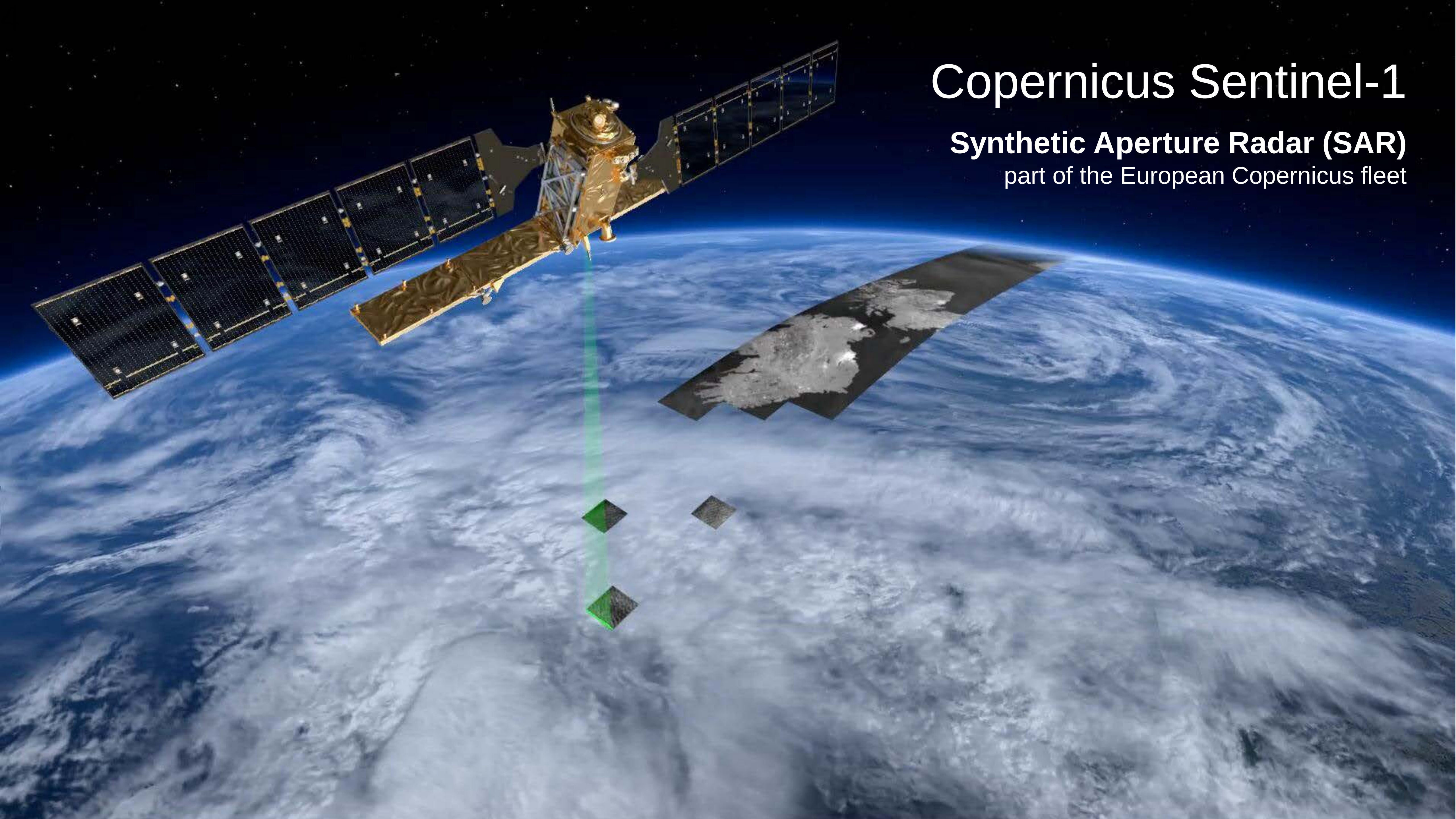
Sampling: 20 m
Repeat coverage: 3 – 12 days

Satellites
Sentinel-1A: 2014 ongoing
Sentinel-1B: 2016-2021
Sentinel-1C: foreseen 2024
Sentinel-1D: foreseen 2025

Achieve process understanding by combining the high temporal sampling capabilities of ASCAT and the high spatial resolution of Sentinel-1

Copernicus Sentinel-1

Synthetic Aperture Radar (SAR)
part of the European Copernicus fleet



CEMS Global Flood Monitoring (GFM) product

- **Sentinel-1** Synthetic Aperture Radar (SAR)
  - Systematic coverage
  - Currently only one satellite
- **Fully automatic processing** of all incoming Sentinel-1 scenes within 8 hours
- **Ensemble** of 3 flood mapping algorithms
  - LIST, DLR, TU Wien
- **Context** through 11 output layers incl.
  - Flood extent
  - Likelihood
  - Exclusion mask
  - Advisory flags

Wagner et al. (2020) Data processing architectures for monitoring floods using Sentinel-1, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2020, 641–648.

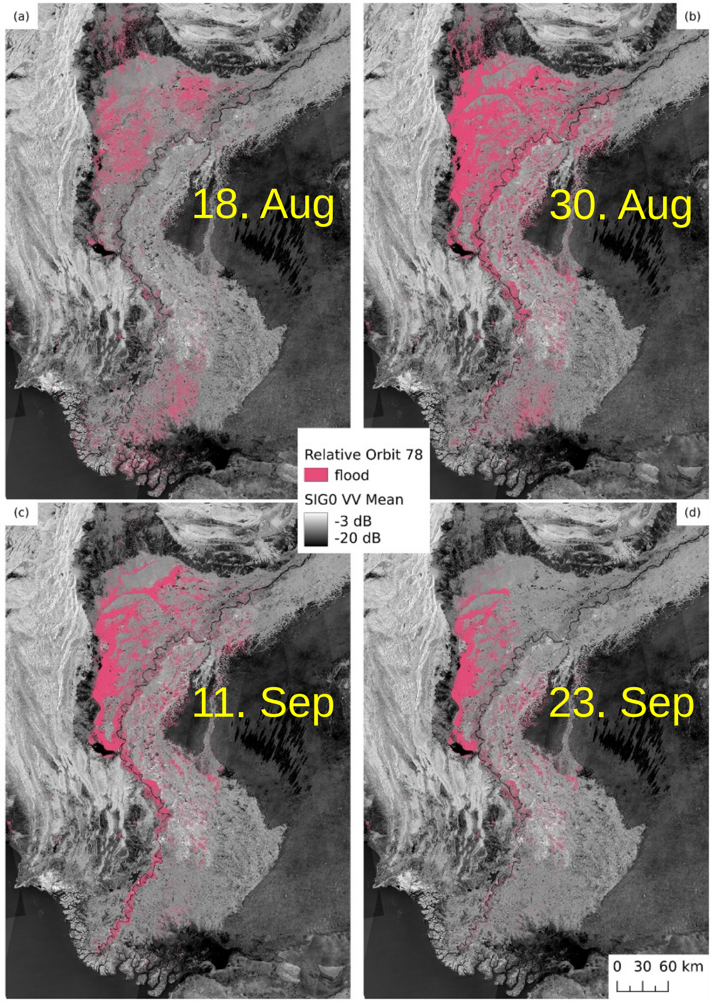
→ **Advantages**
- No time is lost due to human intervention
- Discover unreported events

→ **Disadvantages**
- False alarms
- Processing overhead

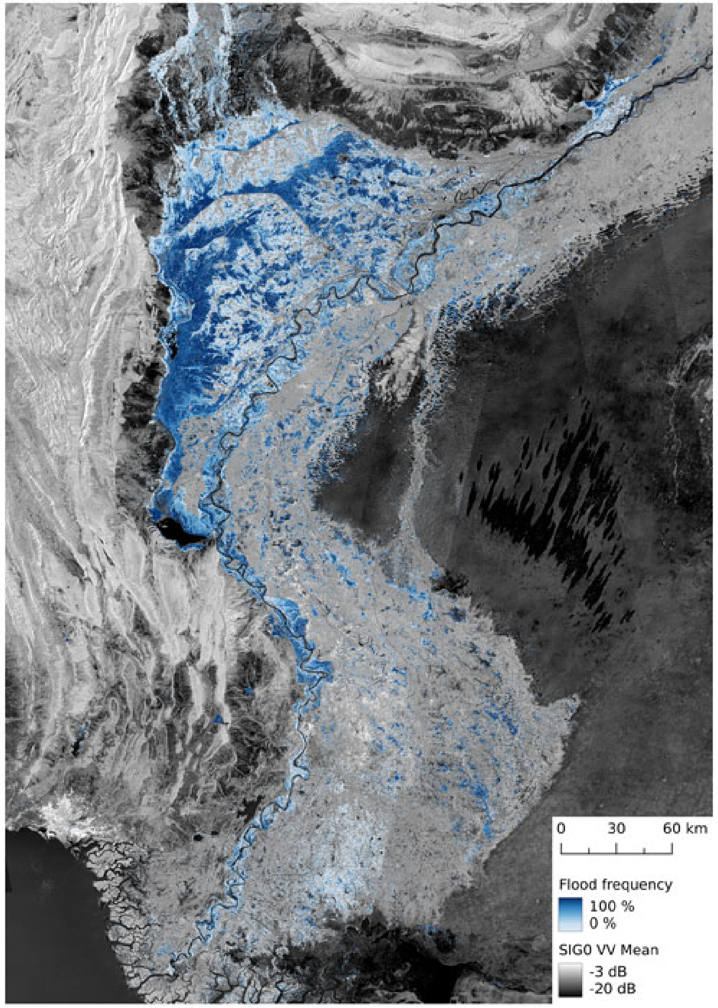
→ **Challenges**
- Accuracy
- Timeliness

