ESA's 2025 Big Data Foundations for Earth Observation Training Course

Cloud-based EO processing - Flood Mapping with Dask

Thais Beham, Davide Festa (TU Wien)
Tobias Raiger-Stachl (eodc gmbh)
23. September 2025

European Commission

researchLatvia
Value Through Knowledge

Ministry of
Education and Science
Republic of Latvia

esa academy

Agenda

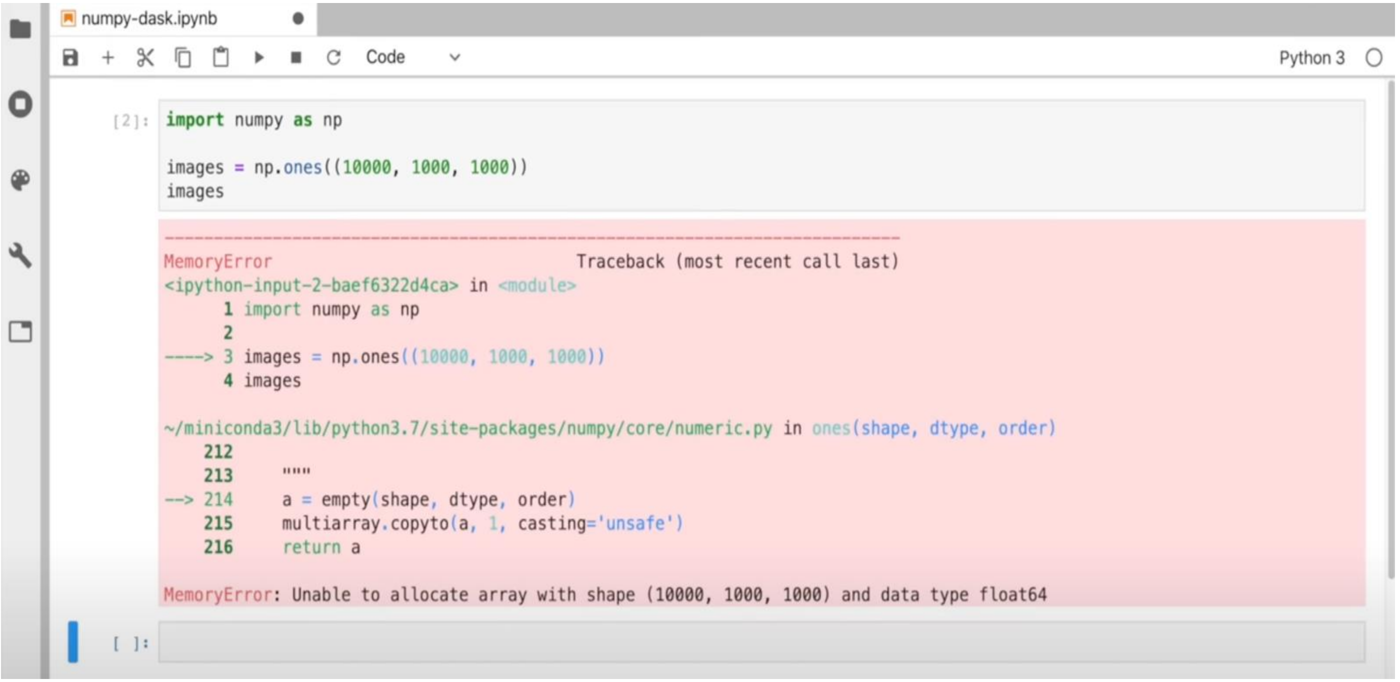
- **Dask**
- **Cloud Computing at eodc**
- **STAC**
- **Sentinel-1 SAR & Water Surfaces**
- **Sentinel-1 Datacube at TU Wien**
  - TU Wien flood maSentinel-1 Procssing
  - Datacube Architecture
- **TU Wien flood mapping algorithm**
  - Modelled parameters from datacube
  - Bayesian classifier
  - Examples

Dask

Dask

- Numpy, Pandas, Xarray are powerful for data science & analytics, but struggle with Big Data without specialised frameworks

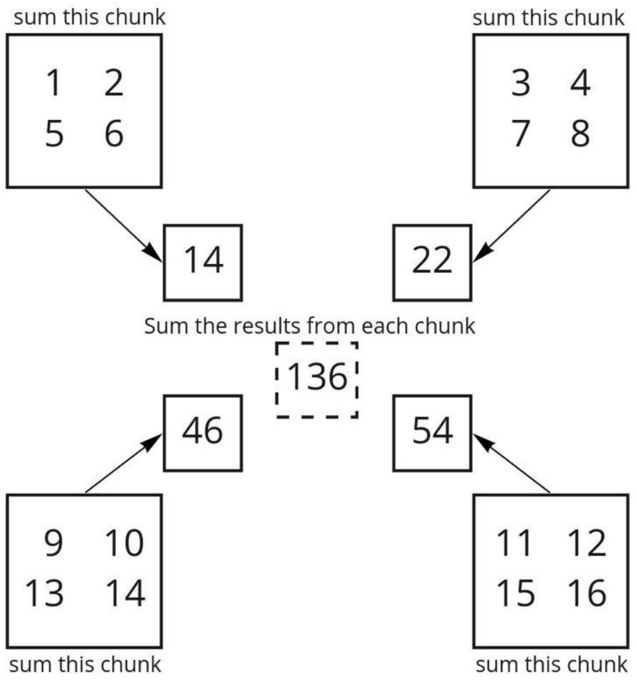
- Working with EO data often triggers MemoryError due to dataset size.



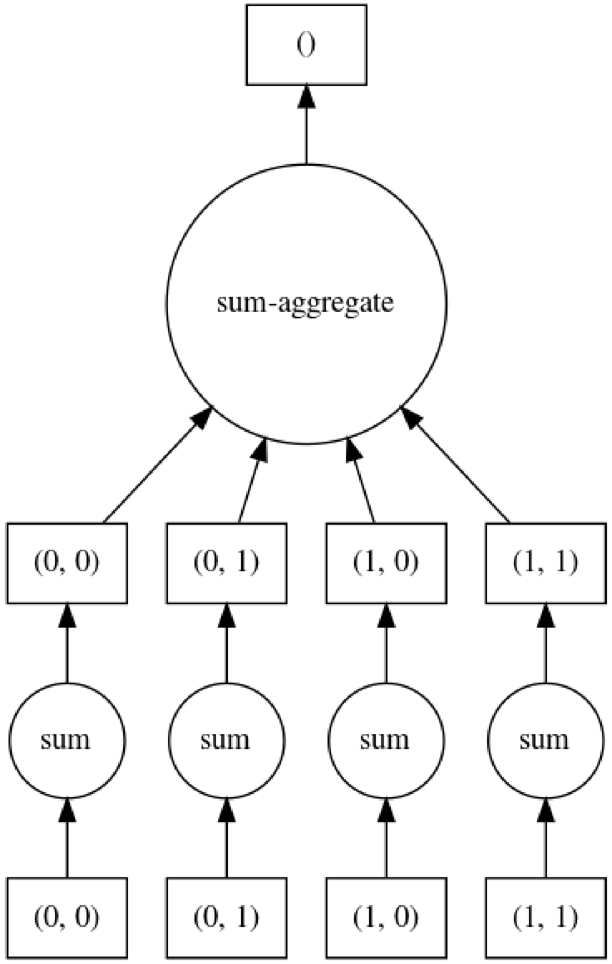
Dask - Chunking

- Break the data into chunks
- Tasks are organised in the order they are performed
- Distribute the computation into several nodes, in case of a many-node cluster.

dask



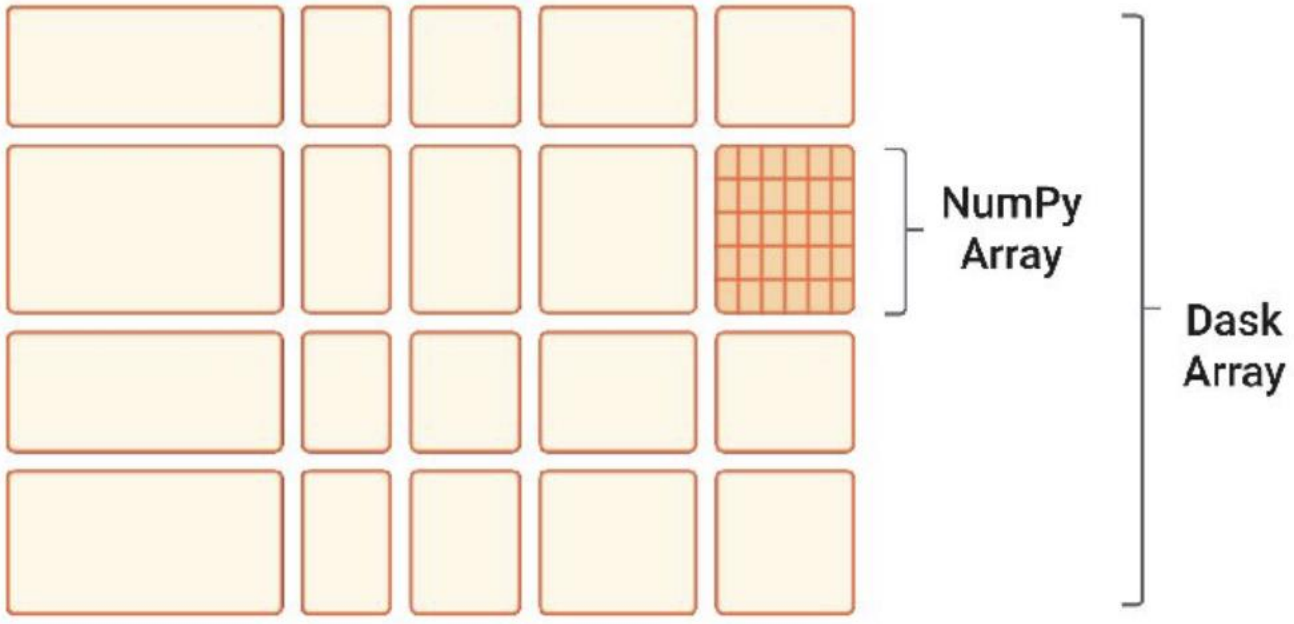
SOURCE: HTTPS://SATURNCLOUD.IO/BLOG/WHAT-IS-DASK/



SOURCE: HTTPS://SATURNCLOUD.IO/BLOG/WHAT-IS-DASK/

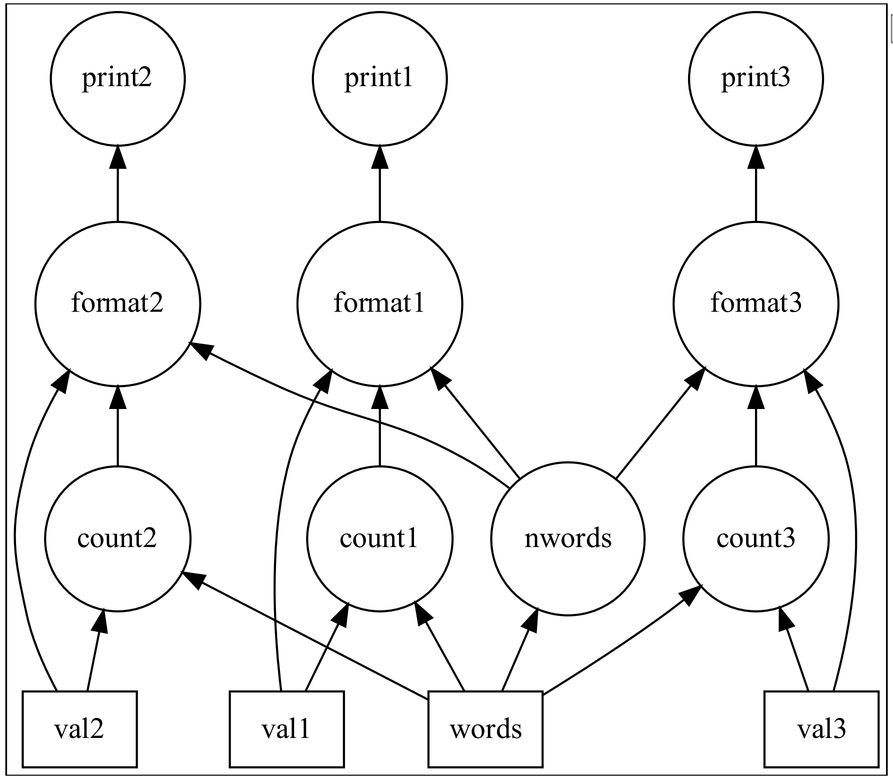
Dask - Chunking

- Data chunks
- Each task runs independently, and intermediate results are only communicated when necessary.