Monitoring of the Flood in Pakistan in 2022

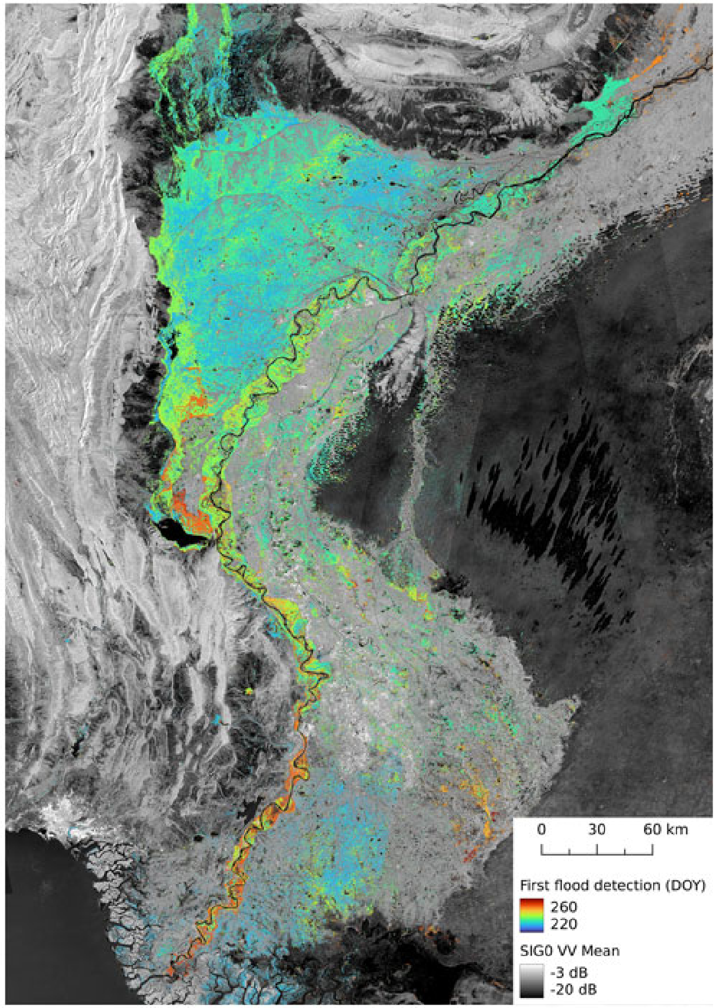
Individual Flood Scenes



Flood Frequencies

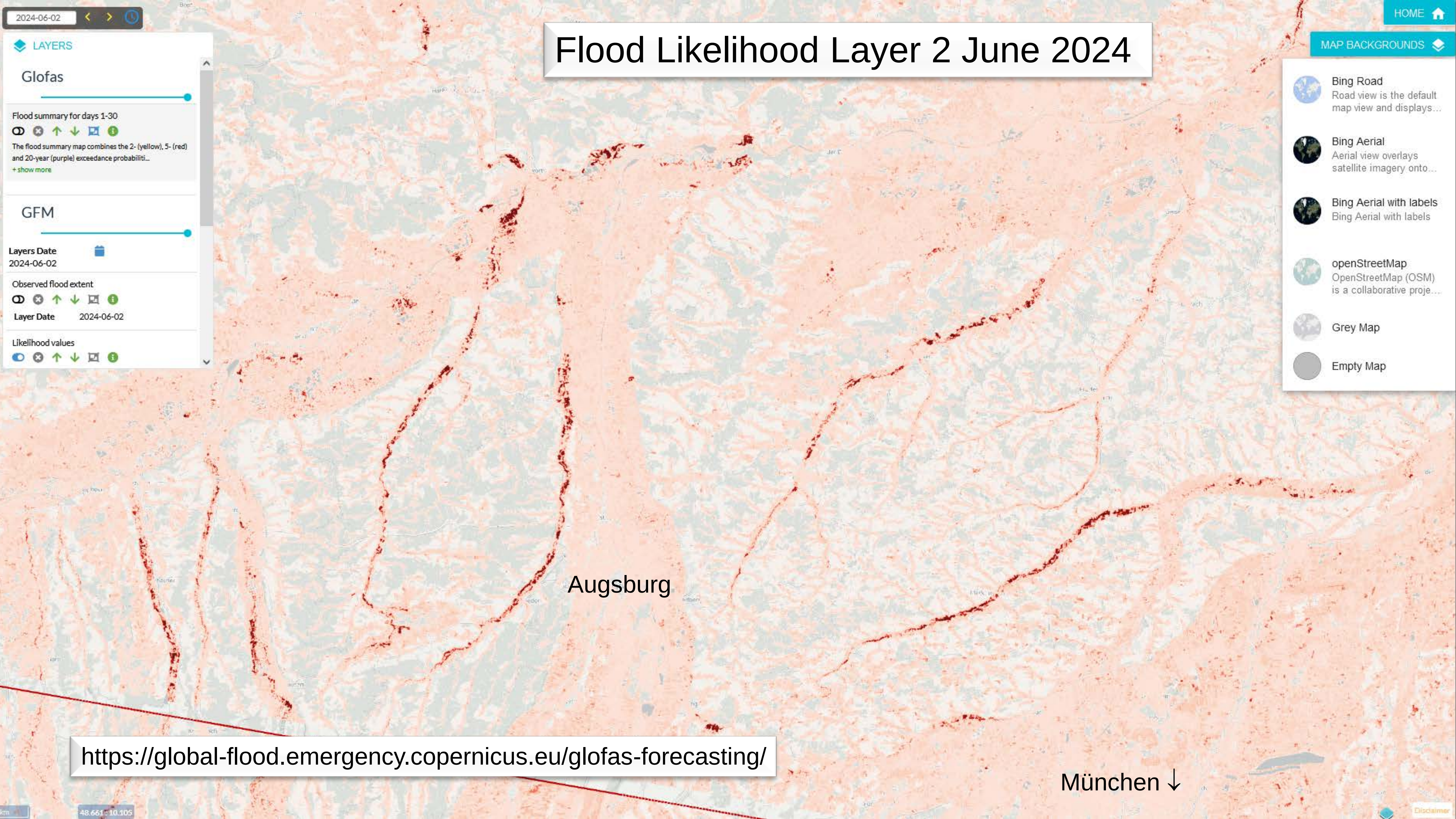


Day of First Flood Detection



Roth et al. (2023) Sentinel-1-based analysis of the severe flood over Pakistan 2022, Natural Hazards and Earth System Sciences, 23, 3305–3317.

Flood Likelihood Layer 2 June 2024

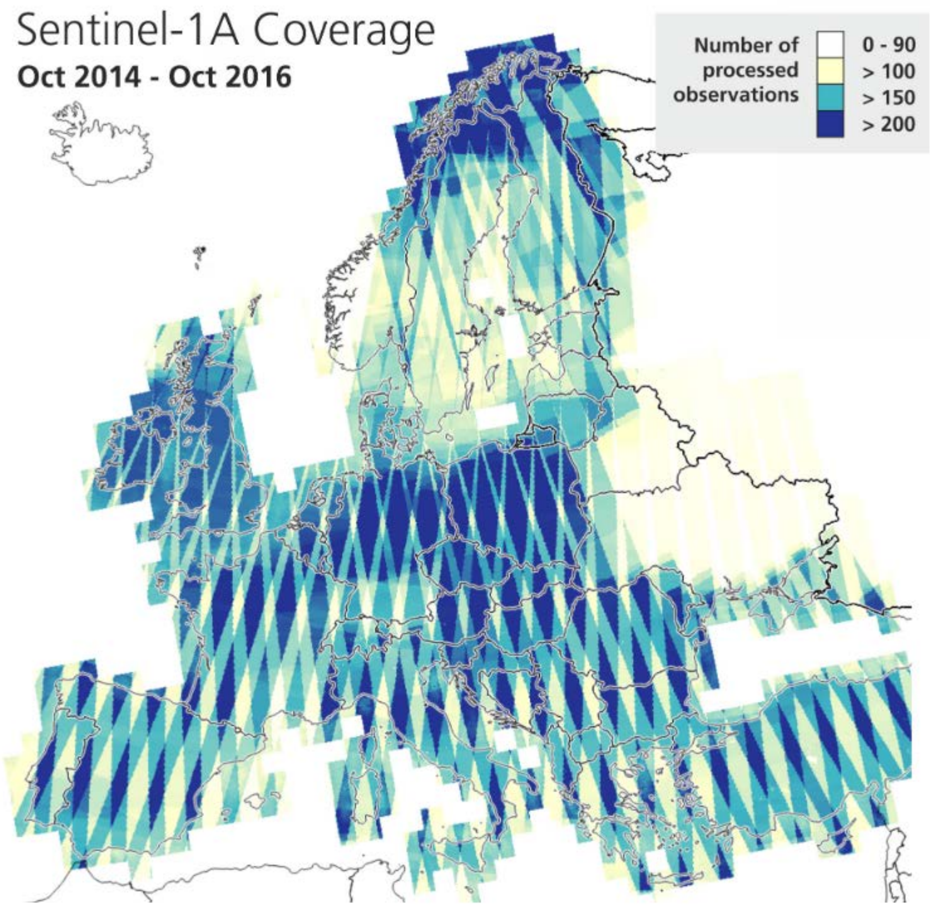


https://global-flood.emergency.copernicus.eu/glofas-forecasting/

CLMS 1km Sentinel-1 Surface Soil Moisture

Global Land Operations

- **Fully automatic processing** of all incoming Sentinel-1 scenes over Europe
  - Daily updated
- **Change detection** model of TU Wien
  - Backscatter is aggregated to 1km before soil moisture retrieval
- **Context** through additional layers incl.
  - Uncertainty estimate
  - Exclusion mask
- There will be a major **update in 2025**
  - Global coverage
  - Seasonal vegetation correction
  - Removal of non-soil-moisture-sensitive 20m pixels
  - Better masking of snow and frost
  - Masking of flood affected areas



Bauer-Marschallinger et al. (2019) Towards global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles, IEEE Transactions on Geoscience and Remote Sensing, 57(1), 520-539.

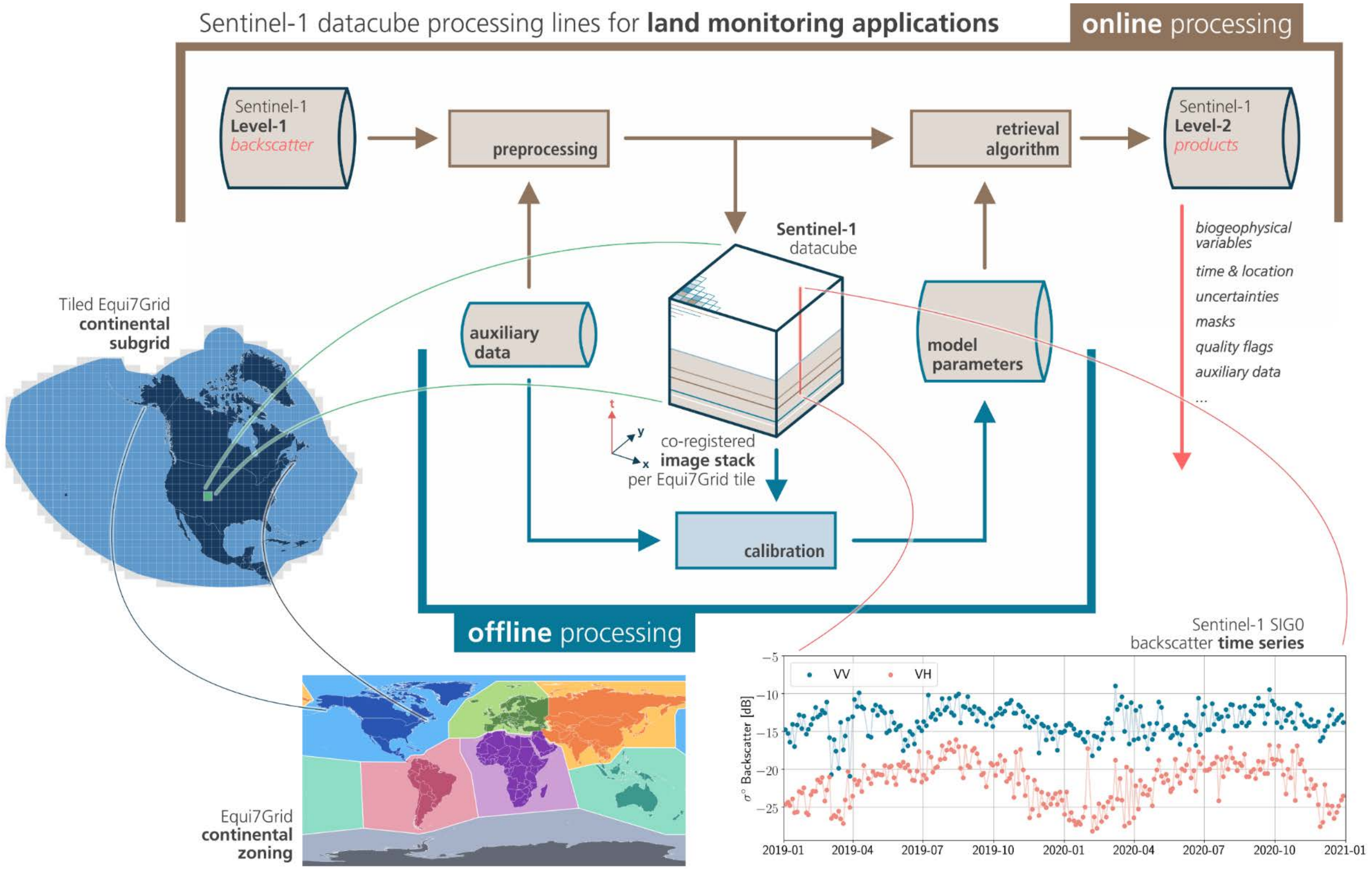
1km Sentinel-1 Surface Soil Moisture Retrievals

Digital Twin Earth (DTE) Hydrology Project

Mean-monthly Sentinel-1 surface soil moisture retrievals for the Mediterranean region from 2017 to 2021.

Quast et al. (2023) Soil moisture retrieval from Sentinel-1 using a first-order radiative transfer model - a case-study over the Po-Valley, Remote Sensing of Environment, 295, 113651.

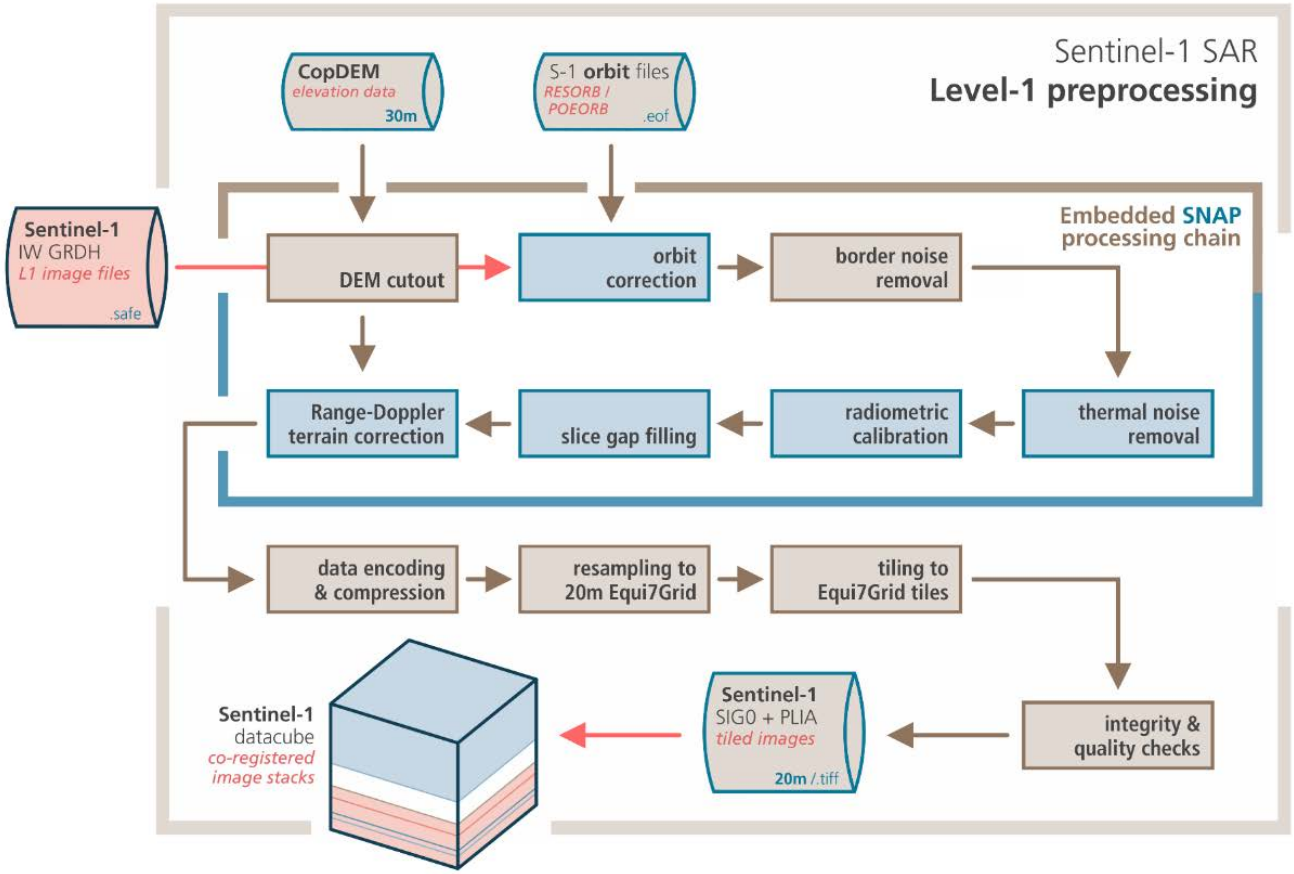
A Datacube for Dynamic Land Monitoring Applications



- Exploits the strengths of dense satellite time series
- Combines online and offline workflows
- Suited for training complex models
  - Physical models
  - AI models
  - Hybrid models

Wagner et al. (2021) A Sentinel-1 backscatter datacube for global land monitoring applications, Remote Sensing, 13, 4622.