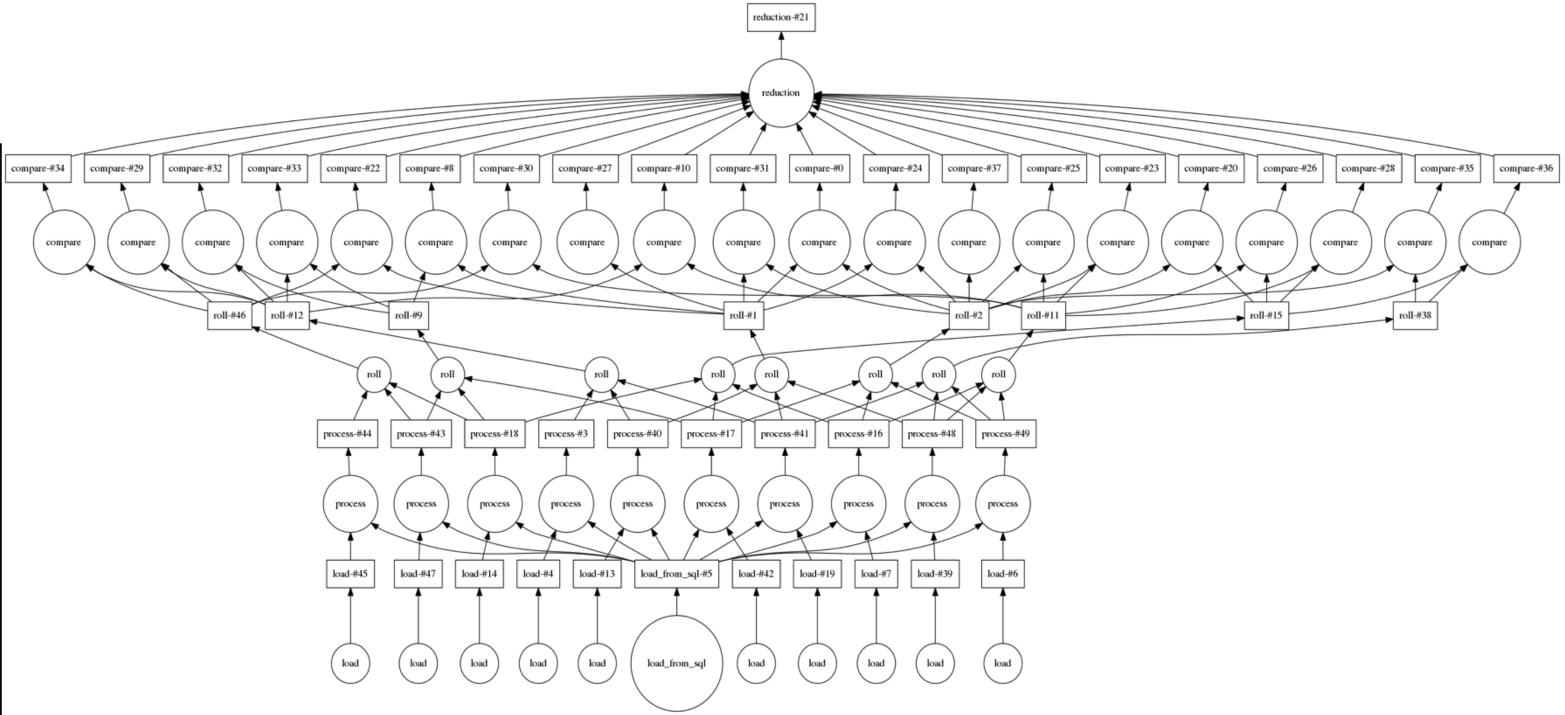


HTTPS://PYTHON.PLAINENGLISH.IO/DASK-PARALLELIZE-PYTHON-0CB653BF9A02

Dask – Task Graph



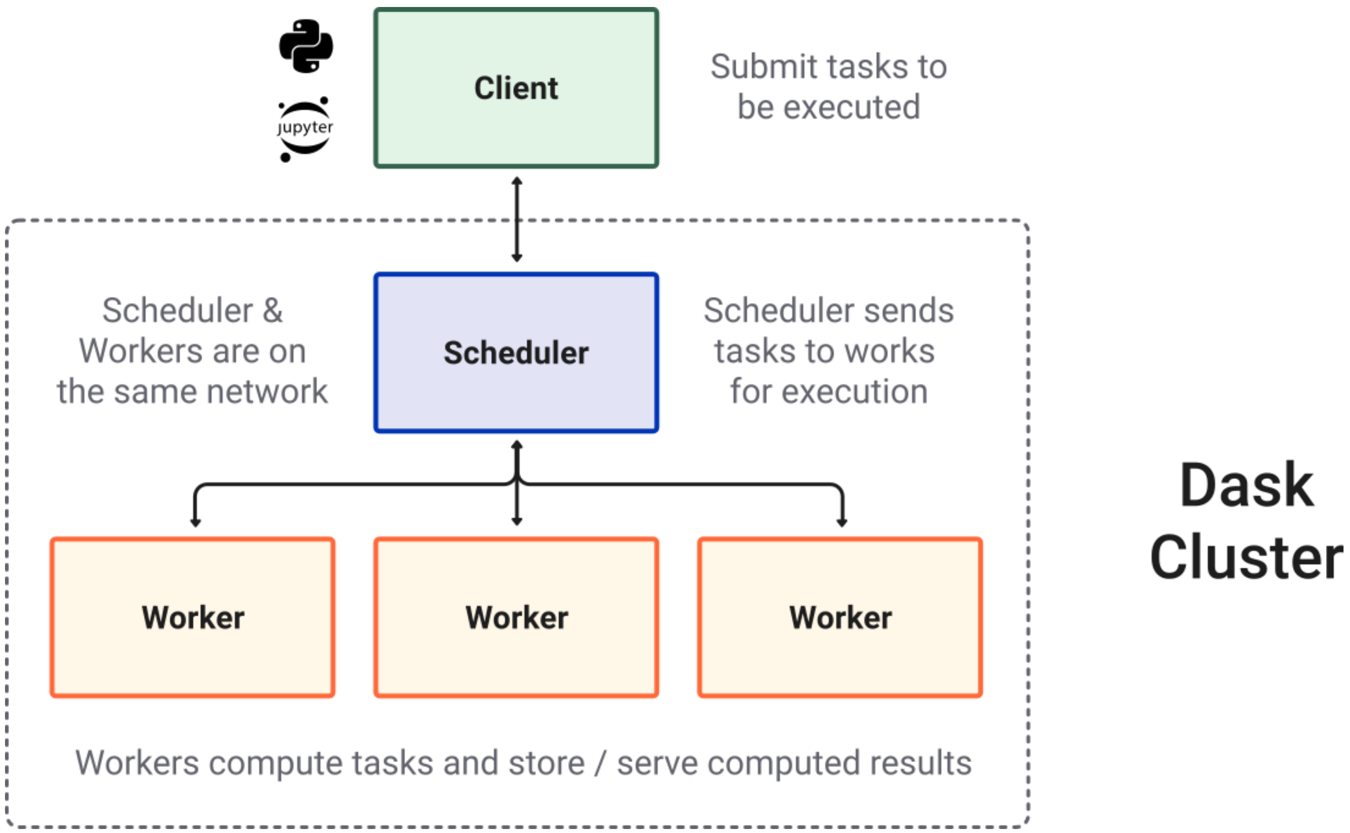
SOURCE: DASK DOCUMENTATION



SOURCE: HTTPS://OZDEMIRHASAN.MEDIUM.COM/PARALLEL-PROCESSING-WITH-DASK-4647A58BD2A5

Dask Scheduler

- Distributed computation
- Fault tolerance
- Adaptative scaling



HTTPS://PYTHON.PLAINENGLISH.IO/DASK-PARALLELIZE-PYTHON-0CB653BF9A02

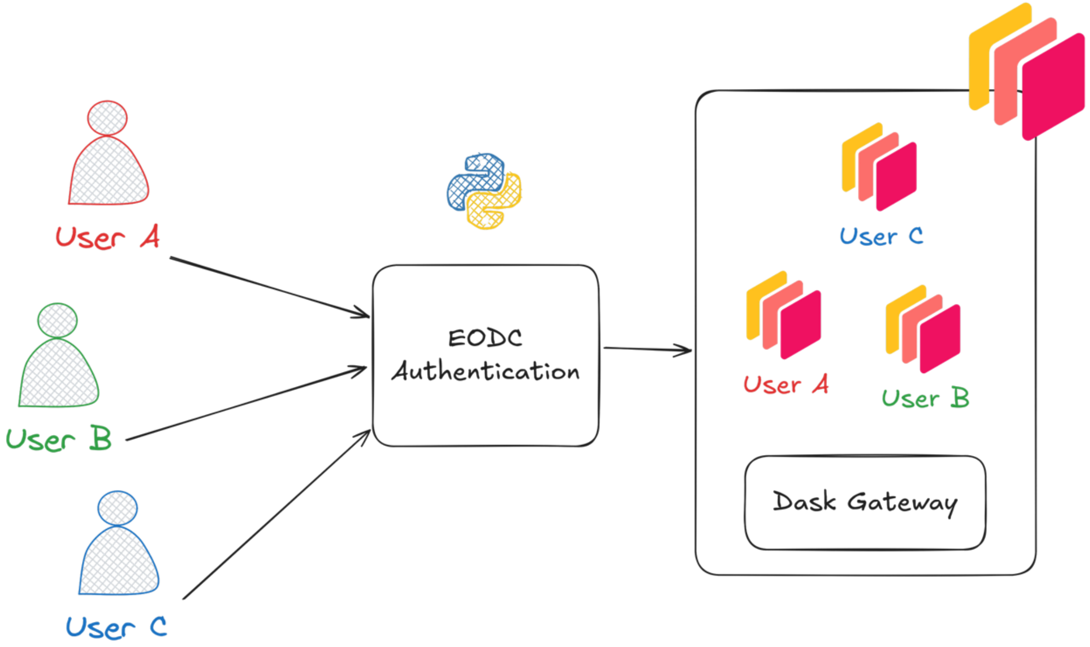
Dask

- Easy to install with pip or conda
- Secure
- Compatible with major cloud computing platforms:
  - AWS, Azure, Google Cloud Platform…
- Dashboard

Cloud-computing at eodc

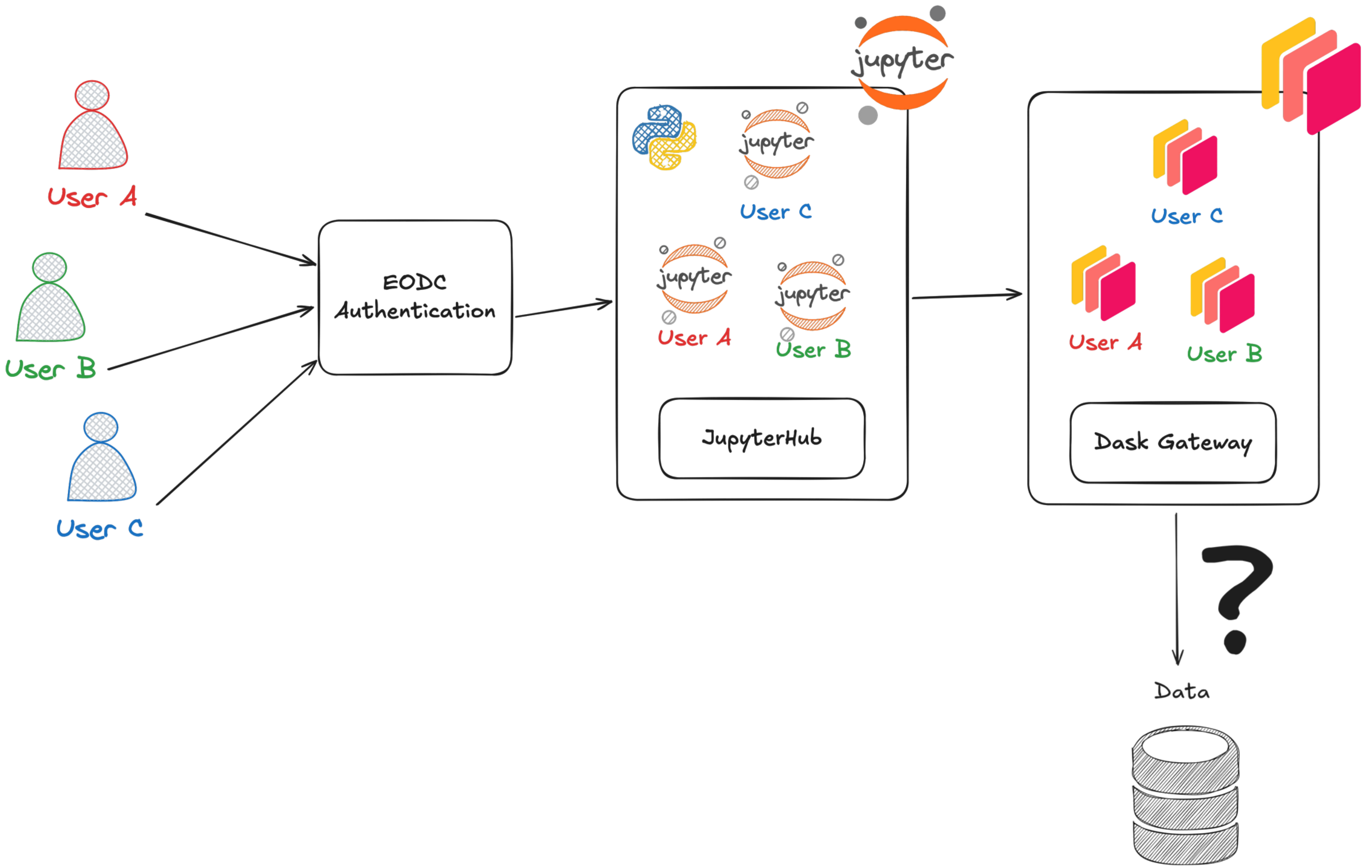
eodc Dask Gateway

- eodc offers Dask as a service
- Users can launch a Dask cluster without requiring to have direct access to any cloud infrastructure
- Lower entrance barrier for users
- Run large scale data analysis on demand in a scalable environment



eodc JupyterHub

- Provide users access to pre-configured environments
- Reduce burden with installation and maintenance tasks



STAC

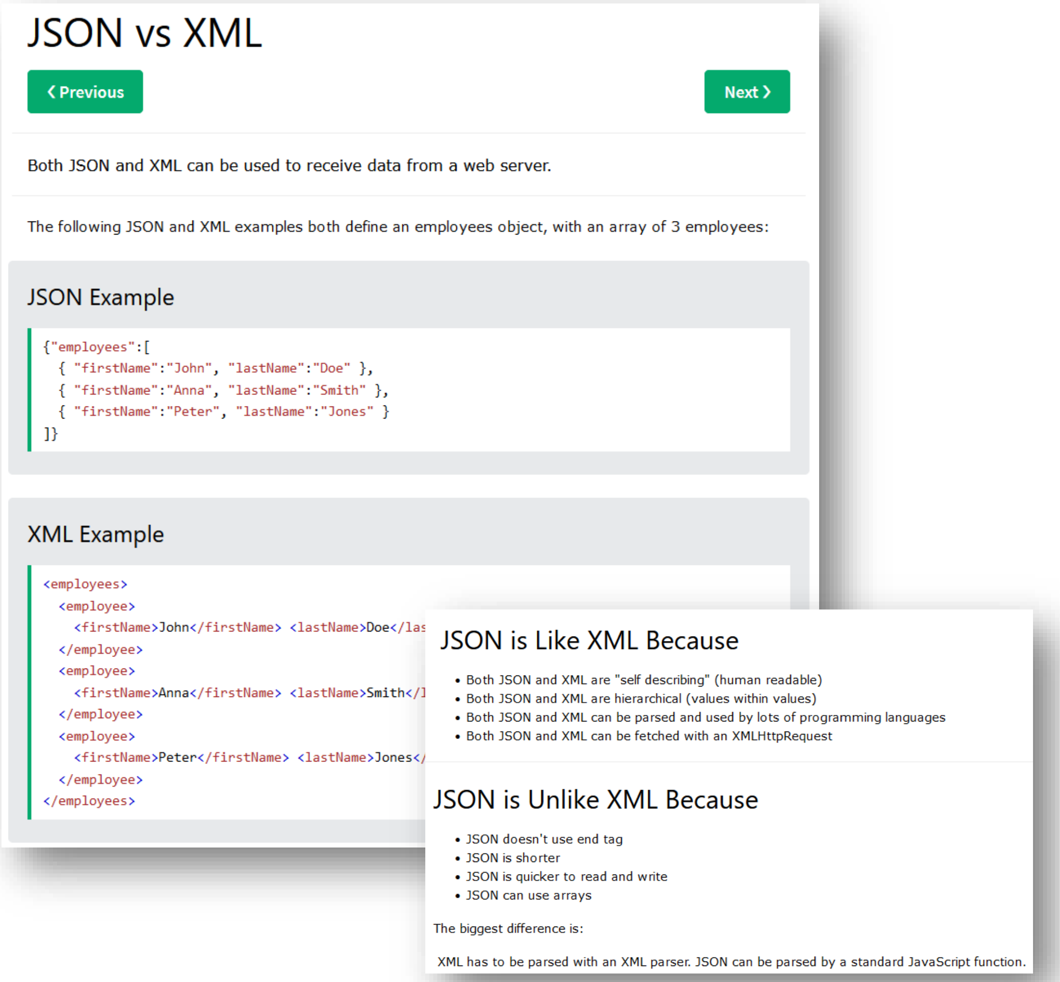
STAC Motivation

- **Spatio Temporal Asset Catalogs**

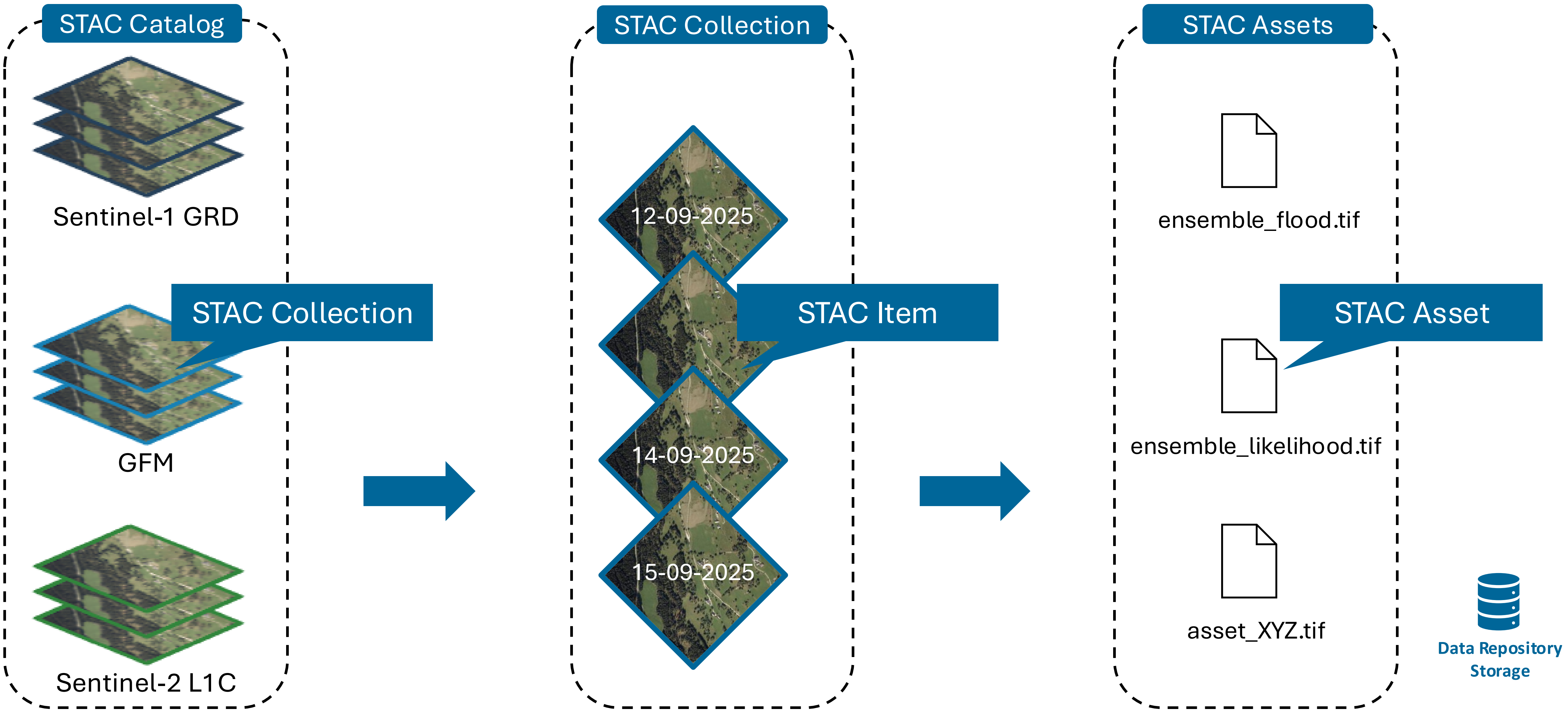
- **Simplicity:**
  STAC uses JSON for straightforward metadata management.