Global Sentinel-1 Backscatter Datacube



Data Volume in TB

Level-1 Sentinel-1 IW GRD data

Year	Africa	Asia	Europe	NA	Oceania	SA	Total
2015	12.7	15.1	22.0	6.2	4.9	5.3	66.2
2016	20.6	19.2	31.9	11.5	6.6	9.0	98.8
2017	45.0	53.9	71.8	31.4	18.4	23.1	243.6
2018	48.0	58.1	70.3	35.3	20.2	24.7	256.6
2019	94.4	61.1	119.9	38.5	21.1	26.9	361.9
2020	97.3	63.3	130.7	41.4	21.3	28.6	382.6
Total	318.0	270.7	446.6	164.3	92.5	117.6	1409.7

20 m Sentinel-1 datacube

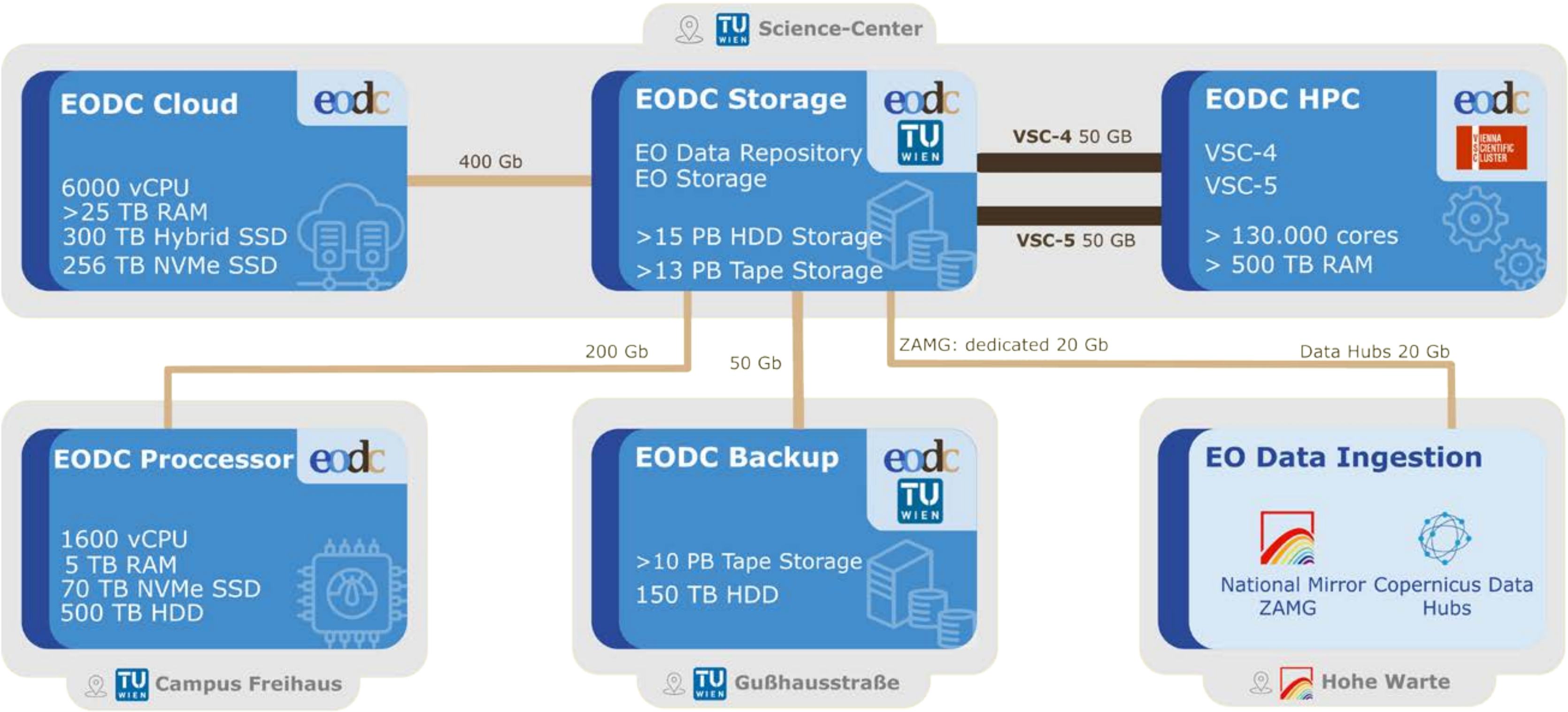
Year	Africa	Asia	Europe	NA	Oceania	SA	Total
2015	2.5	2.9	4.3	1.2	1.1	1.0	13.0
2016	4.4	4.0	6.4	2.5	1.5	1.9	20.7
2017	9.8	11.9	14.6	6.9	4.3	4.9	52.4
2018	10.3	12.8	12.8	7.6	4.7	5.2	53.4
2019	16.9	19.4	23.5	13.4	7.6	8.6	89.4
2020	17.3	20.1	25.0	14.6	7.7	9.4	94.1
Total	61.2	71.1	86.6	46.1	26.9	31.0	323.0

Wagner et al. (2021) A Sentinel-1 Backscatter Datacube for Global Land Monitoring Applications, Remote Sensing, 13, 4622.

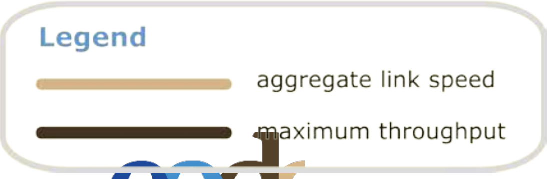
Earth Observation Data Centre

Collaboration for Earth Observation

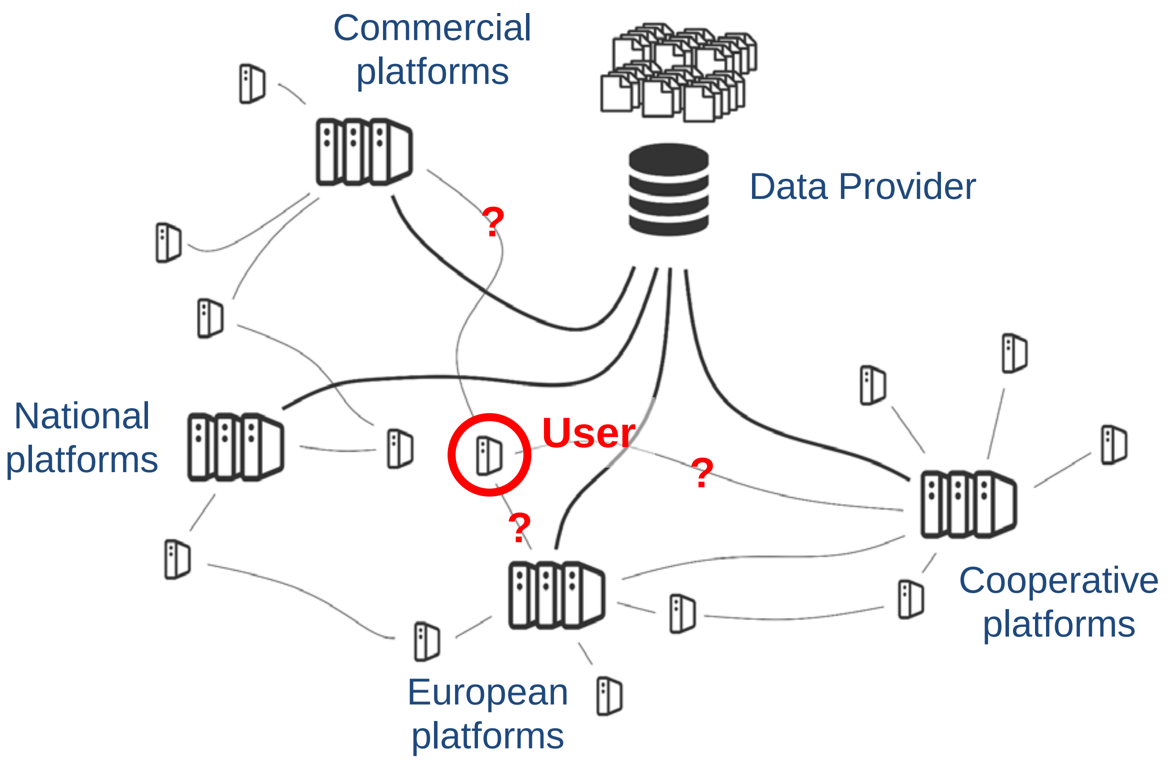
Petabyte Storage
Supercomputing
Cloud Platform



Wagner et al. (2014) Addressing grand challenges in earth observation science: The Earth Observation Data Centre for Water Resources Monitoring, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals), Volume II-7, 81-88.



The Proliferation of Non-Interoperable Big Data Platforms

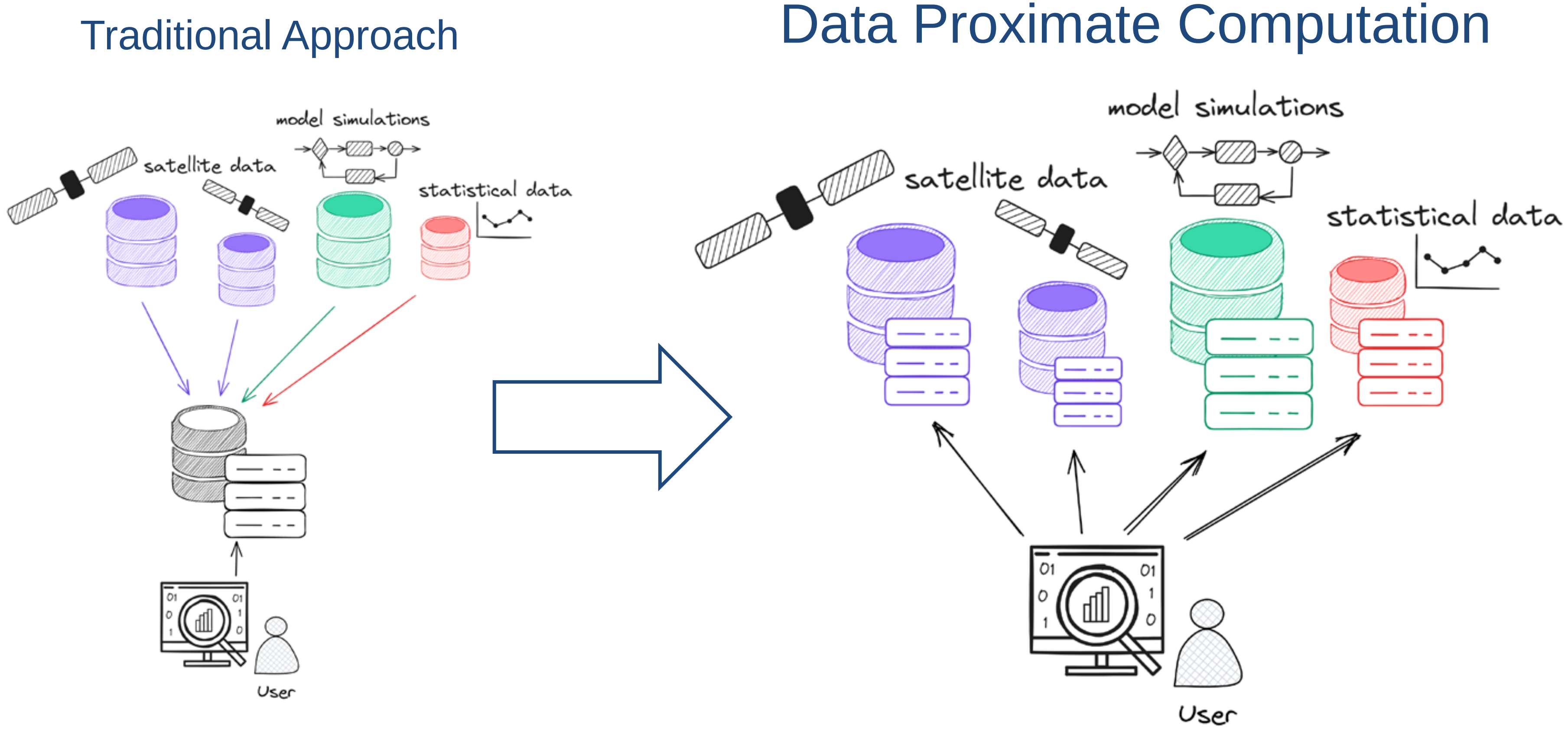


- Many different types of non-interoperable Earth observation platforms have emerged
  - Some thrived – others failed
- Bigger is not always better
  - Big data collections are nice to have but more important is to offer the best data
  - The diversity and complexity of scientific algorithms calls for very flexible solutions
  - Most users works with only small data volumes, i.e. data transfer is usually not the bottleneck
- There is no monolithic solution that can do it all

- Science has started to tackle these problems, e.g.
  - Schramm et al. (2021) The openEO API - Harmonising use of Earth Observation cloud services using virtual Data Cube functionalities, Remote Sensing, 13, 1125, 21p.
  - Backeberg et al. (2022) An open compute and data federation as an alternative to monolithic infrastructures for big Earth data analytics. Big Earth Data, 7(3), 1–19.

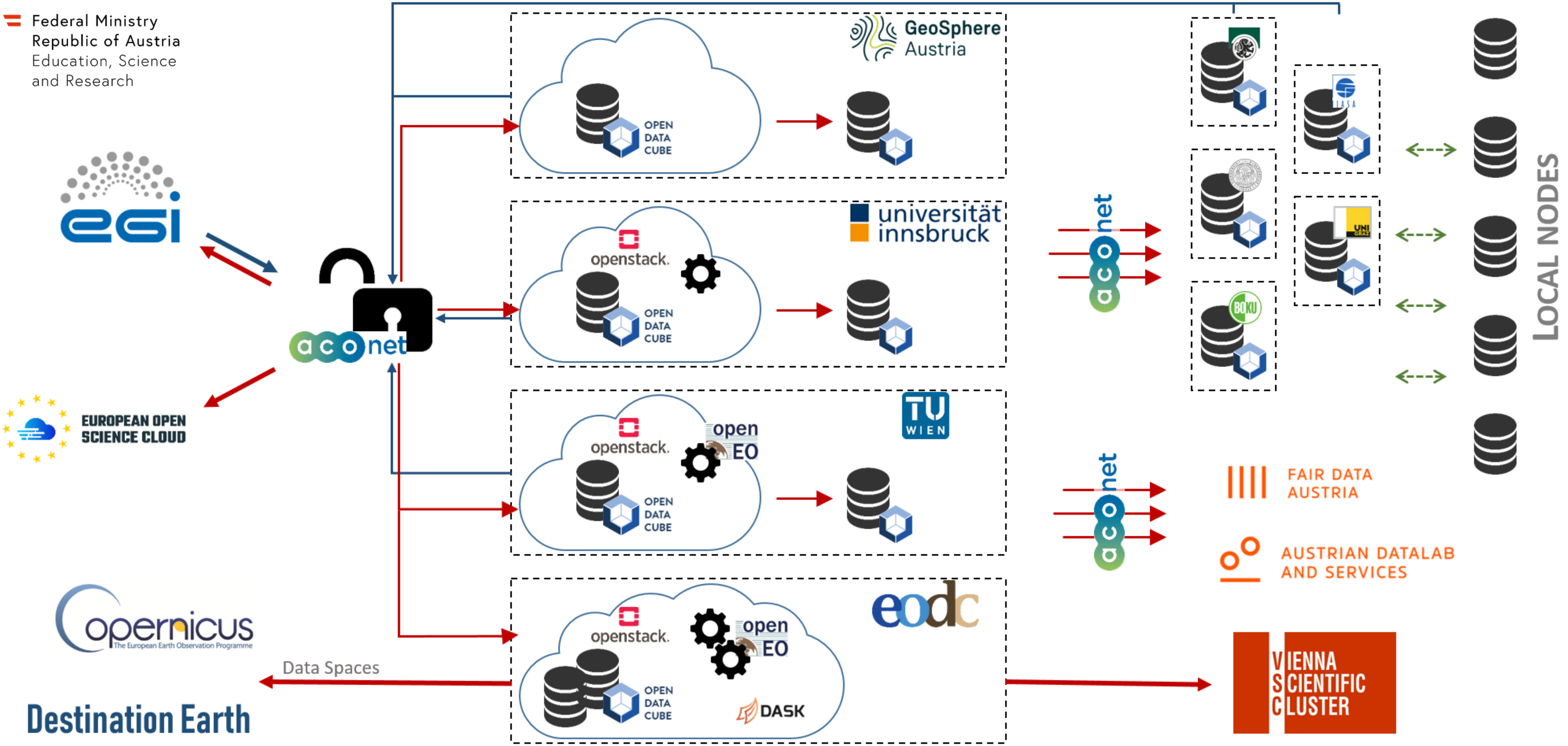
Towards a Web of FAIR Data and Services

- Powerful open source solutions has become available that will help turning the vision of the European Open Science Cloud into reality



Reimer et al. (2023) Multi-cloud processing with Dask: Demonstrating the capabilities of DestinE Data Lake (DEDL), Proceedings of the 2023 Conference on Big Data from Space (BiDS'23), 161-164.

Cloud4GEO



Wagner et al. (2023) Federating scientific infrastructure and services for cross-domain applications of Earth observation and climate data, BiDS’23 Proceedings, 93-96.

Strategies to Scale on Cloud Platforms

- Utilizing cloud platforms and open source tools, we're developing different strategies to cope with the processing needs of ever growing data volumes

Better Algorithms

Algorithm 1 Streamed Pearson r

$n, \mu_a, \mu_b, M2_a, M2_b, C \leftarrow 0$
for each $tile$ in datacube stream **do**
  $n \leftarrow n + 1$
  $\delta_a \leftarrow tile - \mu_a$
  $\delta_b \leftarrow tile - \mu_b$
  $\mu_a \leftarrow \mu_a + \frac{\delta_a}{n}$
  $\mu_b \leftarrow \mu_b + \frac{\delta_b}{n}$
  $\delta_{2a} \leftarrow tile - \mu_a$
  $\delta_{2b} \leftarrow tile - \mu_b$
  $M2_a \leftarrow M2_a + \delta_a * \delta_{2a}$
  $M2_b \leftarrow M2_b + \delta_b * \delta_{2b}$
  $C \leftarrow C + \delta_a * \delta_{2b}$
end for
$r \leftarrow \frac{C}{\sqrt{M2_a * M2_b}}$