- **Web Compatibility:**
  By adopting JSON, STAC aligns with web standards, promoting interoperability.

- **Client Support:**
  Supported by a variety of client libraries.

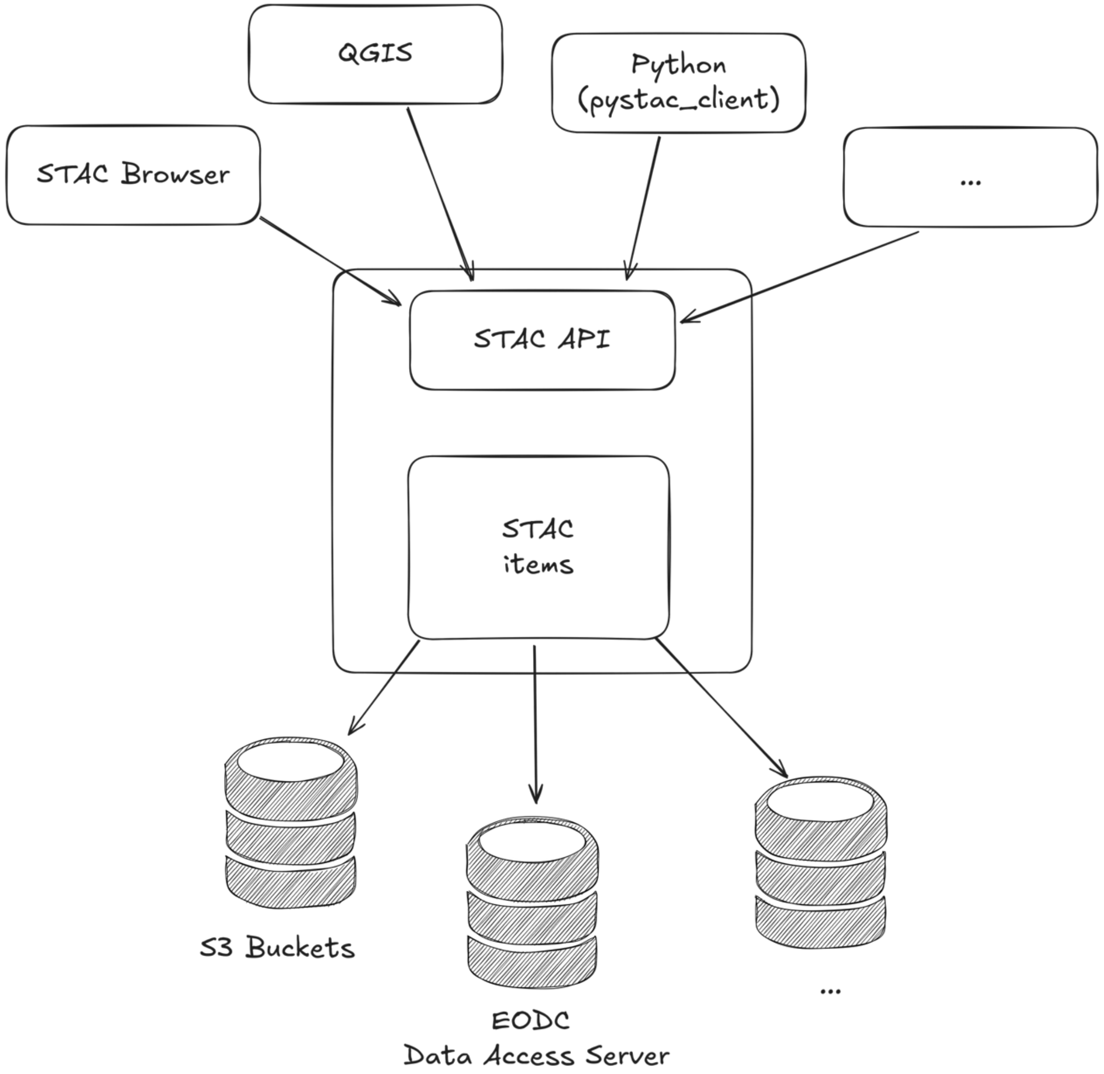


STAC entities

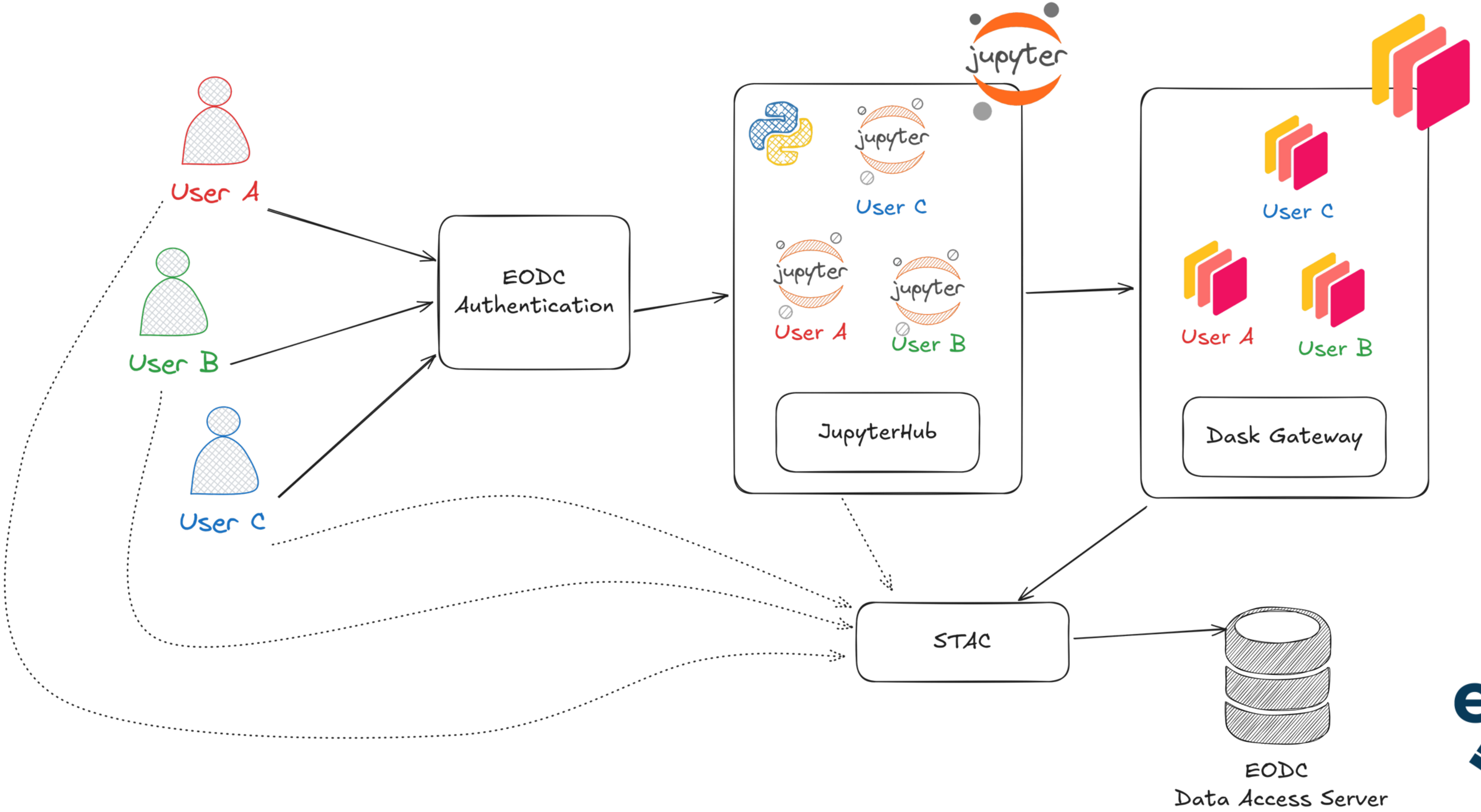


STAC != Data Access

- STAC makes data queryable and observable
- STAC **is not** a data access method
- STAC provides us with links to assets which can live wherever



Cloud-Computing at eodc



Do you want to see it in action?

- Join the hands-on workshop in the afternoon hosted by Thais and Davide

- eodc webinars online on YouTube
  - eodc Webinar: How to make use of eodc services for data discovery and access
  - eodc Webinar: How to utilise eodc data processing services

- Check out the eodc Knowledgebase

Sentinel-1 SAR & Water Surfaces

Sentinel-1 – a game changer for radar data

A two-satellite constellation

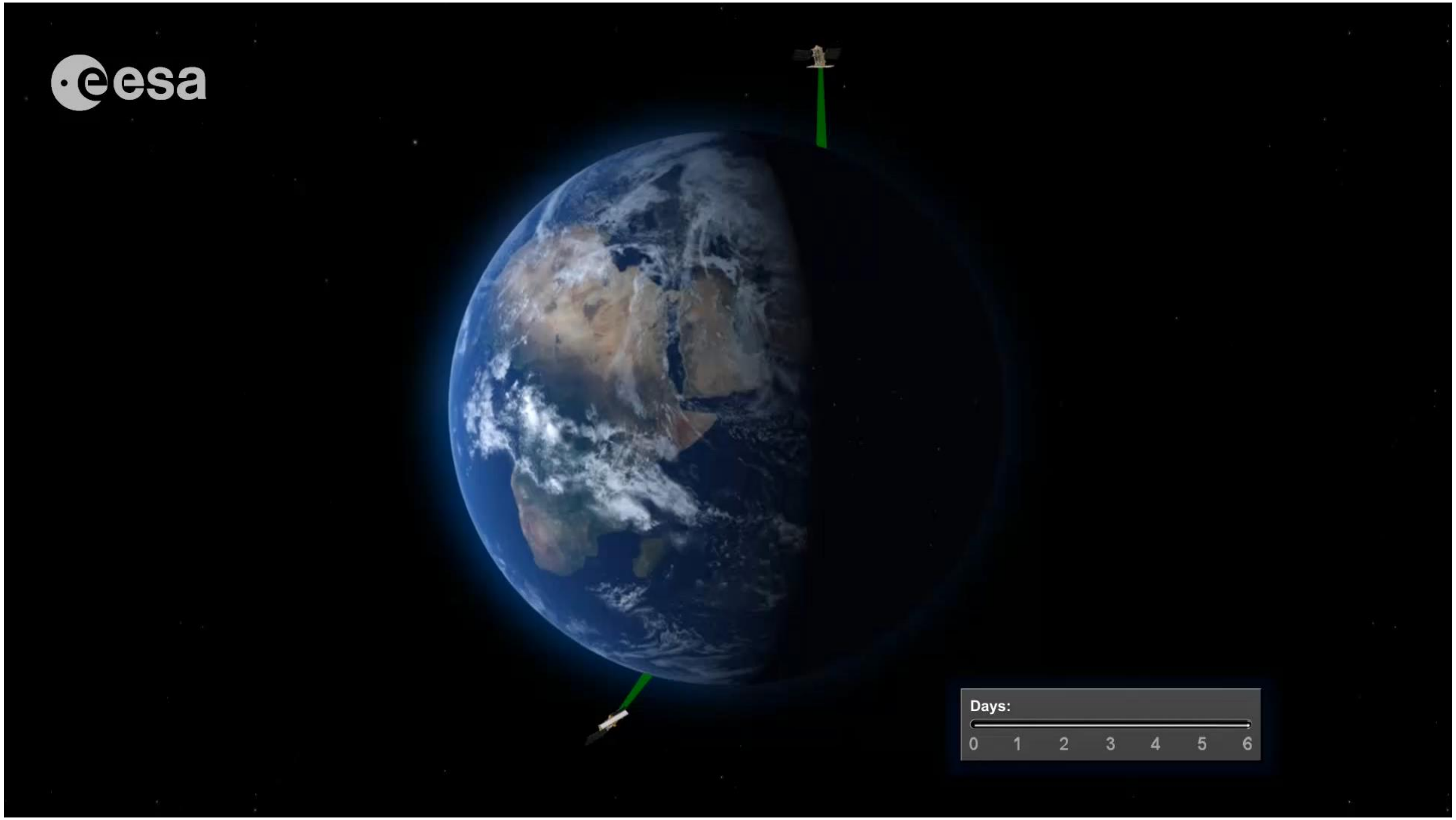
- Sentinel-1A: 2014 – present
- Sentinel-1B: 2016 – 2021
- Sentinel-1C: 2024 – present
- Sentinel-1D: 2025?

Fixed orbits

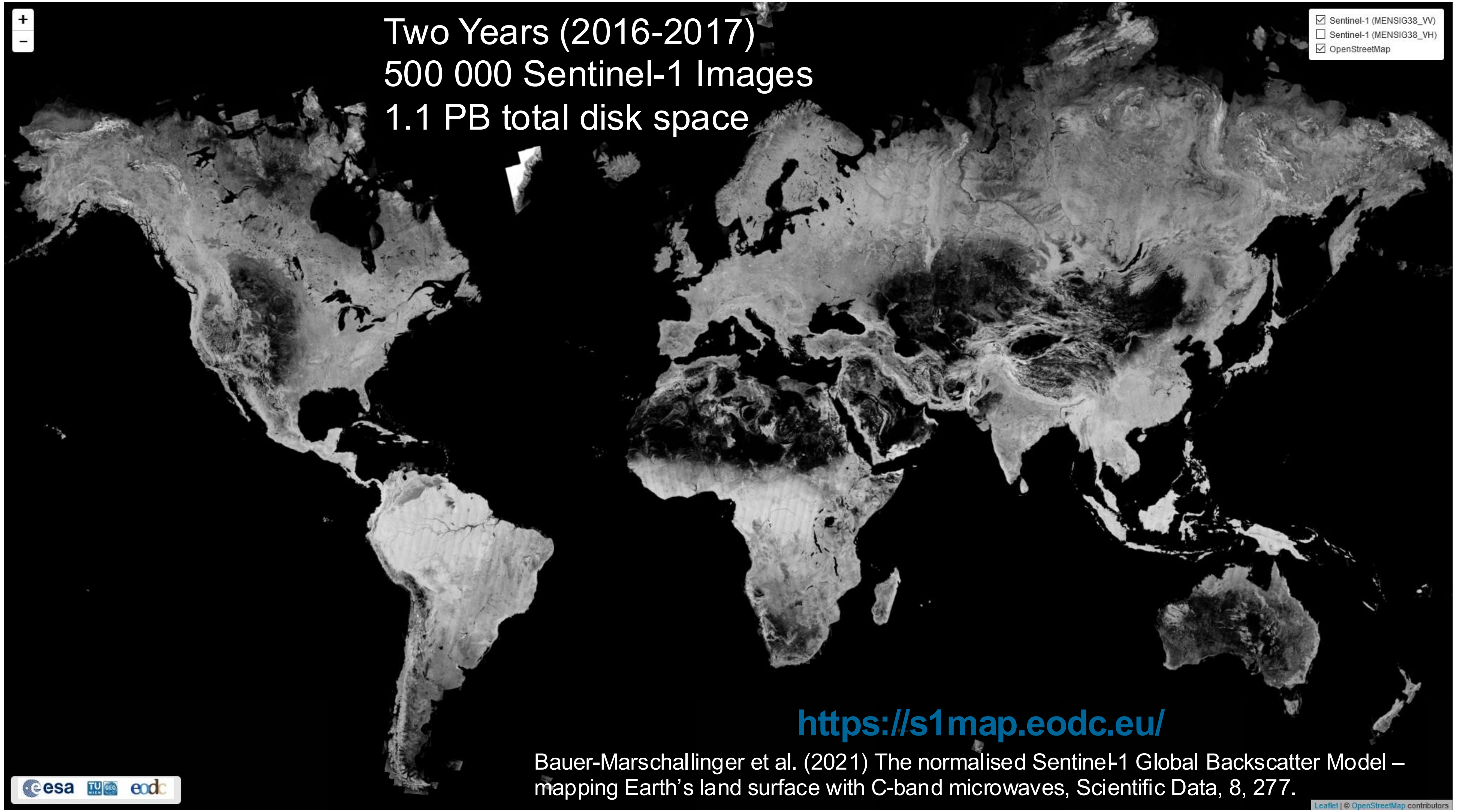
- 180° separation
- 12-day repeat cycle
- 175 locally distinct relative orbits

Applications

- Interferometry
- Soil moisture
- Water Surfaces
- Vegetation (density)
- Land Cover
- ...



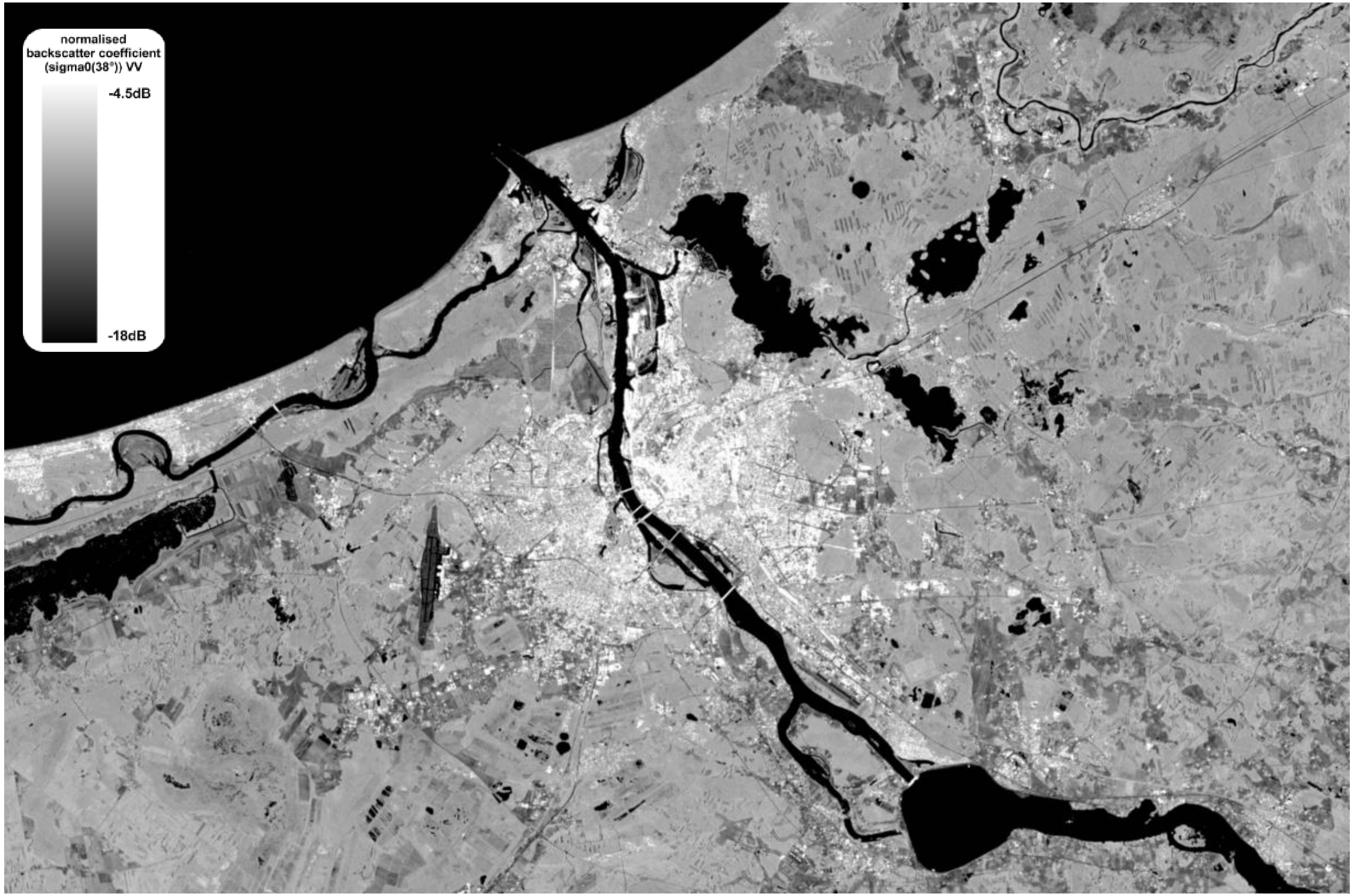
Mapping Earth´s land surface with Sentinel-1



The radar point of view

Sentinel-1 SAR

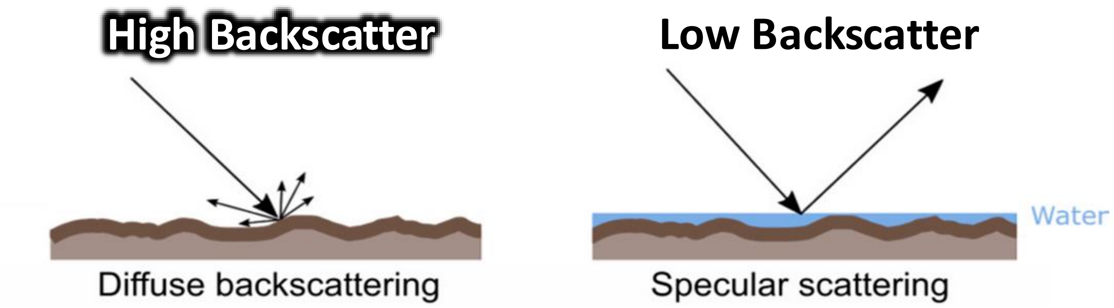
- C-band system (5.4 GHz ~ 5 cm)
- High sensitivity to the electrical properties of water
- Strong contrast compared to dry surfaces
- Operates over land in *Interferometric Wide Swath (IW) Mode*
  - 5 m x 20 m resolution
  - Europe: 1.5 - 4.5 day acquisition frequency
  - Wolrdwide: 3 – 12 day acquisition frequency



Flood mapping with SAR

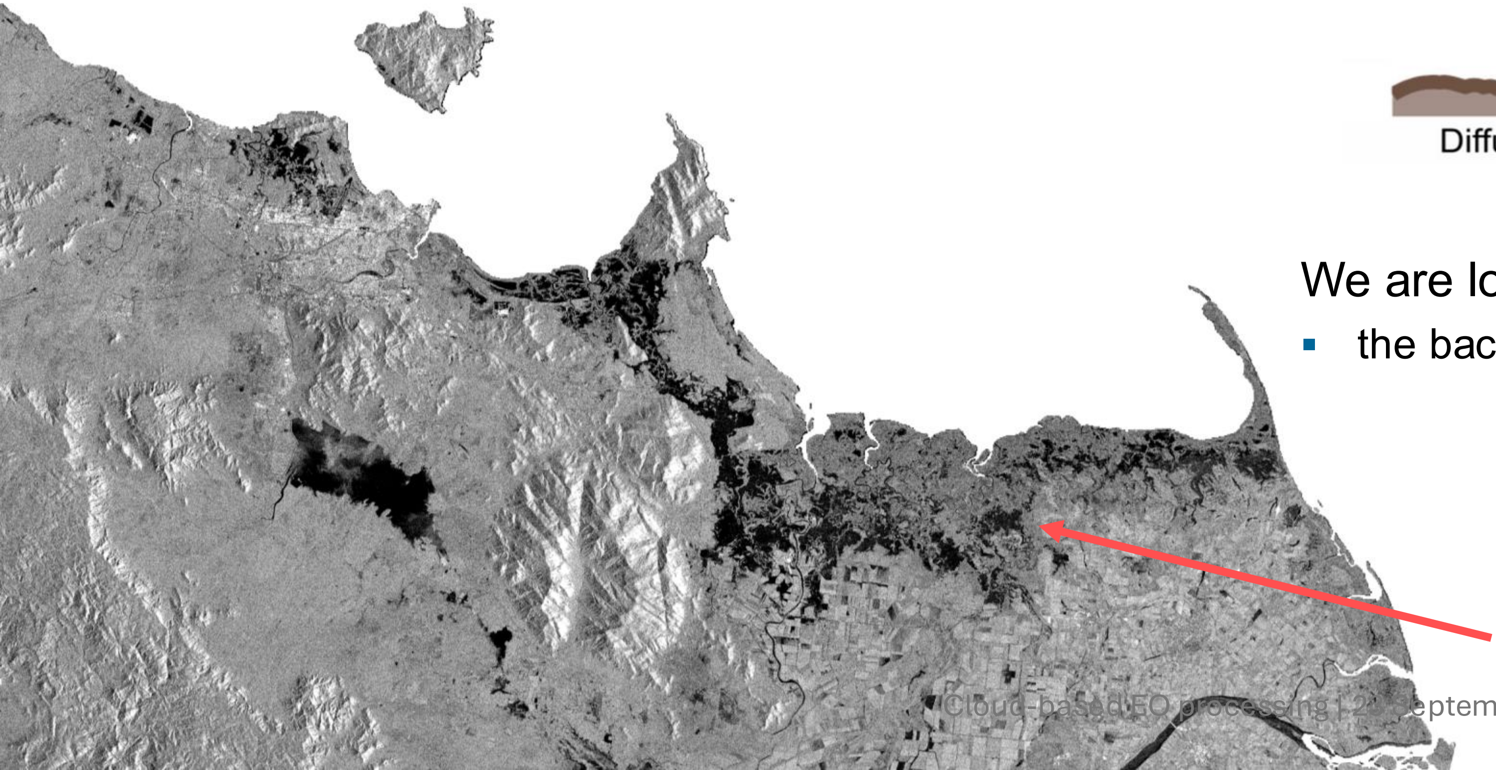
- Sentinel-1 is an **active** microwave sensor
- Operating Day and night
- All weather conditions
- Effective in discriminating water on ground

Figure modified from Ottinger and Kuenzer (2020)



We are looking for significant reductions in the SAR signal

- the backscatter coefficient decreases by approximately -18dB.

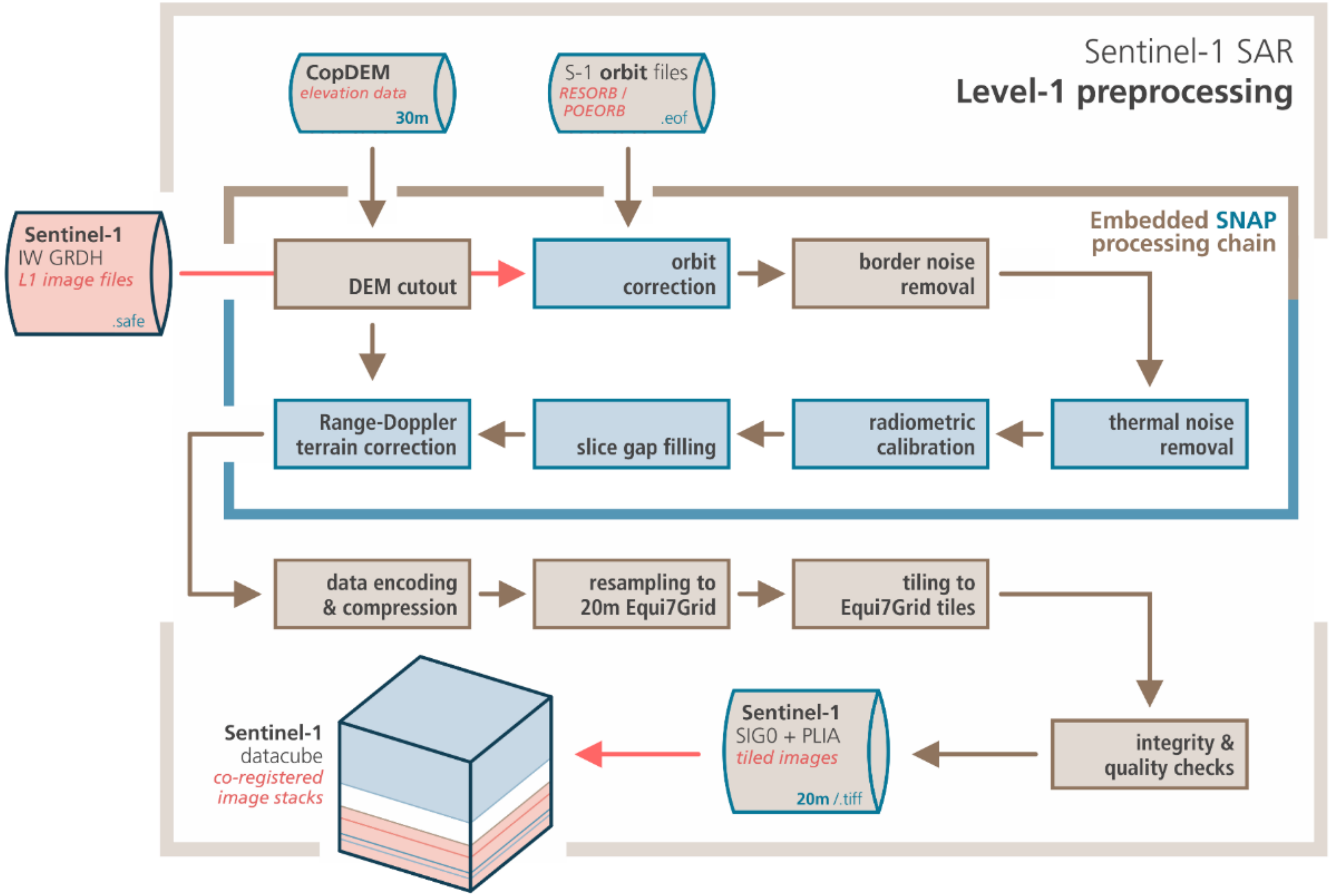


Severe flooding in Queensland, Australia observed by Sentinel-1 on 30 January 2019