Better Hardware

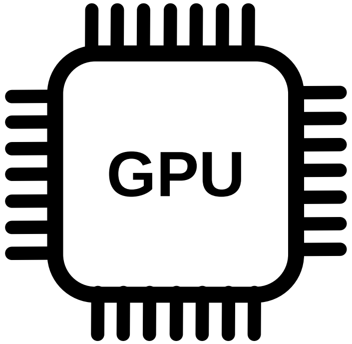
Better Frameworks

dask

xarray

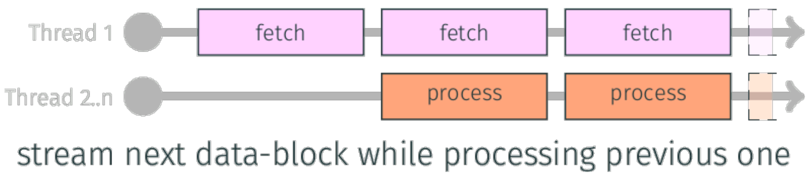
Zarr

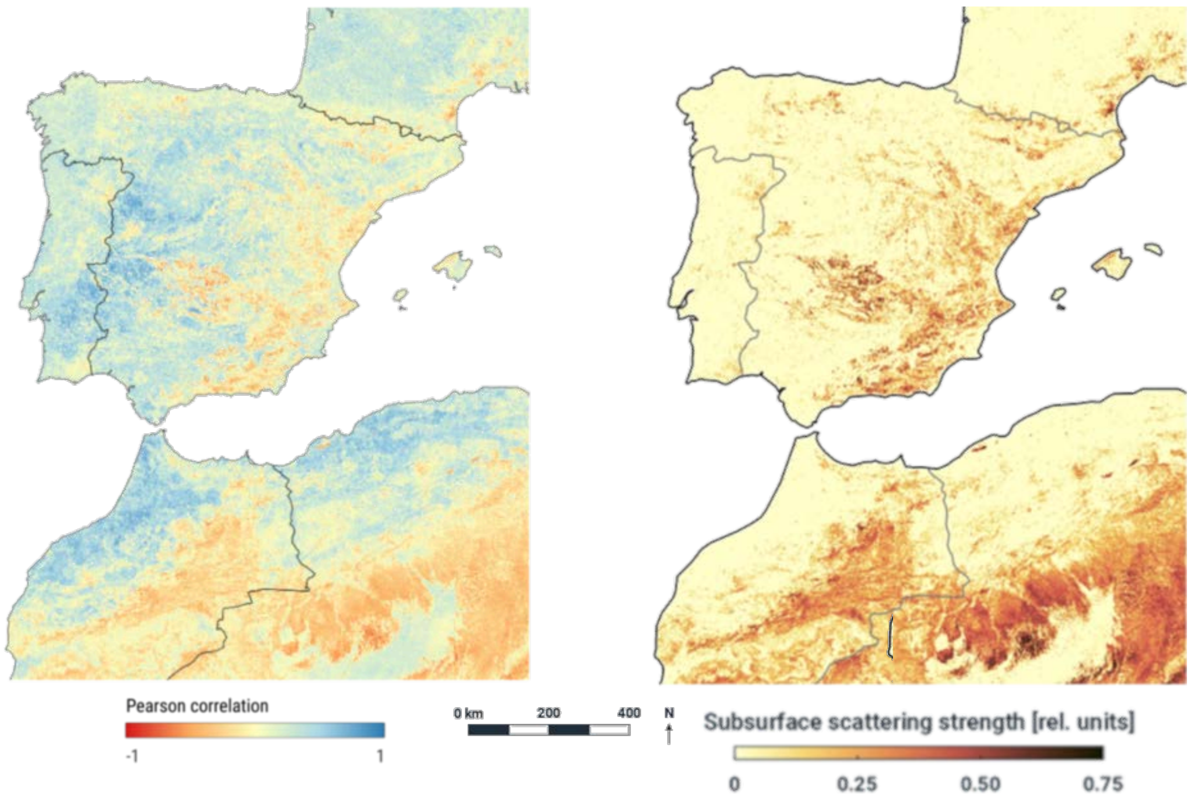
Specialized Hardware



Scaling by Algorithm

- Streaming algorithm to calculate Pearson correlation
- Optimized to work efficiently on GeoTif stacks, a common data format choice
- On a VSC4 Node with 64 cores we achieve a throughput of: 13.6 Gbps
- Calculating soil moisture sensitivity-, and subsurface scattering maps became feasible



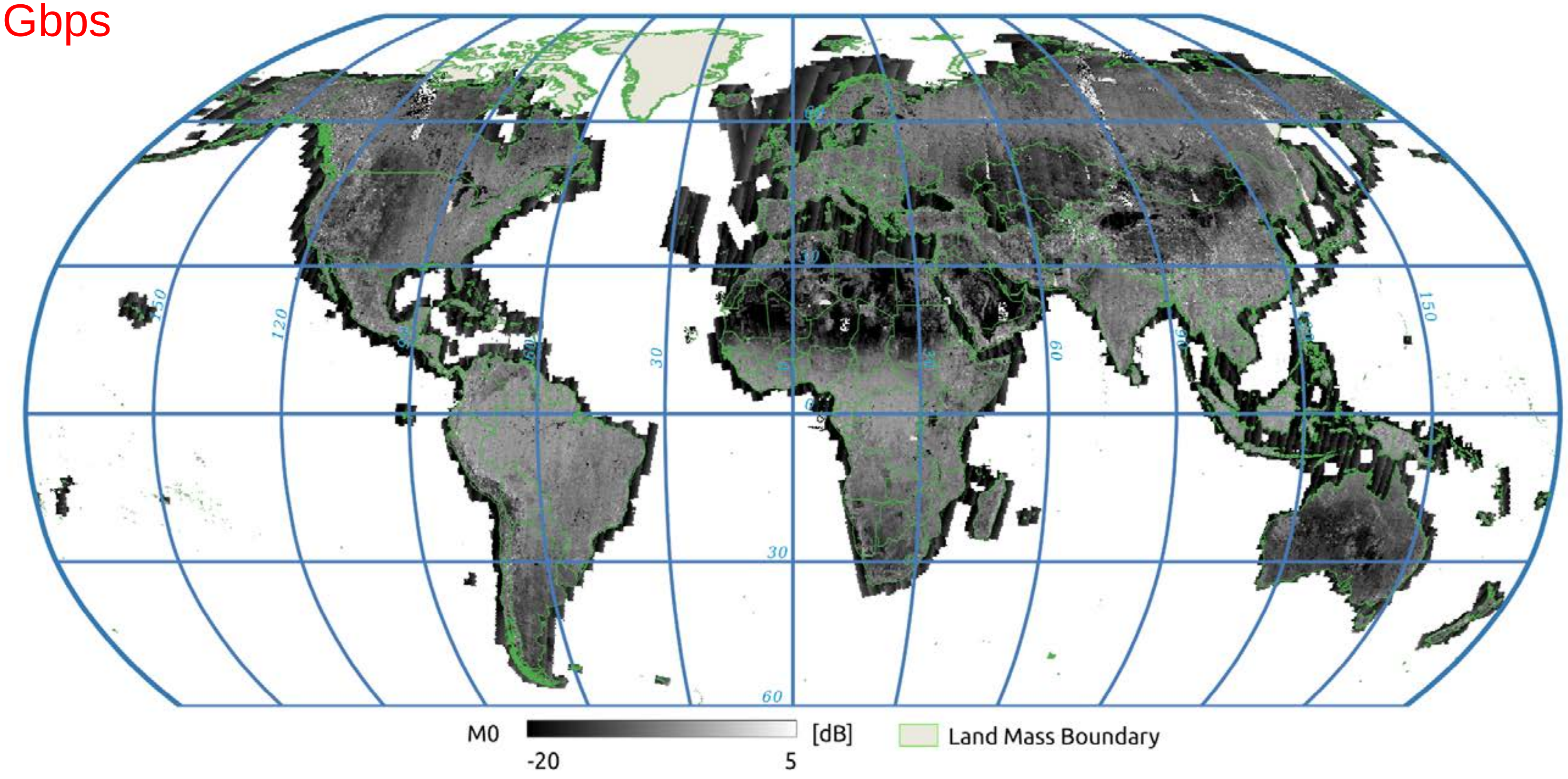


Wagner et al. (2024) Mapping subsurface scatterers from SAR backscatter time series, Proceedings EUSAR 2024, Munich, Germany, 23-26 April 2024, 93-97.

Raml et al. (2023) Enabling global scale Sentinel-1 time series analysis through streaming, Proceedings of the 2023 Conference on Big Data from Space (BiDS'23), 29-32. doi: 10.2760/46796.