Sentinel-1 Datacube at TU Wien

Sentinel-1 Processing

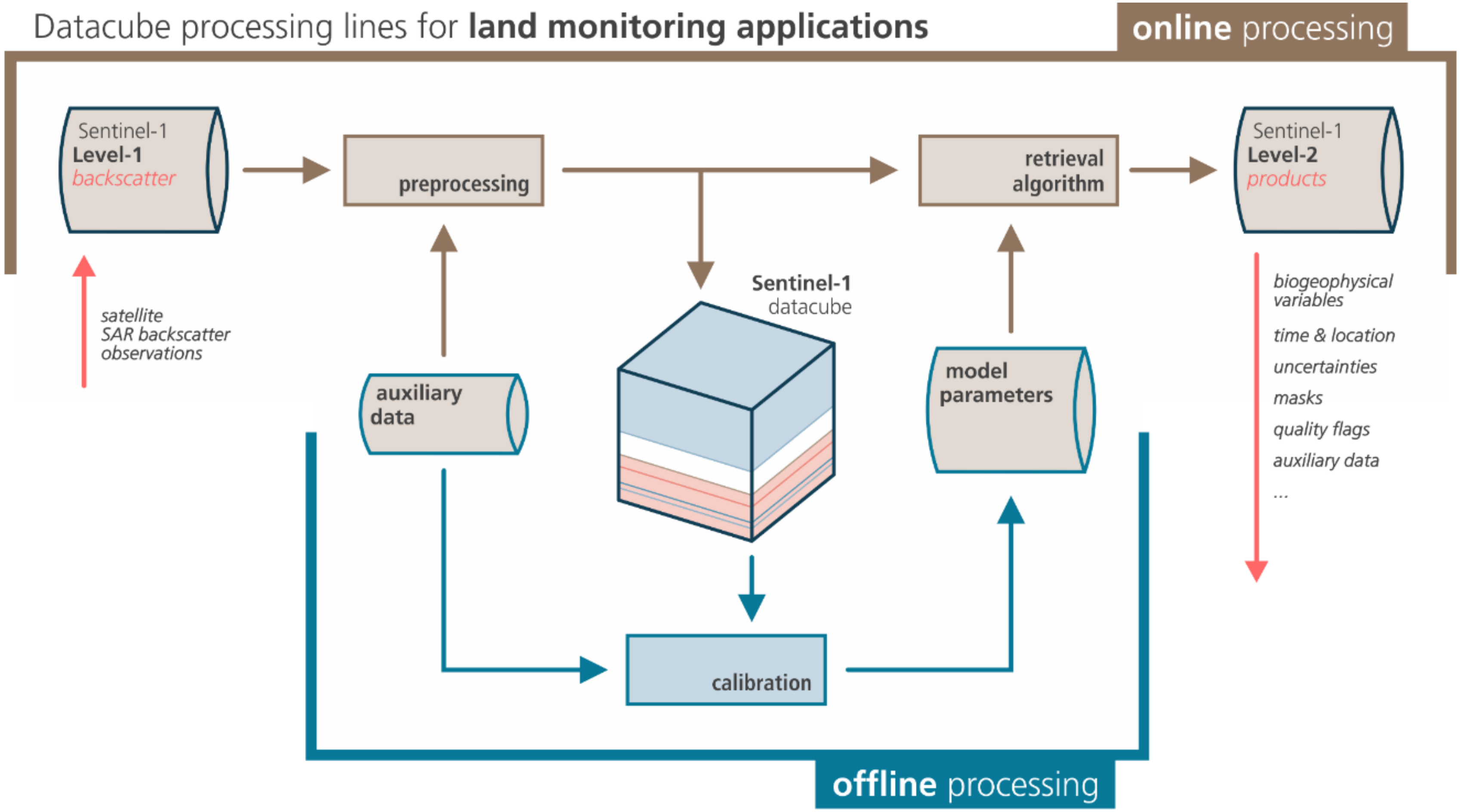
- High computing power required
- Level-1 file collection from eodc and 30m Copernicus DEM + EGM 2008
- Pre-processing using a GPT-graph SNAP within Python wrapper:
  - Apply external orbit data
  - Border noise removal (Ali et. al, 2018)
    - for data until ~2018
  - Thermal noise removal
  - Radiometric calibration
  - Slice gap filling
  - Range-Doppler terrain correction
- Output (registered as a STAC item):
  - Cloud-optimized GeoTIFF containing encoded dB values (Int16, scale factor=0.1)
  - Equi7Grid T3-tiles at 20m sampling



Datacube: Processing & Model parameters

Based on scientific findings

- Algorithms for...
  - SAR processing
  - Variable estimation
  - Water surfaces
  - Soil moisture
  - Vegetation density
  - and more
- Operational services
  - global
  - in near real-time (NRT)
  - freely accessible

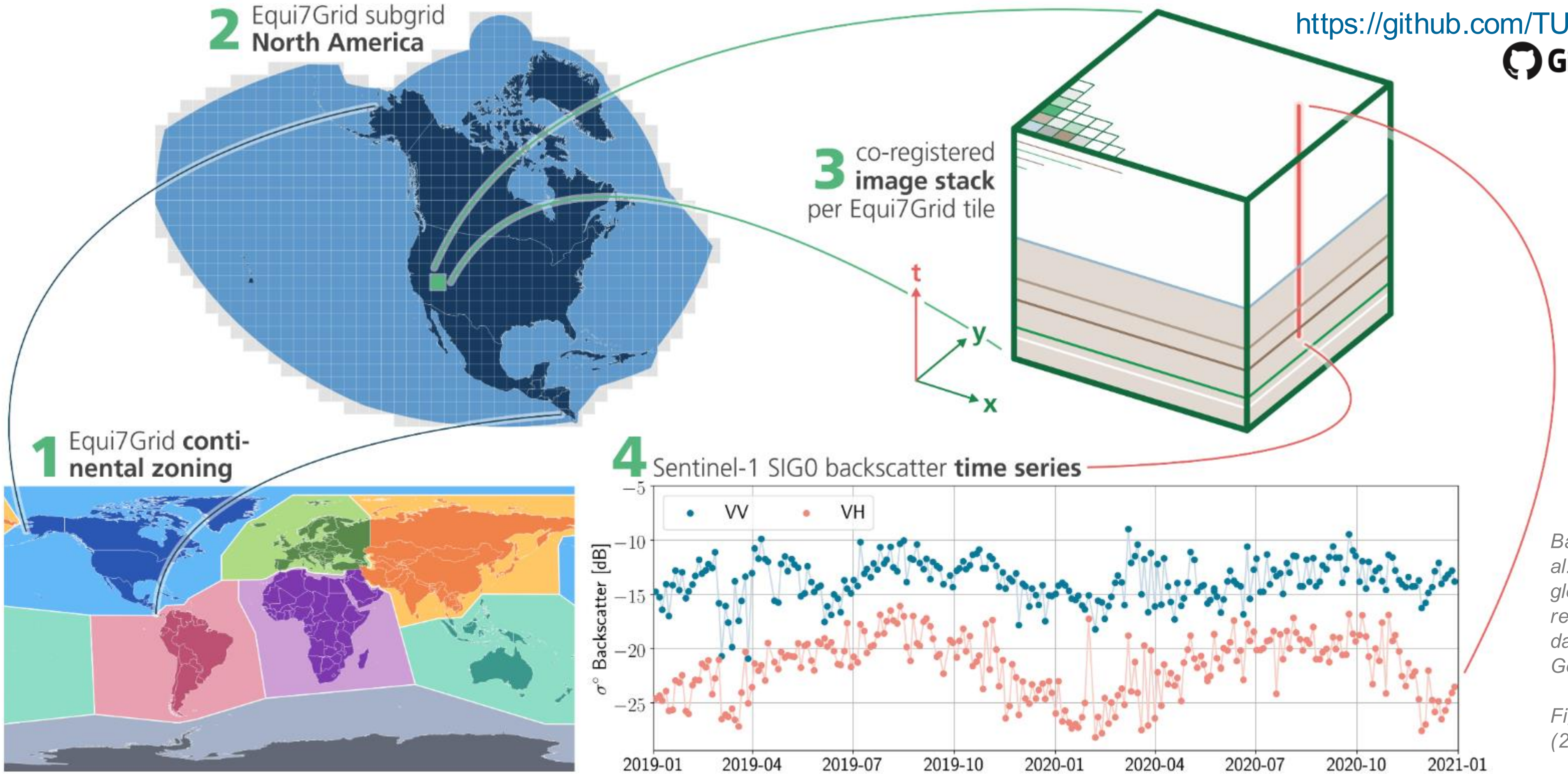


Datacube architecture: Equi7Grid & Time Series

Sentinel-1 ARD datacube: Concept of **Equi7Grid data structure** & **time series access** | Example for T3-tile over the USA

https://github.com/TUW-GEO/Equi7Grid

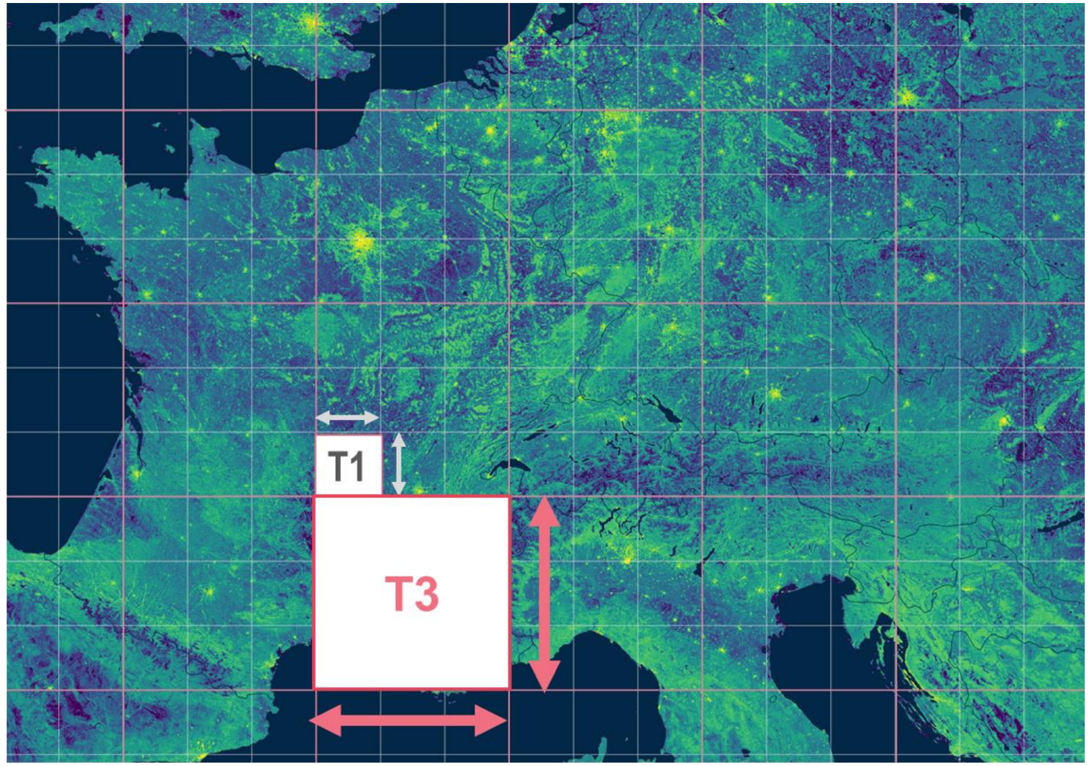
GitHub



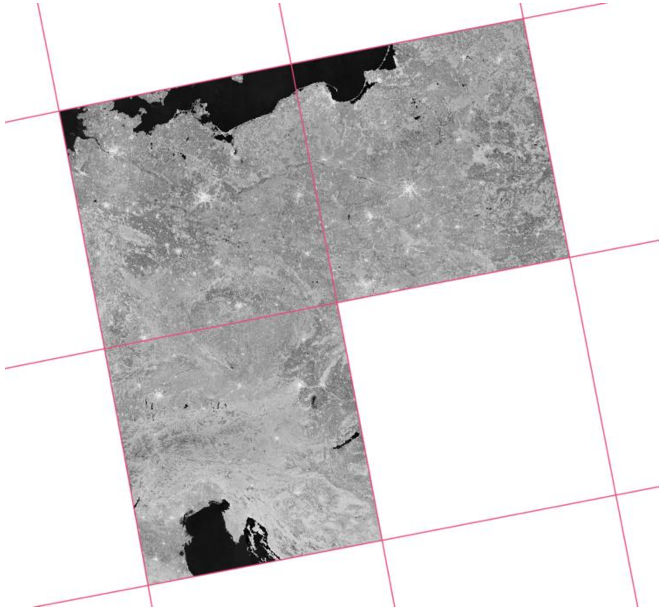
Bauer-Marschallinger et al. (2014) Optimisation of global grids for high-resolution remote sensing data, Computers & Geosciences, 72, 84-93.

Figure in Wagner et al. (2021).

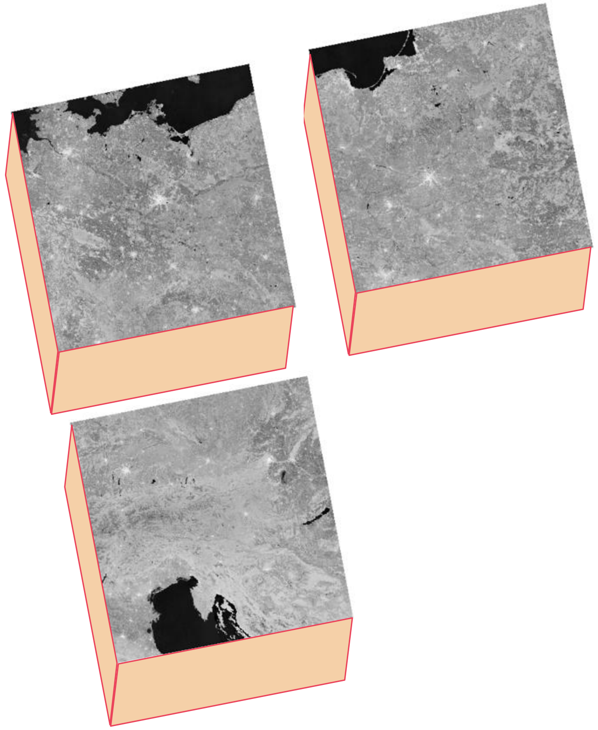
Equi7Grid: Tile Structure



geographic logic



file-system logic



- One folder per tile
  - datacube is populated with tiled & formated files (mainly GeoTIFFs)
    - with attached metadata

- Georeferenced image stacks per tile
  - time via time-stamp in filename
  - Easy to search/query, sort/filter, display/plot, read/write

TU Wien Flood Mapping Algorithm

TU Wien Flood mapping algorithm I

- Binary classification into flood/no-flood
  - based on Bayes decision from a priori probabilites
  - likelihood values associated to decision
  - *used for NRT operations in CEMS GFM*

- Parameters from datacube
  - Flood model: linear model from water body samples
  - No-flood model: harmonic time series analysis → seasonality (DOY)
  - Masking for no/low SAR sensitivity and challenging topography

Flood mapping algorithm based on parameters from Sentinel-1 datacube

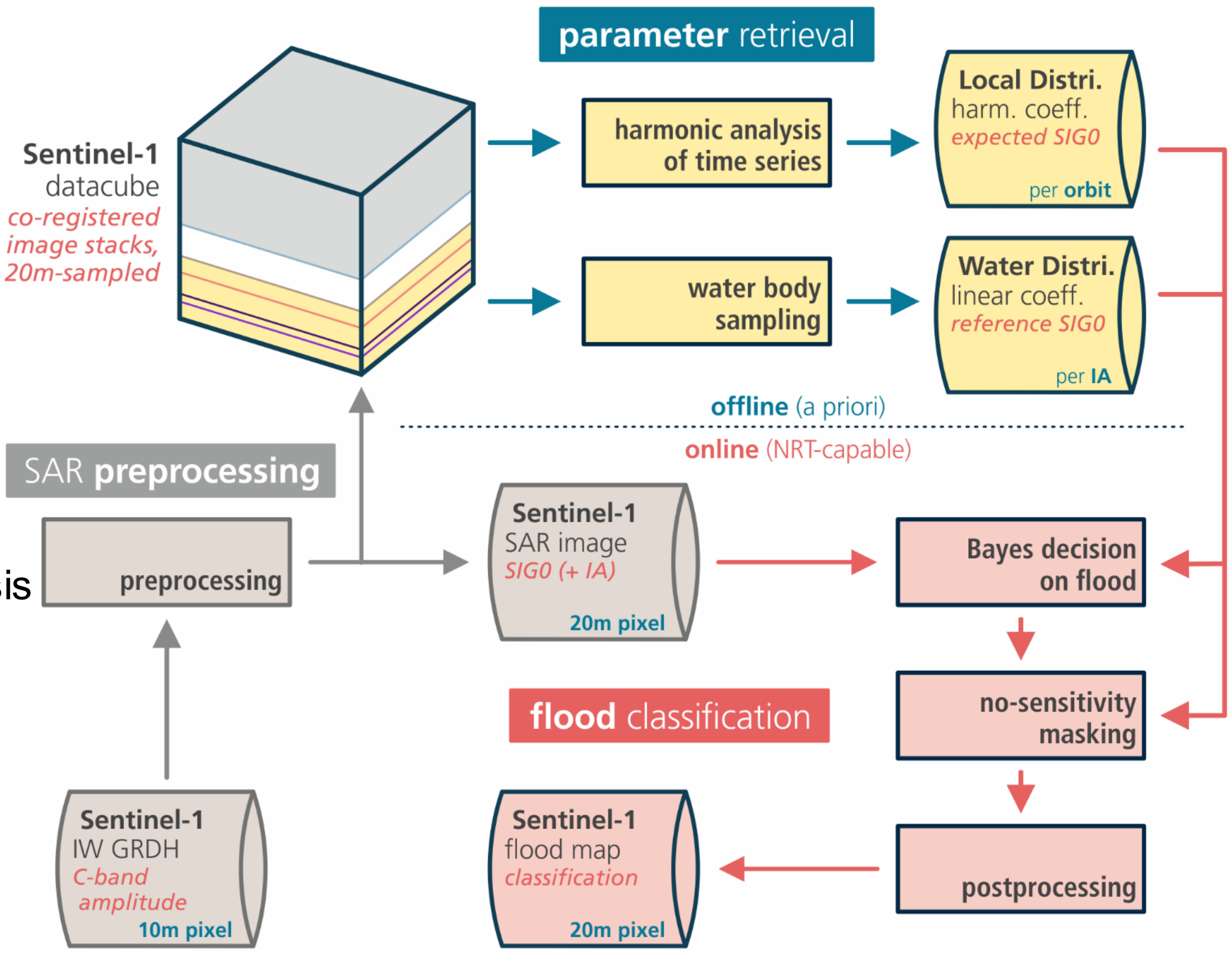
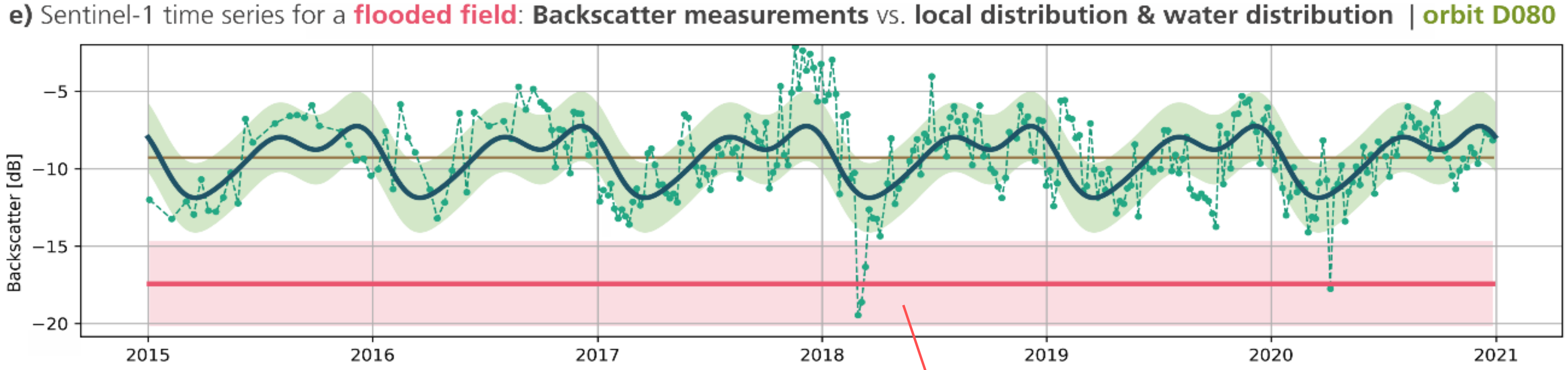