Scaling by Hardware

- Improvements of global Sentinel-1 data analysis with optimized algorithms:
  - Throughput on VSC 3: ~0.23 Gbps
  - Throughput on VSC 5: ~1.75 Gbps
- Significant speed up of global parameter calculation, from ~7 days to < 1 day
  - Using ~30 – 50 VSC 5 nodes

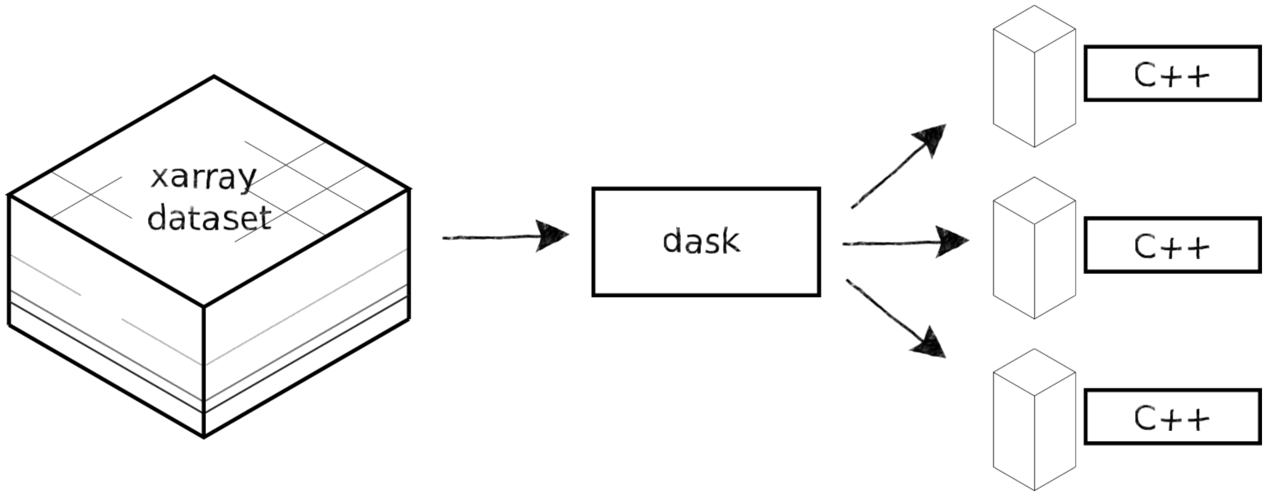


Harmonic model fitted to 3 years of Sentinel-1 data: Estimated mean backscatter

Tupas, M., et.al. (2024). Enabling Global Flood Monitoring with SAR Datacubes and the Vienna Scientific Cluster. Austrian-Slovenian HPC Meeting 2024 – ASHPC24, Grundlsee, Austria, 11-13 June 2024.

Scaling with Open-Source Frameworks

- We utilize open source frameworks such as dask and xarray to scale up on cloud platforms
- We can focus on the bottleneck, and optimize just that
  - i.e. reimplemeting the slow part in C++
- On a machine with 16 cores we achieve a throughput of: 1.3Gbps
- Enables calibration of more complex models like radiative transfer globally for operational use



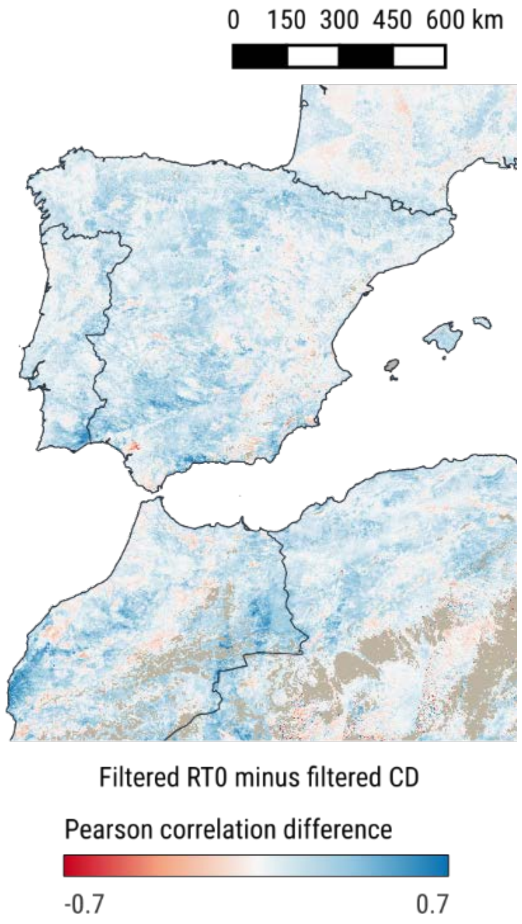


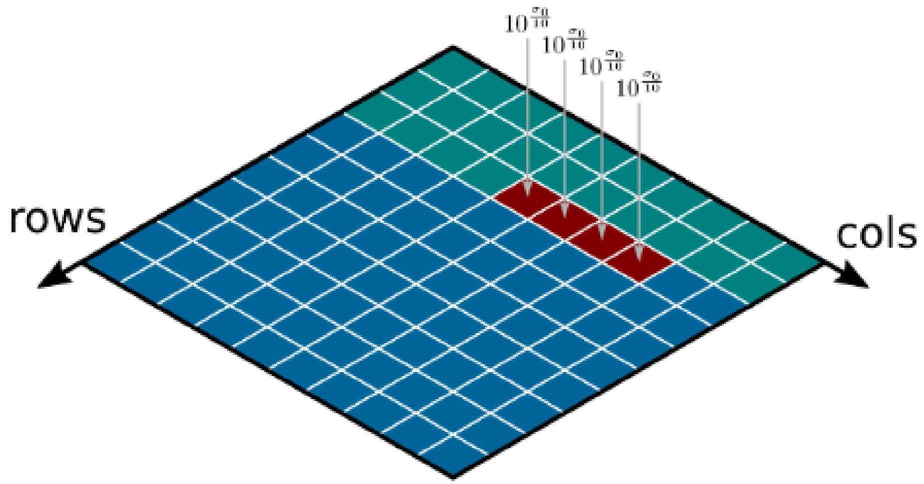
Figure from publication in preparation

Raml et al. (2023). Unleashing the power of Dask with a high-throughput Trust Region Reflectance solver for raster datacubes, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-7765

Next steps: Scaling with Specialized Hardware

- GPUs are up and coming, driven by AI research and applications
- Remote sensing can exploit this to run their own processing chains more effectively
  - i.e. preprocessing radiometric terrain corrected backscatter in seconds
- Investigated porting simple algorithms to GPU using Python and CUDA in G-Pet Use Case
- Initial test show speedups from 2x to 5x using trivial GPU implementations
  - Suggests potential for further performance increases in the order of magnitudes

Achieving significant speed ups for easily parallelizable operations on high-res raster data

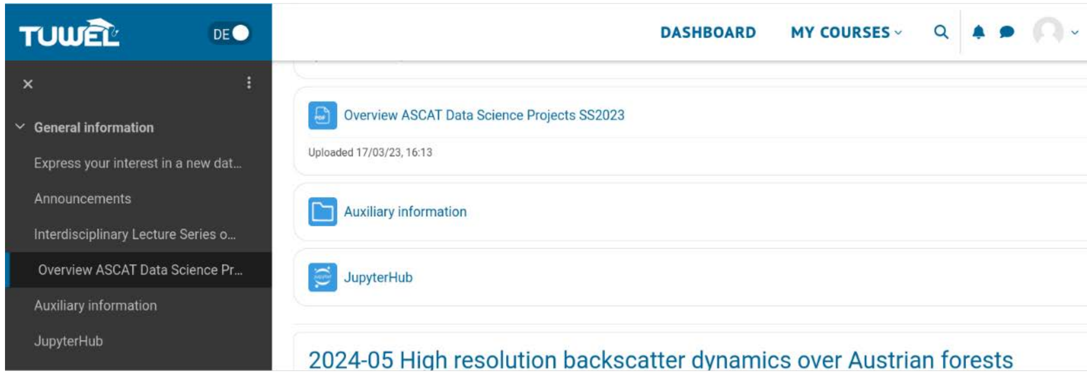


GPU - dB to linear

Teaching and Learning Earth Observation with Jupyter

Benefits for the student

- Engagement
- Participation (Hub)
- Understanding
- Preparedness
- Lower barriers



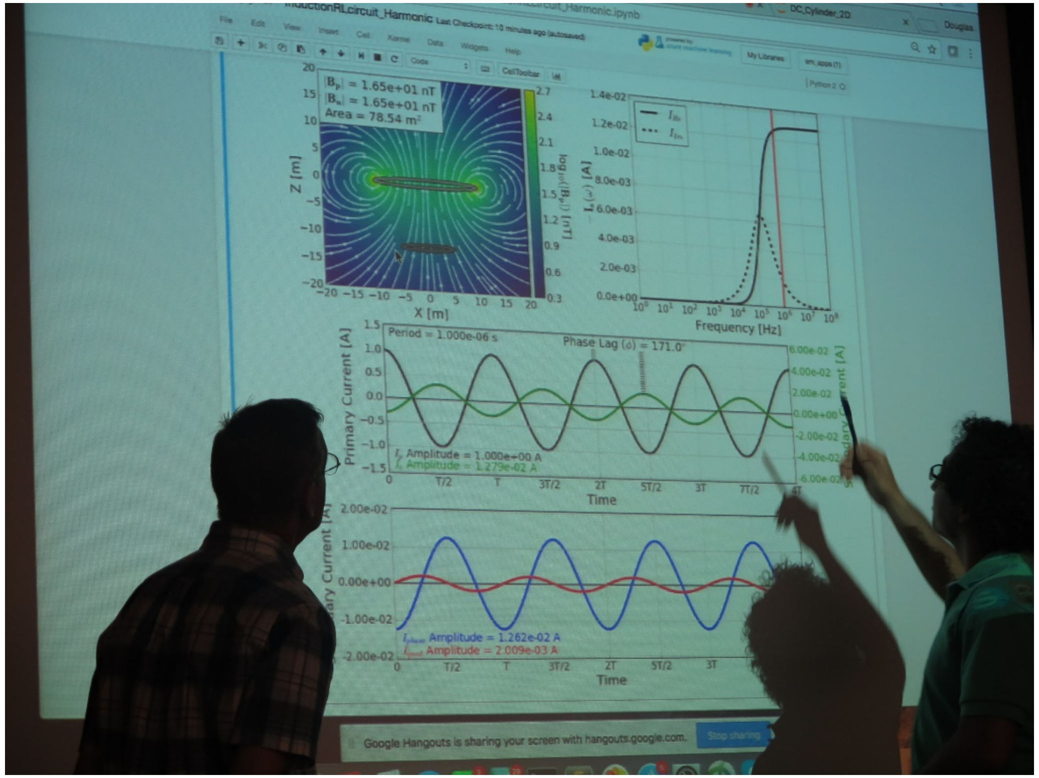
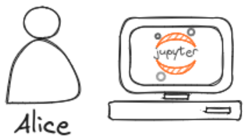