Figure from Bauer-Marschallinger et al. (2022) Satellite-Based Flood Mapping through Bayesian Inference from a Sentinel-1 SAR Datacube, Remote Sensing, 14, 3673

TU Wien Flood mapping algorithm II

- Bayesian decision per (20 m) pixel



- The **Non-flood distribution** is determined by a harmonic model of the local SAR signal — the "SAR seasonality," i.e., for all 365 days of the year, computed using time series analysis from the datacube archive.

- The **flood distribution** is derived from a globally applicable linear SAR model for water surfaces.

- Decision: flood or no flood at this pixel at this time.
- Morphological post-processing for noise suppression.

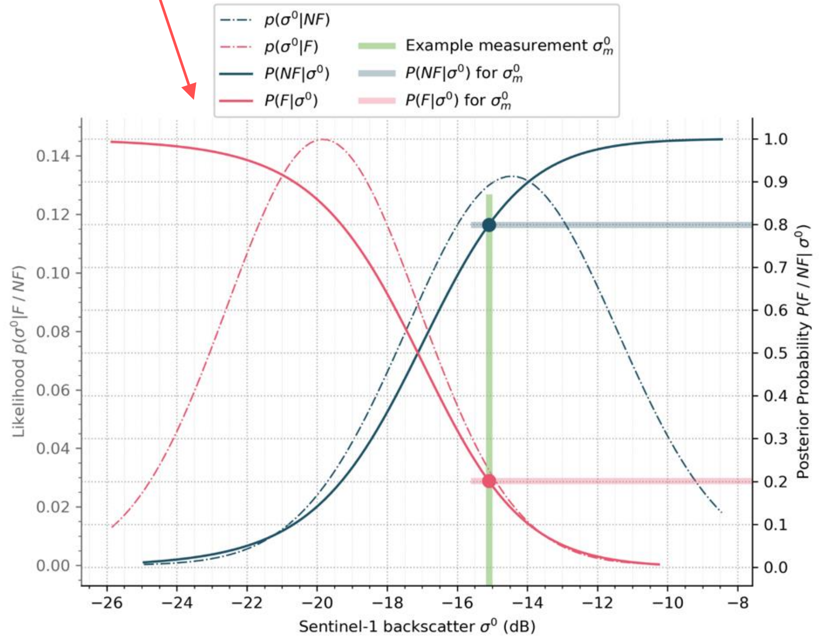


Figure from Bauer-Marschallinger et al. (2022)

The non-flood backscatter probability distributions

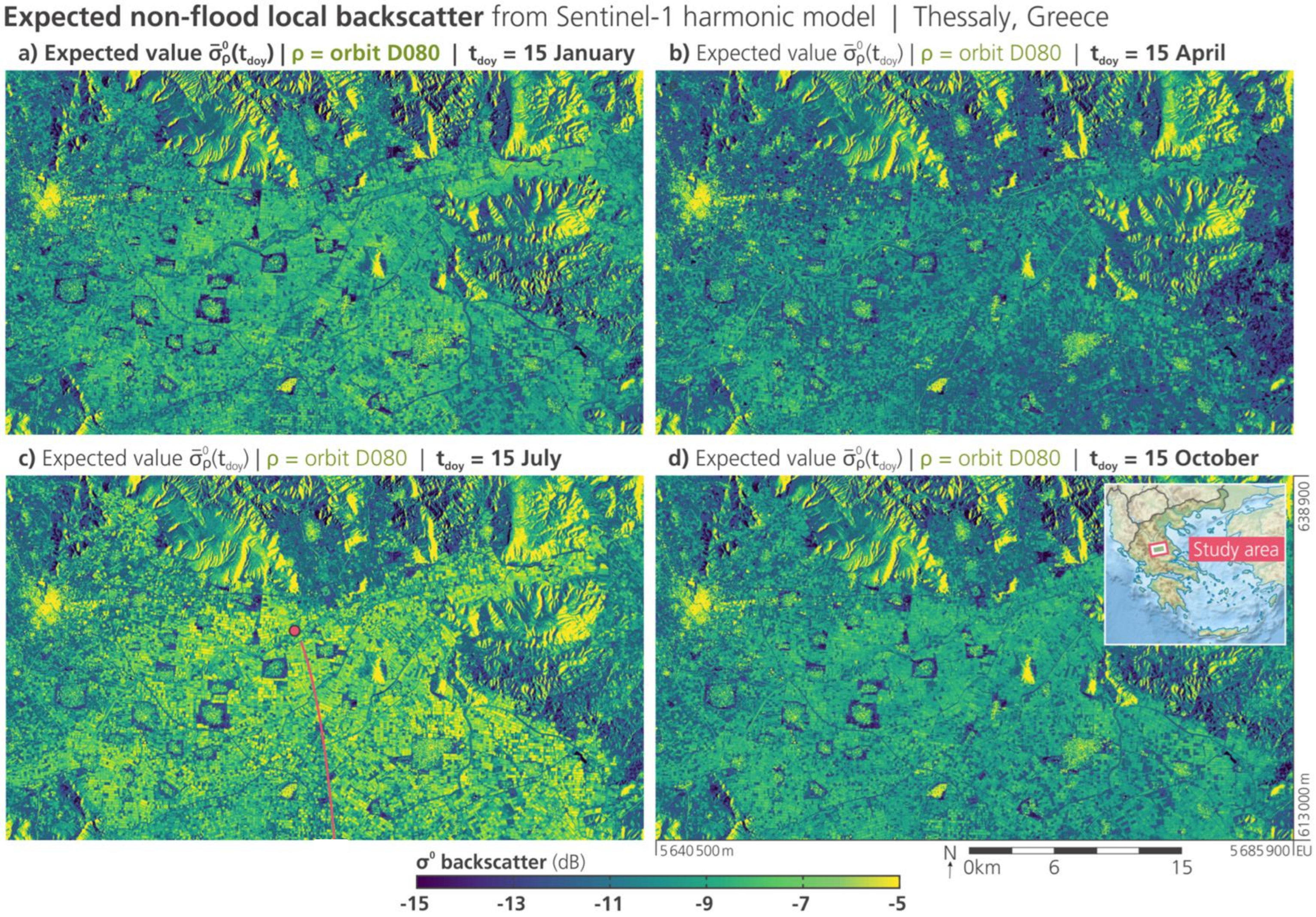


Figure from Bauer-Marschallinger et al. (2022)

The flood backscatter probability distributions

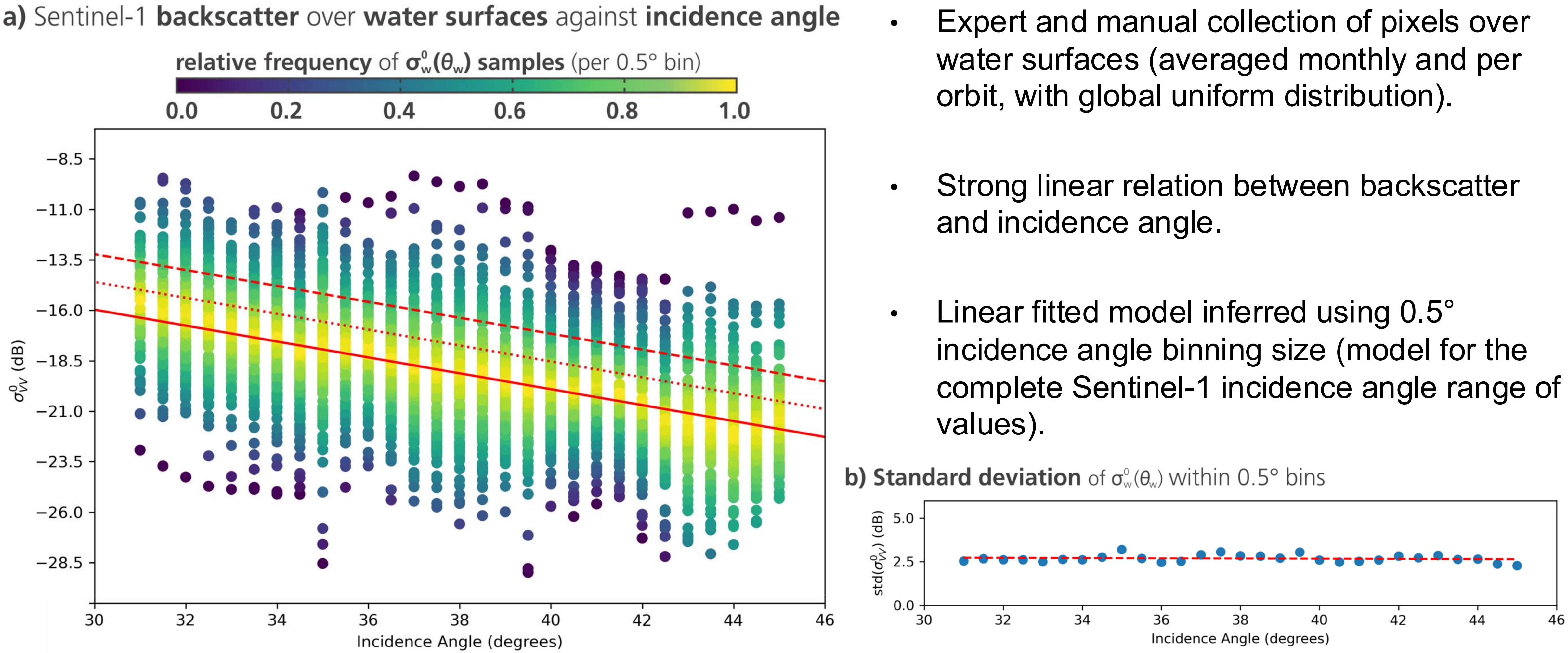


Figure from Bauer-Marschallinger et al. (2022)

- Expert and manual collection of pixels over water surfaces (averaged monthly and per orbit, with global uniform distribution).
- Strong linear relation between backscatter and incidence angle.
- Linear fitted model inferred using 0.5° incidence angle binning size (model for the complete Sentinel-1 incidence angle range of values).

TU Wien Flood mapping algorithm: Thessaly example

- Bayesian decision provides:

- Flood mapping (or no flood)
- a posterior probability of the binary classification

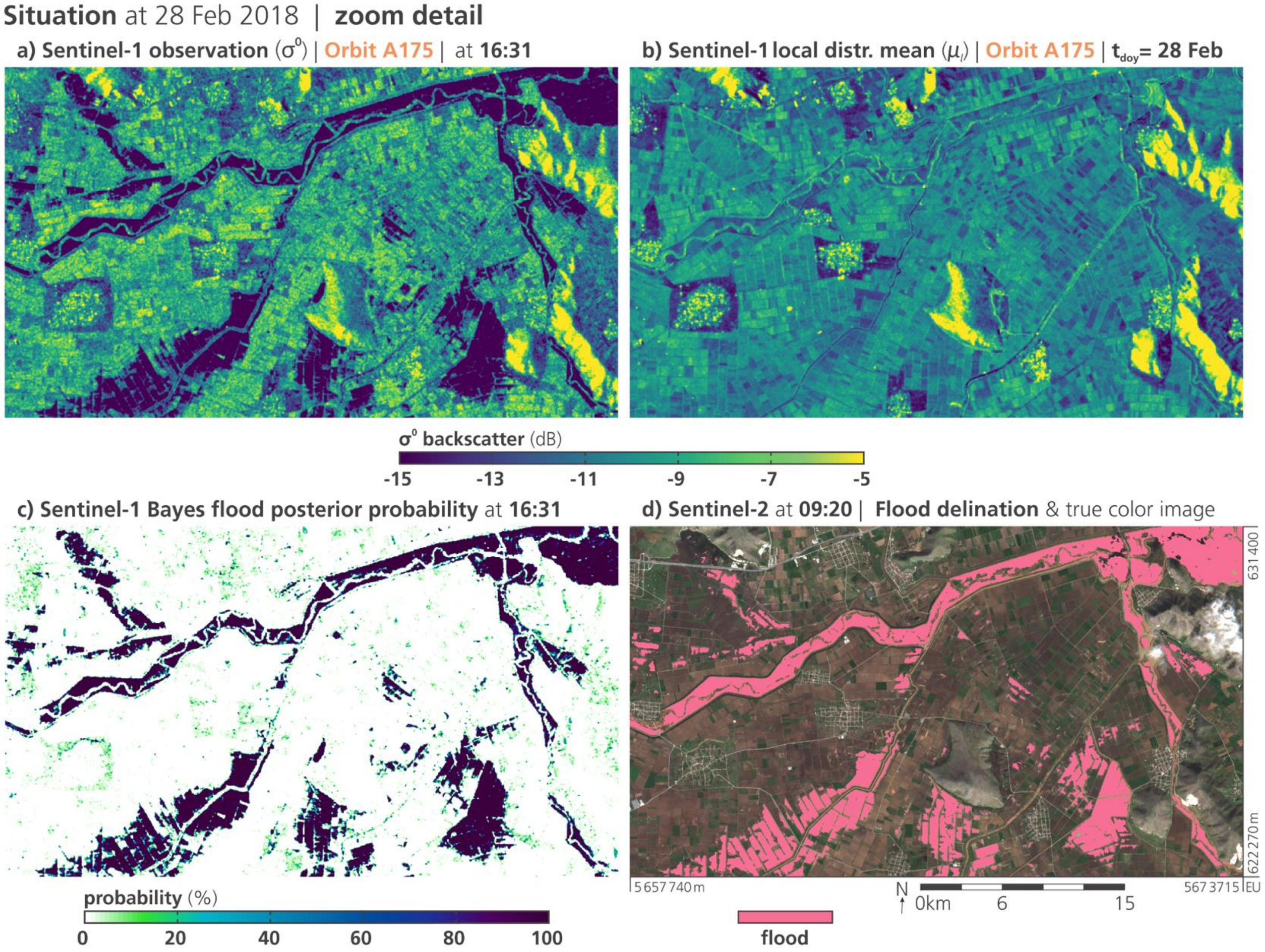
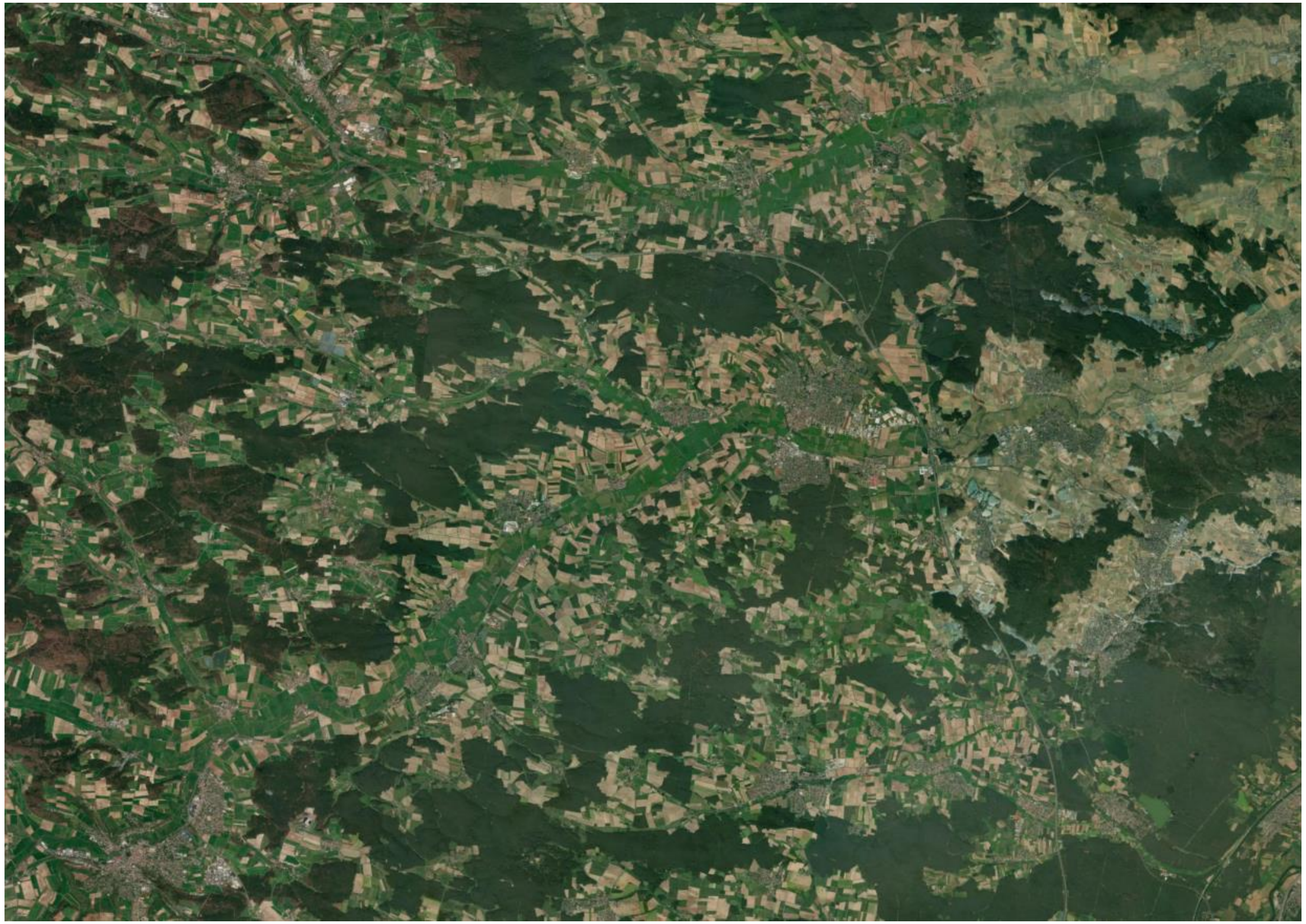


Figure from Bauer-Marschallinger et al. (2022)

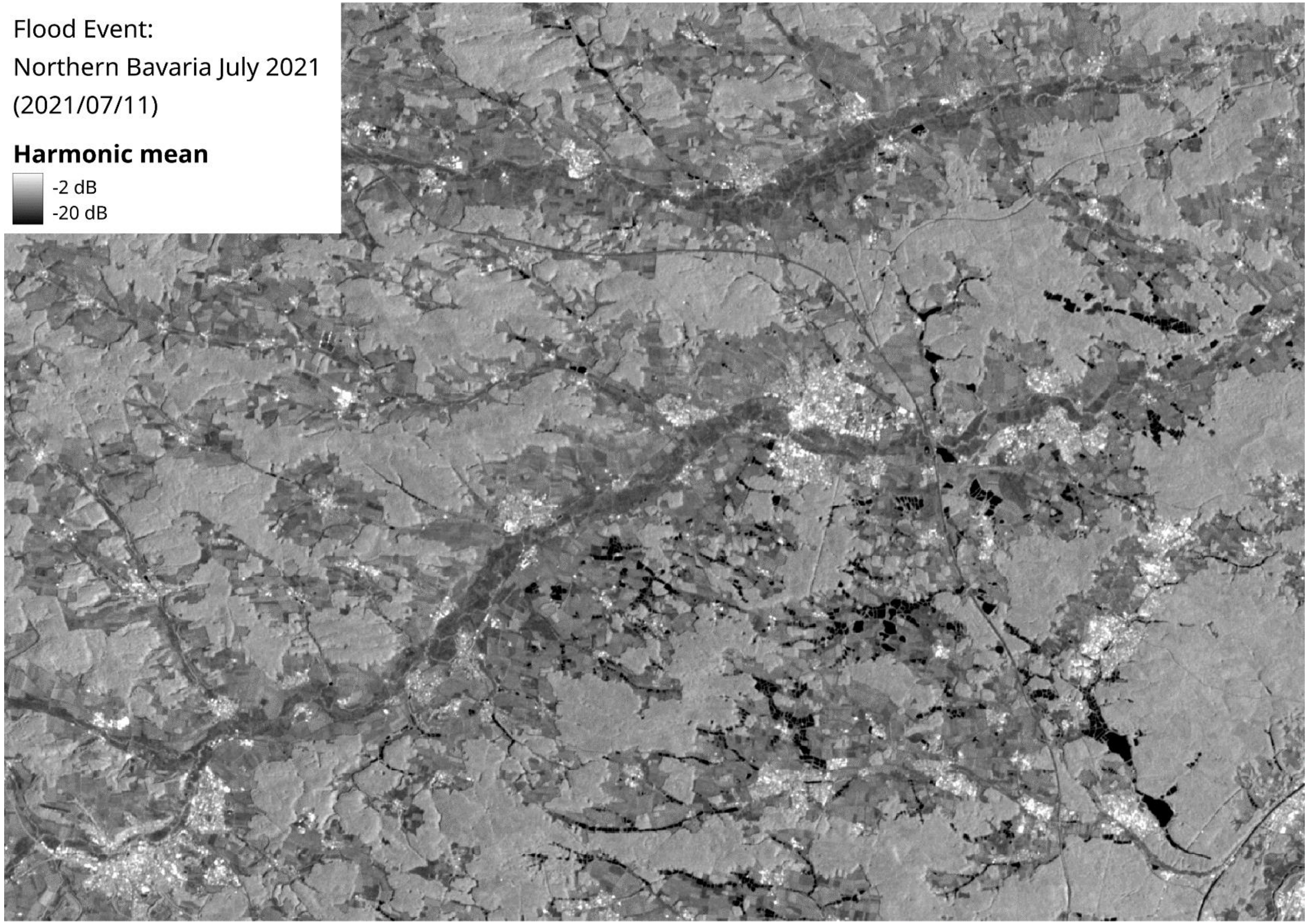
Germany Flood July 2021

“Centennial Flood 2021”
Example near Erlangen,
Bavaria

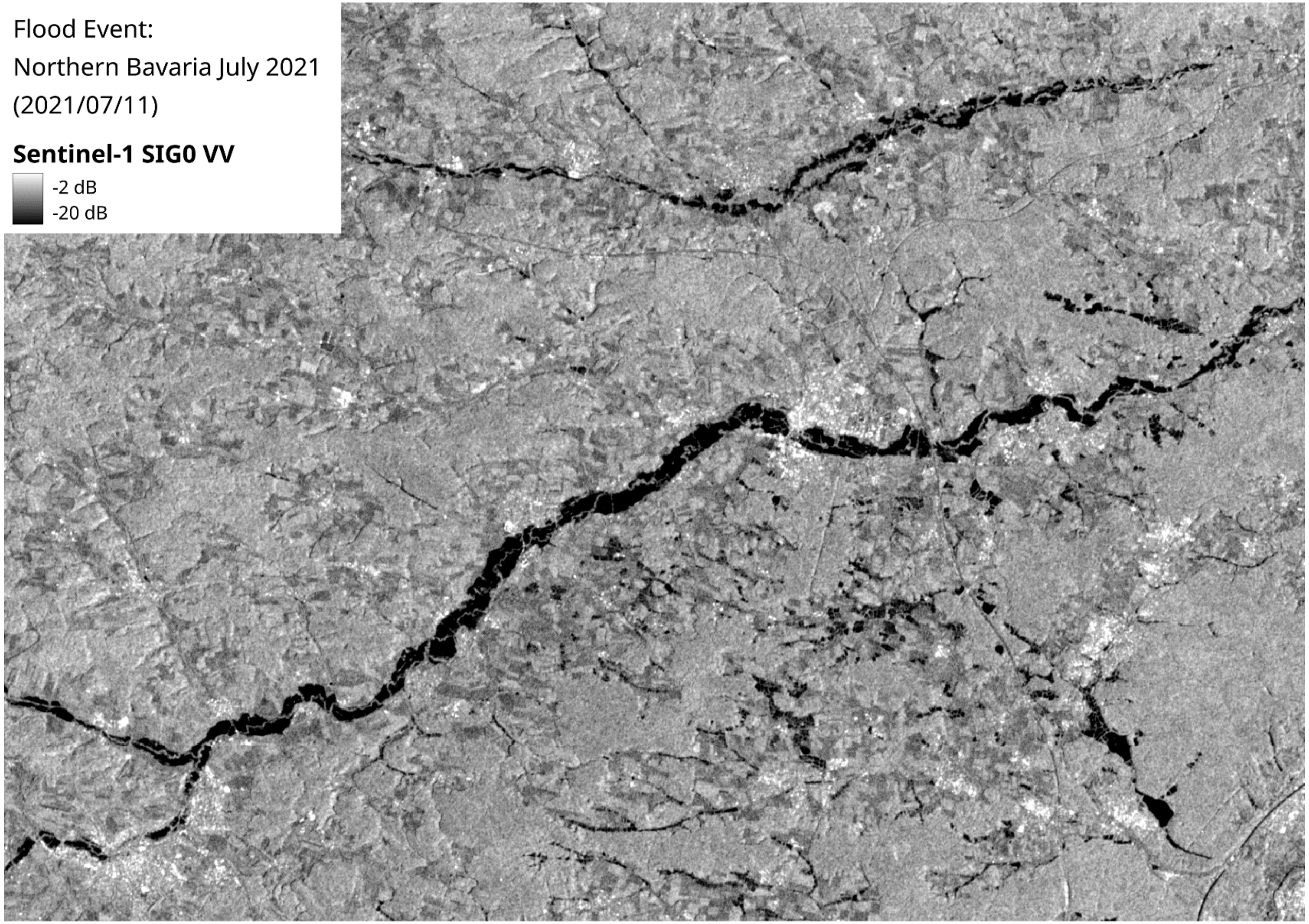


https://www.nordbayern.de/region/erlangen/hochwasser-vollig-verhindern-lasst-sich-das-nicht-1.11215408

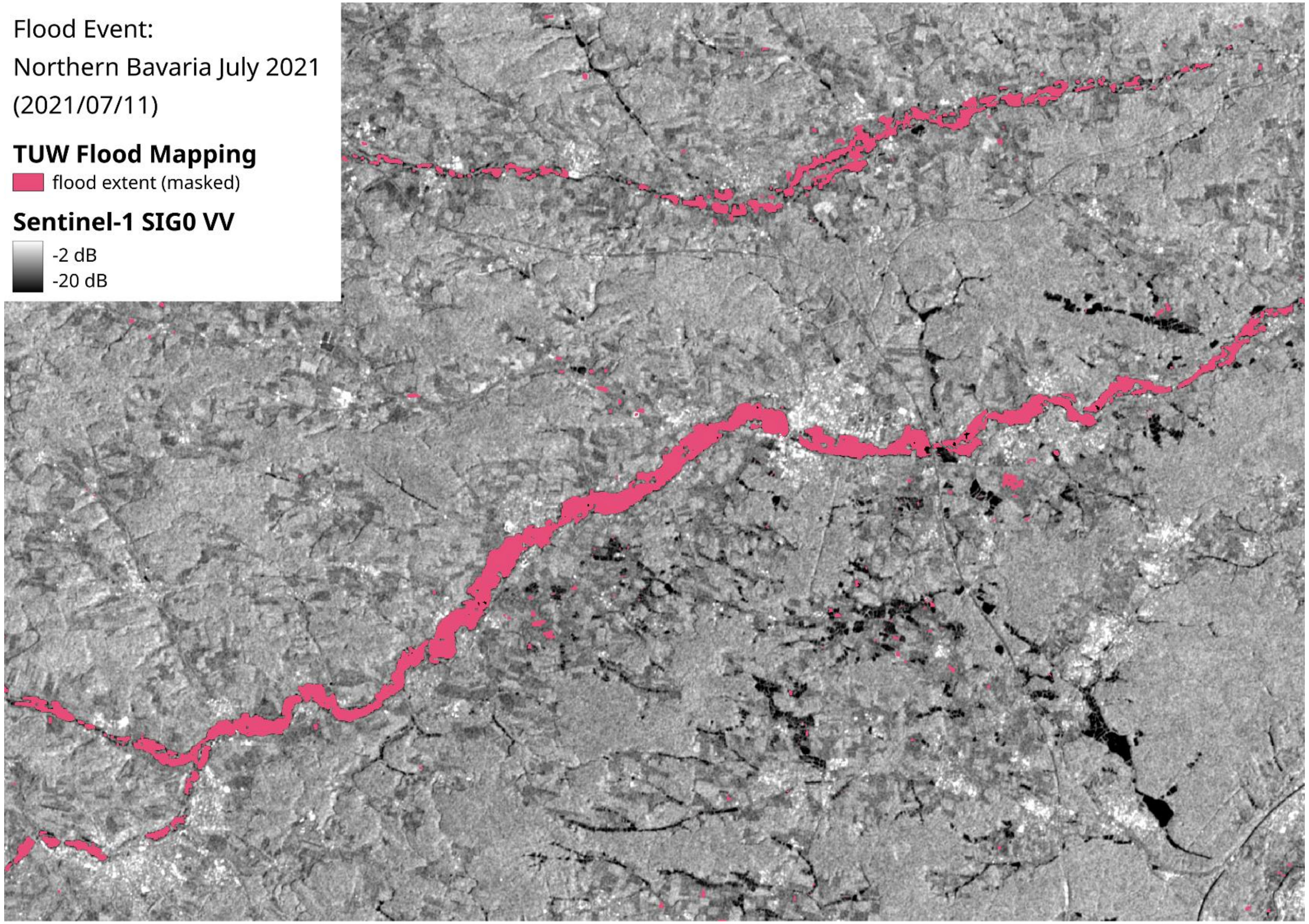
Germany Flood July 2021 – Backscatter Harmonic mean