Photo credit: Seogi Kang

EODC JupyterHub



Limited data

Processing limited by student's hardware

"Unlimited" data but static

Processing not limited by student's hardware

Dynamic data created by external experts

Up-to-date real world examples

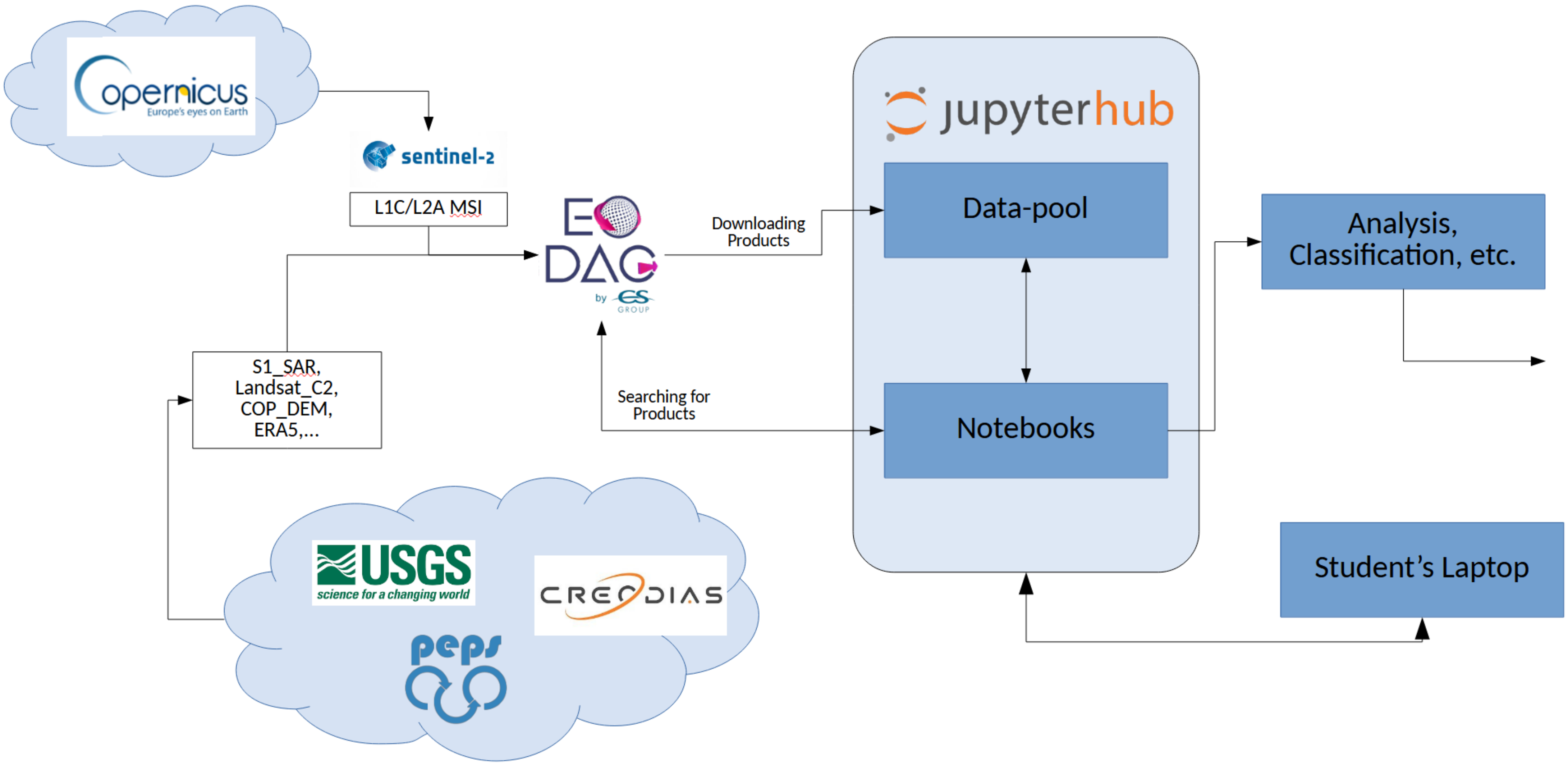
Processing specialised for big EO data

The World is your dataset

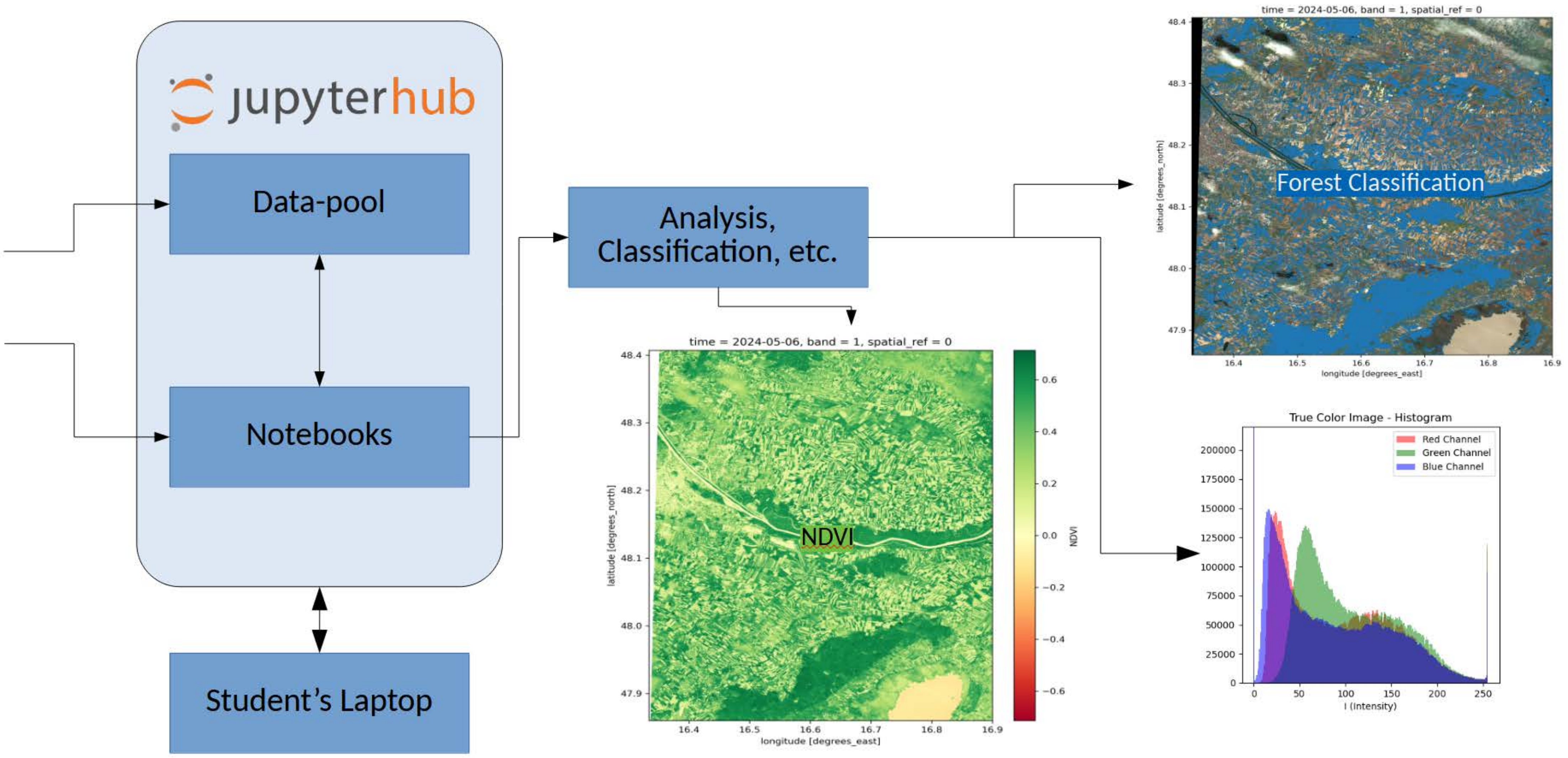
Computing close to the data

Increasing access to EO data

Notebooks for Student Projects and Thesis Work



Notebooks for Student Projects and Thesis Work cont.



Final Thoughts

- Only a small percentage of scientific studies exploit the potential of Big Earth Science Data
  - Cloud platforms have improved the situation, but most users are happy to just work with relatively small data sets

- More and more extremely powerful satellite missions are becoming available
  - Lack of resources for the proper care and exploitation of the data is a big concern

- The Web of FAIR data and services is becoming a reality
  - There is no need to just work on one monolithic Big Data platform

Get ready for it!

Acknowledgements

BMBWF: Cloud4GEO | Copernicus: Land Monitoring Service, Emergency Management Service | ESA: DTE Hydrology and 4DMED | FFG: DWC-Radar, S1Floods.AT, ScaleFloods | Vienna Business Agency: FAIR2Earth