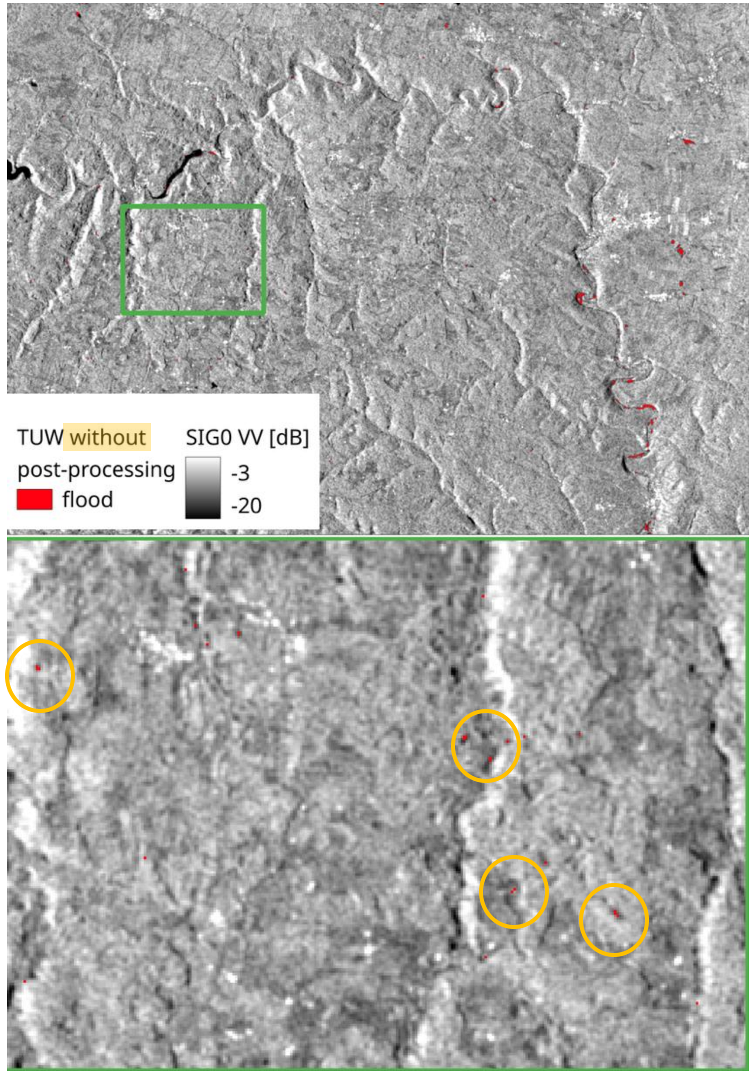
Germany Flood July 2021 – Acquisition during flood

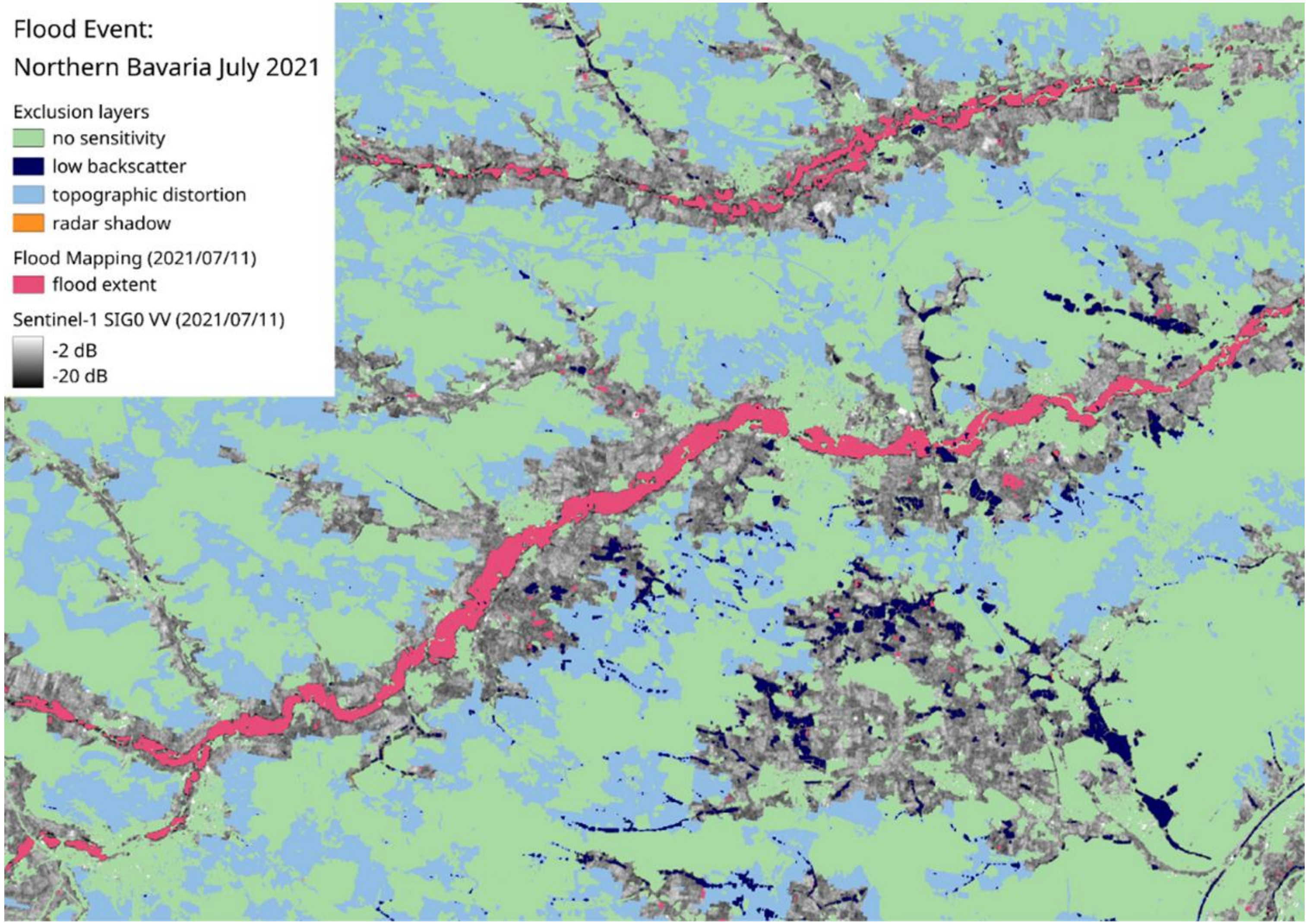


Germany Flood July 2021 – Flooded water detection



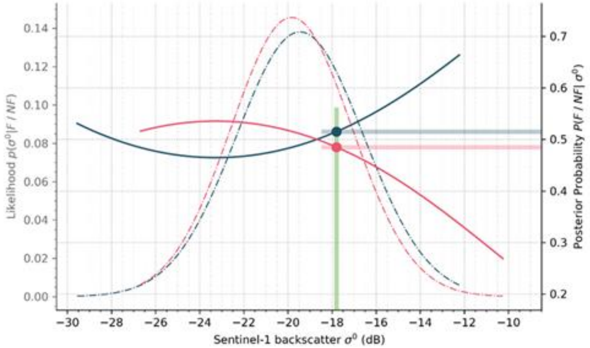
TU Wien Flood mapping algorithm: post processing



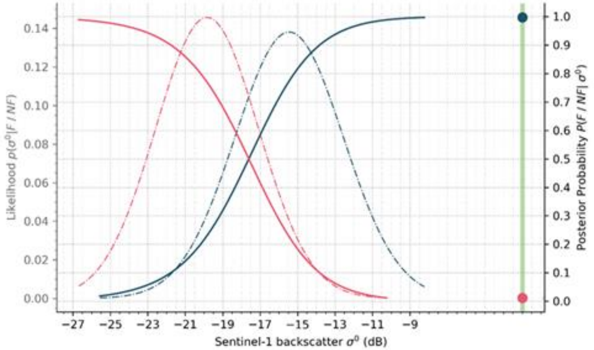


TU Wien Flood mapping algorithm: challenging cases

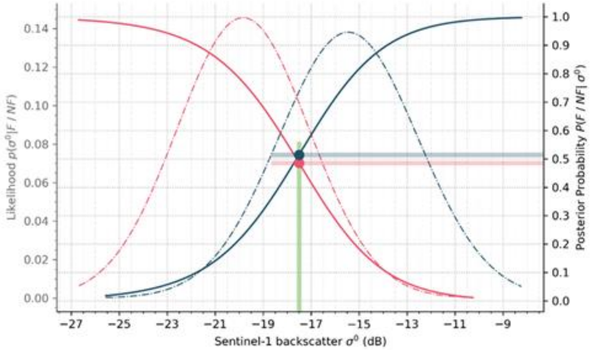
- Conflicting distribution

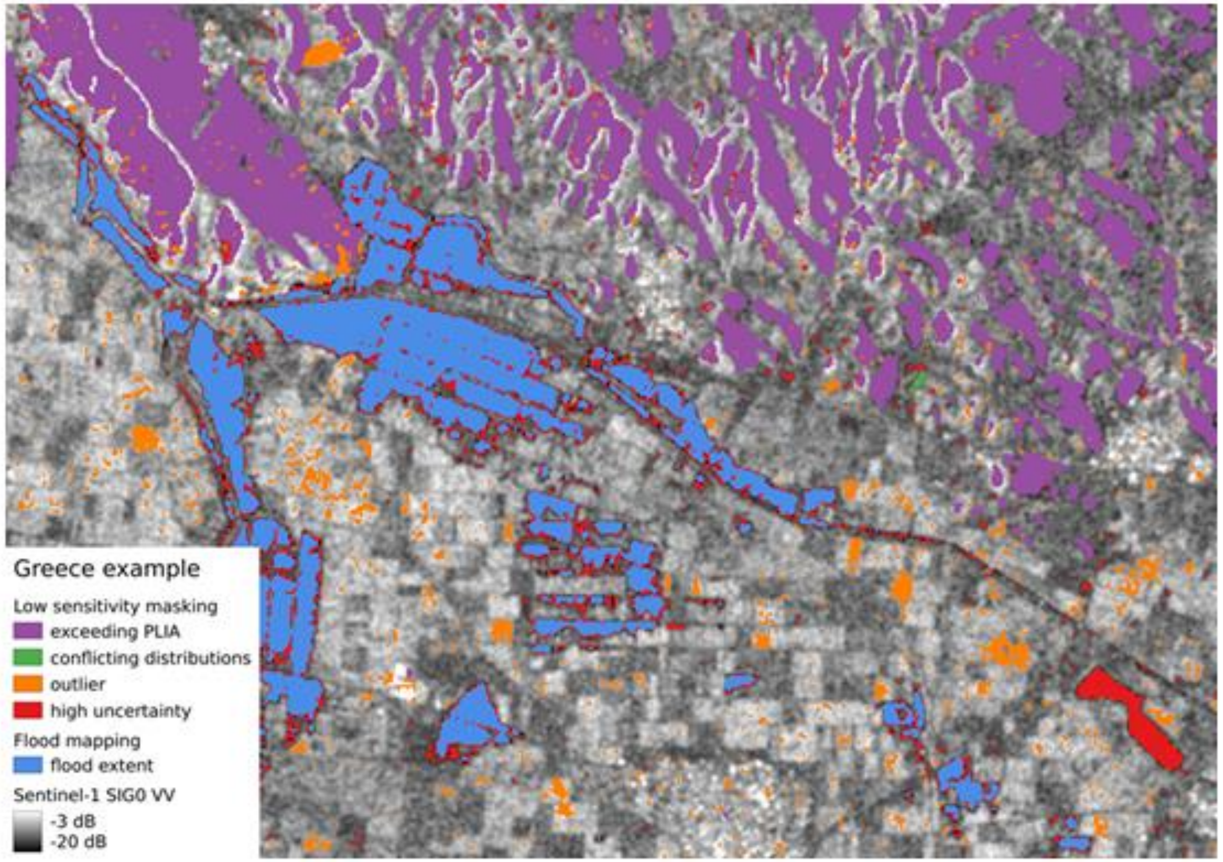


- Detected outlier



- High uncertainty



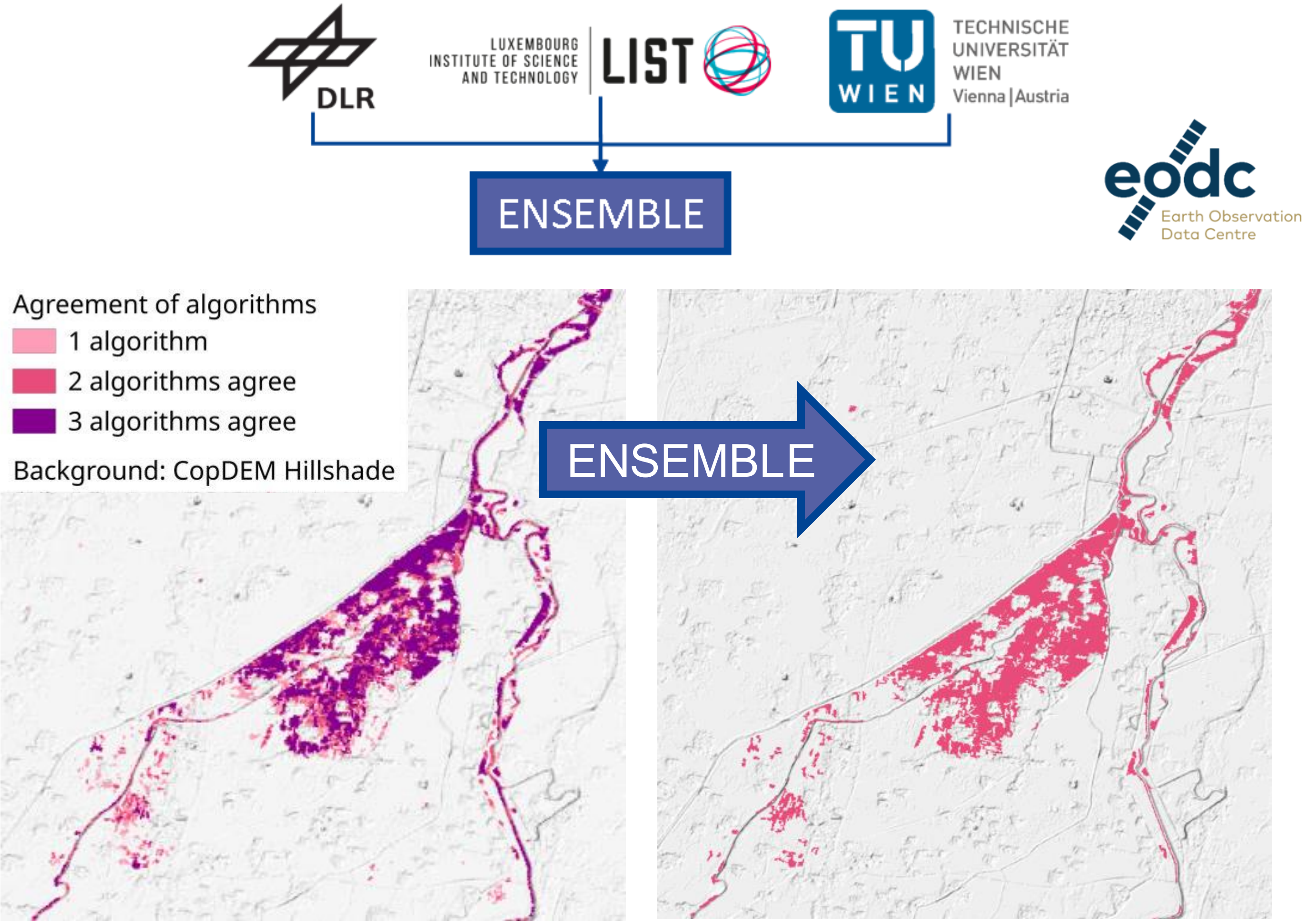


TU Wien Flood mapping algorithm: role in CEMS GFM

- **3 algorithms for Sentinel-1 flood mapping**

- **DLR**: Image classification using fuzzy logic with post classification and region growing
- **LIST**: Change-detection using hierarchical split-based approach
- **TUW**: Bayesian classifier informed by full per-pixel Sentinel-1 signal history (harmonic model)

- **Robustness through ENSEMBLE approach**
  - 3 independent algorithms
  - At least two algorithms must agree
  - Joint flood likelihood layer



Davide Festa
Technische Universität Wien
Department of Geodesy and Geoinformation
Research Group Remote Sensing
davide.festa@geo.tuwien.ac.at
www.tuwien.at

Thais Beham
Technische Universität Wien
Department of Geodesy and Geoinformation
Research Group Remote Sensing
thais.beham@geo.tuwien.ac.at
www.tuwien.at

Tobias Raiger-Stachl
eodc Earth Observation Data Centre GmbH
Vienna, Austria
tobias.raiger-stachl@eodc.eu
www.eodc.eu

Thank you!

Questions?